A Feasible Arbitrage-Free Regime-Switching Model of the Term Structure

Wachindra Bandara Richard Munclinger

March 9, 2012

Abstract

We present an affine, arbitrage-free, regime-switching dynamic Nelson-Siegel model of the term structure. We begin the development of the model in continuous time by presenting a class of general affine hidden Markov models of the term structure. We highlight what assumptions are necessary to reach tractable versions in this class such as the Dai, Singleton and Yang (2007) model and our Nelson and Siegel model. We propose a multi-regime approximate Kalman filter to estimate hidden Markov models. We then estimate the proposed arbitrage-free hidden Markov Nelson Siegel model on historical yield curve data. We contrast the model to single-regime alternatives and conclude that our model performs well in-sample. We find, using likelihood ratio tests, that regimes are driven by long term means and measurement and transition covariance matrices. The regimes conform to the periods of expansionary and restrictive monetary policy, but do not coincide exactly with recessions. This suggests that monetary policy responses over the past three business cycles have been effective, but that those responses have persisted long after the recessions they intended to address have ended.

JEL classification: C53, C58, E43

Keywords: Term structure, Nelson-Siegel models, regime switching, arbitrage free models, Kalman filter

Dynamic Nelson and Siegel models are popular in describing the dynamic properties of the term structure of interest rates. These models were pioneered by Nelson and Siegel (NS) (1987), who proposed a flexible parameterization that describes the cross-sectional shape of the yield curve. The static structure of the model was augmented with dynamic factors by Diebold and Li (2006). Their dynamic Nelson and Siegel model (DNS) adequately describes both the cross-sectional and the time-series properties of the yield curve. Although NS models originated in the yield curve fitting literature, they have been linked to affine term structure models (ATSMs) though the work of Christensen, Diebold and Rudebusch (CDR) (2009, 2010). CDR demonstrated that an arbitrage-free Nelson and Siegel (AFNS) model can be derived if certain restrictions are imposed on the general ATSM structure (characterized by Dai and Singleton (2000)). AFNS models have the advantage, relative to DNS models, of having a clear theoretical link to other equilibrium term structure models, and of imposing desirable arbitrage-free restrictions on the time series dynamics of yields.

Nelson and Siegel models have several advantages over other classes of term structure models. DNS and AFNS models can be cast into a state space representation and are therefore relatively easy to estimate via a Gaussian Kalman filter (KF). Estimation is performed using historical term structure data and consequently, the model reflects both the cross-sectional and time-series properties of interest rates. This is in contrast to arbitrage-free models (such as Hull and White (1993), Heath, Jarrow and Morton (1992)) that are calibrated to cross-sectional term structure data only and to ATSMs that require normalizing and other, typically arbitrary, restrictions for estimation. As a result of their flexibility and tractability, DNS and AFNS model are well suited for risk management and interest rates forecasting. They are also used heavily by institutions that do not require exact recovery of market prices, such as central banks and ministries of finance (for public debt management).

AFNS models, however, have several shortcomings that are shared by many equilibrium term structure models. In particular, AFNS and DNS models assume that the density of interest rates is normal and that the interest rate process is stationary. Nonstationarity can best be understood in this context as the tendency of financial time series to behave differently in recessions than in expansions. Both are strong assumptions and have important implications for risk management. The fact that interest rates are not normal is illustrated in figure 1. The top panel plots the kernel density estimator for short and long-term rates based on monthly data from April 1991 to August 2010. The density estimates are not normal and are bi-modal, a clear indication of non-stationary behavior. The second panel of figure 1 plots the density estimate of the slope factor in the DNS and AFNS models estimated on the same data. This indicates that the bi-modality in interest rates is directly reflected in the dynamic factors.

Nonstationarity in macroeconomic time series has been documented at least since the work of Hamilton (1989, 1990), who pioneered the regime-switching methodology. His work was extended by Kim (1994, 1996), Kim and Nelson (1998, 1999A, 1999B), Gray (1996) and many others. Regime switching models have been used to model a wide variety of macroeconomic phenomena, including interest rates, inflation, economic growth and stock prices. The regime

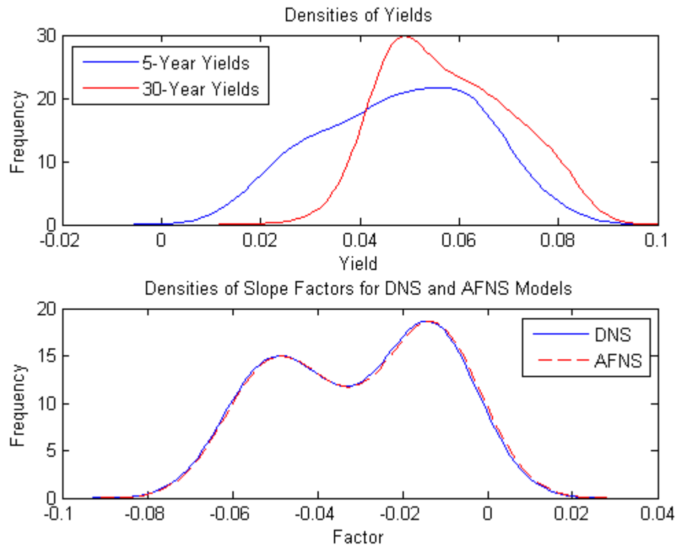

Figure 1: Densities of Yields and Single Regime Models: The top panel of this figure presents the kernel density estimate of the 5-year and 30-year yields and the lower panel plots the densities of the factors in the DNS and AFDNS models.

dependent dynamics have been often attributed to the business cycle. A hidden Markov component was introduced into term structure modeling by Lee and Naik (1998) and Hansen and Paulsen (2000). The origin of the methods based on a marked point process, used in Wu and Zhang (2010) and this paper can be traced to Landen (2000). These methods were also used in Bansal and Zhou (2002) and Bansal, Tauchen and Zhou (2003) and Dai, Singleton and Yang (DSY) (2007). They tested the performance of hidden Markov models and concluded that they are superior, in some respects, to their single-regime counterparts.

In this paper, we address the inability of these models to accurately describe the dynamics of interest rates. First, however, we present a solution to a generalized affine arbitrage free hidden Markov model. The solution takes the form of a pair of ordinary differential equations (ODEs), as is typical for affine models. The general solution however is not practically viable under most circumstances. It contains multiplicative terms which depend of the discrete jumps of the hidden Markov process and are difficult to simplify. We therefore impose several restrictive assumptions under the risk neutral measure to arrive at tractable models.

In this way, we show the assumptions on the general specification that are necessary to obtain the DSY model. Further restrictions akin to those imposed by CDR result in a Nelson and Siegel version of the hidden Markov model. This model maintains the tractable Nelson and Siegel structure and arbitrage free restrictions, while allowing greater flexibility than single-regime models to capture multi-modality, heteroscedasticity and other characteristics of the interest rates.

There have been several papers that model the macro-economy using dynamic Nelson-Siegel models as proposed by CDR highlighting their prominence in the academic and practitioner literature. This paper fits in with a number of investigations that deal with Nelson-Siegel models and extend the work of Diebold and Li (2006) and CDR. A recent paper in this strand of the literature is Bernadell, Coche and Nyholm (2005). These authors extend the work of Diebold and Li to accommodate regime-switching parameters. Another recent paper is Zantedeschi, Damien and Polson (2011) who extend CDR to allow an unknown number of regime changes.

Our paper differs from the existing work on this topic in several important ways. First, we begin the model in continuous time and derive a general continuous-time solution for a general class of affine term structure models. In contrast, DSY begin in discrete time, by assuming a form of the pricing kernel. Second, we use the framework of Nelson and Siegel and results in CDR to infer reasonable restrictions on the regime-switching term structure such that our model is *feasible*. We think of feasibility in this context as the ability of the model to be fitted to the entire term structure (more than a small number of yields) with relative ease. The KF methodology ensures that our model can be fitted to an arbitrary (exceeding the minimum based on the number of latent factors) number of yields. The NS and CDR restrictions ensure that the number of parameters is manageable. In particular, no normalizing or ad hoc restrictions on the term structure dynamics are necessary in our model.

We estimate our model on treasury strip data from 1991 to 2010 and benchmark its in-sample performance to DNS and AFNS models. We use the approximate-maximum-likelihood algorithm (AML) of Kim and Nelson (1999C) and assume that the latent factors are uncorrelated and that transitions under both the real and risk neutral measures are homogeneous. The regime-switching arbitrage-free Nelson and Siegel model outperforms all alternative models, even after we control for the number of parameters in each model. We find that the shape of the term structure implied by the regime switching model is more consistent with the DNS model, in that short term yields are matched well cross-sectionally. The λ parameter and therefore the factor loadings of the regime switching model are also more similar to the DNS model. The estimated yield adjustment term in our model is significantly different to the adjustment in CDR and affects each regime differently.

We identify an expansionary and a contractionary monetary policy regime. These regimes coincide with NBER recession dates, but in general begin at or after the start of a recession and continue long after the official recession is over. This indicates that monetary policy is effective in combating recessions. However, these findings also hint at the fact that monetary policy effects persist long after recessions are over. The regimes appear to be driven by the mean and variance of the factor dynamics, as well as the state dependent measurement covariance error matrices. We find that mean reversion differences are not significant across states. We show, through the use of likelihood ratio tests that significant difference between the regimes exist and that the market price of risk in indeed priced.

The paper is organized as follows. Section 1 presents the generalized affine hidden Markov model in a continuous time setting. Section 2 discusses the restrictions that are necessary to arrive at our feasible version of the general specification. Section 3 outlines the approximate likelihood estimator via Kalman filter and its application to a general regime switching term structure model. Data is summarized in Section 4 and the results of DNS and AFNS model estimation are given in Section 5. Section 6 summarizes our findings from the estimation of the hidden Markov Nelson and Siegel term structure model. In particular Section 6 covers: (i) regime identification and interpretation; (ii) market prices of risk; (iii) comparison to DNS and AFNS models; and (iv) robustness checks via the likelihood ratio test. Section 7 concludes. Appendices can be found at the end of the document.[1]

1 General affine hidden Markov model

Dai and Singleton (2001) characterize affine models and consider a general specification with a general solution due to Duffie and Kan (1996). Although the goal in this paper is not to characterize all possible forms of affine hidden Markov models, it is instructive to begin with a general specification and work progressively towards a tractable model though introducing restriction on the general form. We therefore begin by considering the general specification of affine models in Dai and Singleton (2000) in combination with a general regime switching process.

We define a market embedded in a filtered probability space $(\Omega, \mathcal{F}, ((\mathcal{F}_t)_{t\in[0,\mathcal{T}]}), \mathcal{P})$. We assume that the basis admits a n-dimensional standard Wiener process W and a m-dimensional Markov process, under the risk neutral measure Q. The Wiener process drives the dynamics of the state variables X, which follow a general affine diffusion given by

$$dX_t = \tilde{\kappa}(s_t)(\tilde{\theta}(s_t) - X_t)dt + \Sigma(s_t)\sqrt{V(t, s_t)}d\tilde{W}_t. \tag{1}$$

$\Sigma(s_t)$ is a non-diagonal, and possibly asymmetric matrix, $\tilde{\kappa}(s_t)$ is the speed of mean reversion, $\tilde{\theta}(s_t)$ is the long term mean of X_t and $V(t, s_t)$ is a diagonal matrix with each element of the diagonal given by

$$V_{ii} = a_i(s_t) + b_i(s_t)'X_t. \tag{2}$$

$d\tilde{W}_t$ is an n-dimensional correlated Brownian motion. This specification is identical to that in Dai and Singleton (2000), with one exception. Each coefficient is dependent on a continuous time Markov chain s_t, with a Markov state space $\{0, 1, 2, ..., (m-1)\}$. An n-factor regime switching affine model is also based on the assumption that the instantaneous short rate is a linear function of state variable $X(t) = (X_1(t), \dots, X_n(t))'$,

$$r(t) = \delta_0(s_t) + \delta_1(s_t)'X(t). \tag{3}$$

Following the specification in Landen (2000) and Wu and Zeng (2010), we assume that regime

[1]In this paper we use the terms hidden Markov model, regime switching model and AFNSRS model interchangeably.

transitions are governed by the stochastic differential equation

$$ds_t = \int_E \zeta(z)\tilde{\mu}(dt, dz), \tag{4}$$

where the mark space E is defined as $E = \{(i, j) : i \in \{0, 1, 2, , m-1\}, j \in \{0, 1, 2, , m-1\}, \forall i \neq j\}$. $\tilde{\mu}$ is a marked point process on the mark space $(E, 2^E)$, where 2^E is the $\sigma-$ algebra of E. Each point in the process $z \in (i, j)$ represents a transition from regime i to regime j and is described as a point in E. Any marked point process can be uniquely defined through its stochastic intensity kernel. We define the stochastic intensity kernel of E by

$$\tilde{\gamma}(dt, dz) = \tilde{h}(z, X(t-))\mathbf{I}\{s(t-) = i\}\epsilon_z(dz)dt, \tag{5}$$

where $\tilde{h}(z, X(t-))$, with $z = (i, j)$, is the transition intensity from regime i to j. The function $\tilde{\gamma}$ can be interpreted as the conditional probability of a shift from regime i to regime j in the interval $[t, t + dt)$, conditional on $X(t-)$ and $s(t-) = i$. Similarly, $\tilde{h}(z, X(t-))$ can be interpreted as the probability of transition within a unit time ($dt = 1$) given $X(t-)$ and $s(t-) = i$.

Finally, we appropriate the form suggested by Wu and Zeng (2010) and assume that the log of the regime switching intensity is also affine in the factors.

$$\tilde{h}(z, X) = e^{\tilde{h}_0(z) + \tilde{h}_1(z)'X} \tag{6}$$

We note that $\tilde{h}(z, X(t))$ is both state and time dependent. This form proves to be particularly tractable since it is consistent with the form of the bond price derived later.

This framework specifies a general affine regime switching model. This structure leads us to propose a solution to the price of a zero-coupon, unit face value, riskless bond in regime $s(t)$ with τ periods to maturity as

$$P(s(t), X(t), \tau) = e^{A(\tau, s_t) + B(\tau, s_t)'X(t)}, \tag{7}$$

where $A(\tau, s_t)$ and $B(\tau, s_t)$ are respectively a vector and a matrix of state dependent coefficients. At maturity, the payoff of the bond must be unity. Therefore, $A(0, s_t) = 0$ and $B(0, s_t) = 0$. It turns out that, consistent with the single regime model, this model can be solved in close form, up to a pair of differential equations.

Proposition 1: *If the framework is given by (1), (2), (3), (4) and (6), the price of a zero unit face value, riskless bond in regime $s(t)$, with τ periods to maturity is given by (7), where $A(\tau, s_t)$ and $B(\tau, s_t)$ solve the differential equations*

$$-\frac{\partial A(\tau, s_t)}{\partial \tau} + [\tilde{\kappa}(\theta(\tilde{s_t})]'B + \frac{1}{2}\sum_{i=1}^{n}[B'\Sigma(s_t)]_i^2 a_i(s_t) + \int_E [e^{\Delta_s A(\tau, s_t)} - 1]e^{\tilde{h}_0}(dz) = \delta_0 \tag{8}$$

$$-\frac{\partial B(\tau, s_t)}{\partial \tau} - \tilde{\kappa}'B + \frac{1}{2}\sum_{i=1}^{n}[B'\Sigma(s_t)]_i^2 b_i(s_t) + \int_E [e^{\Delta_s A(\tau, s_t) + \tilde{h}_0}(\Delta_s B + \tilde{h}_1) - e^{\tilde{h}_0}\tilde{h}_1](dz) = \delta_1 \tag{9}$$

Proof: See Appendix.

1.1 Admissibility and Identifiability

An affine regime switching model is admissible if $V_{ii}(t, s_t) > 0$ for all regimes and $i \in 1 \ldots n$. A model is automatically admissible if $b_i(s_t) = \vec{0}$. In the case where $b_i(s_t) > \vec{0}$, the model will be admissible if $a(s_t) \geq 0$ and when X_t is guaranteed to be positive. This implies that the mean of X_t must be always non-negative and that the variation at the boundary collapses to zero. This can be ensured by requiring that: (i) $\tilde{\theta}(s_t)\tilde{\kappa}(s_t) \geq \vec{0}$; (ii) $\Sigma_{ij}(s_t) = 0$ for $1 \leq i \neq j \leq n$; and (iii) $\kappa_{ij}(s_t) \leq 0$ for $1 \leq i \neq j \leq n$. These conditions are equivalent to the single regime case and must hold for each regime.

The regime switching model, in its most general form is not econometrically identifiable. This is because, like other affine models, the model permits *invariant affine transformations*. Since latent term structure factors are not observed, their scale and location and volatility is not unique. Infinite number of solutions are therefore possible, with all being an affine transformation of any other solution. In order to ground the model parameters, affine models need to be *normalized*. In particular, we have to address: (i) the scale of X_t; (ii) the level of X_t; (iii) interdependencies of state variables; (iv) signs of state variables; and (v) rotation of the Brownian motions. A large number of normalizations are possible and will not be outlined here. Dai and Singleton (2000) and Blais (2009), present some candidates that meet the necessary criteria. In regime switching models, however, parameters in only one out of the possible m regimes must be normalized. The normalization in one regime ensures that the factors are pinned down in other regimes as well.

1.2 Change of measure and market prices of risk

In this section, we consider the structure of the pricing kernel, the change of measure, and the uniqueness of the equivalent martingale measure (EMM). The pricing kernel in this framework, as derived in the appendix, is given by

$$\begin{aligned} M_t &= exp\Big(- \int_0^t (r_s + \frac{1}{2}\lambda\lambda')ds - \frac{1}{2}\int_0^t \lambda' d\tilde{W}_t \\ &+ \int_0^t \int_E [\psi(z, X_{t-})\tilde{\gamma}(ds, dz) + log(1 - \psi(ds, dz))\tilde{\mu}(ds, dz)] \Big) \end{aligned} \tag{10}$$

Details of the derivation are given in the appendix.

So far we have specified the dynamics under the risk free measure Q. The estimation of the model takes place under the real world measure P. In order to specify the P dynamics, we assume that the market price of factor risk is affine in X_t and that it is given, in general form, by

$$\lambda(t, s_t) = (\Sigma(s_t)\sqrt{V(t, s_t)})^{-1}(\lambda_0(s_t) + \lambda_1(s_t)X_t). \tag{11}$$

Other parameterizations are possible, but (11) is particularly general since it maintains affine structure for all cases where $b(s_t) \geq 0$. Substituting

$$dW_t = \lambda_t dt + d\tilde{W}_t \tag{12}$$

into (1), where lambda is given in (11), leads to the dynamics of X_t under the physical measure,

$$dX_t = \kappa(s_t)(\theta(s_t) - X_t)dt + \Sigma(s_t)\sqrt{V(t, s_t)}dW_t \tag{13}$$

where

$$\lambda_0(s_t) = \tilde{\kappa}(s_t)\tilde{\theta}(s_t) - \kappa(s_t)\theta(s_t) \tag{14}$$

$$\lambda_1(s_t) = \kappa(s_t) - \tilde{\kappa}(s_t). \tag{15}$$

Market prices of factor risk are not observable and can be inferred from the other model parameters through (14) and (15). We also assume that the market price of regime switching risk has the following structure,

$$1 - \psi_t = e^{\psi_0(s_t) + \psi_1(s_t)' X_t}. \tag{16}$$

Note that the marked point process $\tilde{\mu}$ is characterized by its stochastic intensity kernel $\tilde{\gamma}$ which in turn is fully described by $\tilde{h}$ as shown in (5). The change of measure process can be implemented by using the relationship $\tilde{h} = (1 - \psi)h$.[2] We take the logs of this relationship to get

$$log(\tilde{h}(s(t), X(t))) = log(1 - \psi(t, s(t))) + log(h(s(t), X(t))). \tag{17}$$

Using (6), (16) and choosing

$$\psi_0(s_t) = \tilde{h}_0(s_t) - h_0(s_t) \tag{18}$$

$$\psi_1(s_t) = \tilde{h}_1(s_t) - h_1(s_t) \tag{19}$$

we can define the stochastic intensity kernel $\gamma(dt, dz)$ under the physical measure using the regime switching intensity

$$h(s_t, X) = e^{h_0(s_t) + h_1(s_t)' X}. \tag{20}$$

This discussion shows how the assumed structure for the factors and the regime switching intensity allows relatively easy transition between the physical and risk-neutral measures.

In this paper, we do not discuss the existence and uniqueness of the pricing kernel or EMM. This is a nontrivial discussion in the presence of jump processes. An excellent treatment of this topic is in Bjork, Kabanov, Runggaldier (1997), who show that with a finite mark space of the type we consider in this paper, market completeness is a necessary and sufficient condition for the uniqueness of the EMM. We consider a further discussion on this topic beyond the scope of this paper and assume that the market is (sufficiently) complete in subsequent discussions

[2] The change of intensity is described in Bremaud (1981). We refer the interested reader to Chapter VI for a complete characterization of the mechanism.

In this section we have shown that regime dependence under Q does not necessarily mean regime dependence under P and vice versa. Restrictions on the market prices of risk can be imposed in such a way that this separation of dependence is possible. This fact will be used later to simplify the models under the measure Q, while maintaining some salient characteristics of the term structure under the physical measure P.

2 More tractable models

The general model can be restricted in a number of ways to yield different specifications. Multi-regime models analogous to those characterized in Dai and Singleton (2000) are therefore possible. However, affine models are known to suffer some instabilities. Certain parameters may be highly correlated leading to flat likelihood surfaces. These shortcomings are magnified for regime switching models, where the number of parameters is typically larger than in single regime models. Additionally, there are terms in the general solution that make estimation particularly difficult. In this section we will consider two models which are relatively simple, but allow sufficient flexibility to capture regime switching behavior in interest rates.

We focus our efforts on (9), since as will be shown later, (8) can be conveniently simplified. The particular problem is the integral term, that contains $\Delta_s A(\tau, s_t)$ and $\Delta_s B(\tau, s_t)$ as multiplicative terms. With these terms in the expression, it is not possible to obtain relatively simple solutions and a tractable model. We can eliminate some of this complexity by making some simplifying assumptions.

First, we assume that the speed of mean reversion, $\tilde{\kappa}$ is not regime dependent. Second, we assume that the regime switching intensity $(\tilde{h})$ under the risk neutral measure is not dependent on X_t i.e. $\tilde{h}_1 = \vec{0}$, and third we assume that $V_{ii}(t, s_t)$ for $i \in 1 \dots n$ is independent of both s_t and X_t i.e. $b_i = \vec{0}$ for $i = 1 \dots n$. Additionally, we assume that $a_i(s_t) = 1$ for $i \in 1 \dots n$ which is necessary because $a_i(s_t)$ and $\Sigma(s_t)$ can not be uniquely identified. Although these assumptions seem strong they are not as restrictive as they appear.

Although $\tilde{\kappa}$ is regime independent, an appropriate choice of the market price of risk parameters allows κ to be regime dependent. Therefore, mean reversion under the real world measure will exhibit dependency on the current state. The assumption on the regime switching intensity also imposes a restriction only under the measure Q. Under the risk neutral measure, the Markov chain is homogeneous, while under the measure P it can be heterogeneous and dependent on s_t. The last assumption is binding in that the covariance matrix of factors is independent of the value of s_t under both measures. This implies that the distribution of the factors and therefore interest rates is no longer conditionally (non-central) Chi-squared, but is conditionally Gaussian

Finally, we assume that the factor coefficients are regime independent i.e. $\Delta_s B(\tau, s_t) = \vec{0}$.

Consequently, $\Delta_s B(\tau, s_t) = \vec{0}$ and since $\tilde{h}_1 = \vec{0}$ and we have

$$\int_E [e^{\Delta_s A(\tau,s_t)+\tilde{h}_0}(\Delta_s B(\tau, s_t) + \tilde{h}_1) - e^{\tilde{h}_0}\tilde{h}_1](dz) = 0. \tag{21}$$

We summarize the preceding discussion in the following corollary.

Corollary 1: *Under the framework given in (1)-(6) and the regime independence of $\tilde{\kappa}$ and $B(\tau)$ the factor independence of Σ, $\tilde{h}_1$ and V, (8) and (9) can be written as a system of two differential equations*

$$\frac{dA(\tau, s_t)}{dt} + (\tilde{\kappa}\tilde{\theta(s_t)})'B(\tau) + \frac{1}{2}\sum_{i=1}^{n}[B'\Sigma(s_t)]_i^2 + \int_E [e^{\Delta_s A(\tau,s_t)} - 1]e^{\tilde{h}_0}(dz) = \delta_0 \tag{22}$$

$$\frac{dB(\tau)}{dt} - \tilde{\kappa}B(\tau) = \delta_1 \tag{23}$$

The expressions (22) and (23) are a continuous-time analog to the system of ODEs that DSY obtain in discrete time.

The DSY model is much more general. However, even in this form a large number of parameters need to be estimated, even when the number of regimes is small. We propose additional restrictions that yield an arbitrage free, dynamic, hidden Markov Nelson and Siegel model of the term structure as outlined in corollary 1.

2.1 Nelson and Siegel hidden Markov model

The general form of the Nelson-Siegel model is given by

$$y(T-t) = L_t + S_t(\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}) + C_t(\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}) \tag{24}$$

where L_t, S_t and C_t are the level, slope and curvature factors respectively. We can associate L_t with $X_{1,t}$, S_t with $X_{2,t}$ and C_t with $X_{3,t}$ and set $n = 3$. It is easy to see that the instantaneous yield (as $T - t \to 0$) is given by $r_t = L_t + S_t$. To harmonize the notation, we let $r_t = X_1 + X_2$ and therefore $\delta_0 = 0$ and $\delta_1 = (1, 1, 0)'$.

Additionally, the mean reversion parameter ($\tilde{\kappa}$) and the long run mean ($\tilde{\theta}$) of the state variable diffusion in (1) are chosen such that the system given by expression (23) has the the same structure as equation (24). CDR propose a restriction on $\tilde{\kappa}$ for the single regime case. The convenient assumptions made in the previous section, however, ensure that this restriction also works here. We therefore choose

$$\tilde{\kappa} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix}. \tag{25}$$

With Proposition 1 and the above assumptions in hand, we can state the solution of an arbitrage free Nelson and Siegel hidden Markov model.

Proposition 2: *If the instantaneous risk-free rate is given by* $r_t = X_1 + X_2$ *and state variables follow the diffusion given by*

$$\begin{pmatrix} dX_1 \\ dX_2 \\ dX_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \left[\begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \tilde{\theta}_3 \end{pmatrix} - \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \right] dt + \Sigma \begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \end{pmatrix} \tag{26}$$

the price at time t *of a risk-free pure discount bond with maturity at time* T *is given by* $P(s(t), X(t), T) = e^{A(T,s_t)+B(T,s_t)'X_t}$ *where* $A(T, s_t)$ *and* $B(T, s_t)$ *are given by*

$$\begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} = \begin{pmatrix} -(T-t) \\ -\frac{1-e^{-\lambda(T-t)}}{\lambda} \\ (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda} \end{pmatrix} \tag{27}$$

and

$$A^j_{n+1} = -(\tilde{\kappa}\tilde{\theta}(s_t))'B - \frac{1}{2}B'\Sigma(s_t)\Sigma(s_t)'B - log\left(\Sigma(s_t)^S_{k=0}\pi^{jk}e^{-A^k_n}\right) \tag{28}$$

where A^j_n *denotes the the constant term of a risk-free pure discount bond with* n *periods to maturity in state* j*, and* π^{jk} *is the transition probability from state* j *to state* k*.*

Proof. If $\tilde{\kappa}$ is given by (25), we can write (23) as

$$\begin{pmatrix} dB_1/dt \\ dB_2/dt \\ dB_3/dt \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix}. \tag{29}$$

This system of ordinary differential equations solves to (27), with the boundary conditions $B_i(T, T) = 0$ for $i = 1, 2, 3$. This proves the first statement in Proposition 2.

To show (27), we consider the dynamics of the yield adjustment term in Proposition 1, given by (22). The first term in this expression denotes the rate of change of A with respect to time and can be discretized to yield

$$\frac{dA}{dt} = A^j_{n+1} - A^j_n. \tag{30}$$

This denotes the change in A over a unit period, given that the current state is j.

The final term on the left hand side of equation (22) is

$$\int_E [e^{\Delta_s A} - 1] e^{\tilde{h}_0}(dz) \tag{31}$$

From the definition of regime switching intensity in section 1 , the term $e^{\tilde{h}_0}$ can be interpreted as the probability of a transition from state j, under the risk-neutral measure [3], denoted by π^Q. Introducing this notation yields

$$\int_E e^{\Delta_s A} \pi^Q(dz) - \int_E \pi^Q(dz) \tag{32}$$

We add and subtract the probability of not transitioning, $\pi^Q_{j,j}$ and since $\int_{E \cup \{j,j\}} \pi^Q(dz) = 1$[4] we get

$$\int_E e^{\Delta_s A} \pi^Q(dz) - 1 + \pi^Q_{j,j} \tag{33}$$

If transition does not take place, $\Delta_s A = 0$ and we can therefore move $\pi^Q_{j,j}$ into the integral. Simlifying and using the first order Taylor series approximation of the log function yields

$$\int_{E \cup \{j,j\}} e^{\Delta_s A} \pi^Q(dz) - 1 \approx log \left(\int_{E \cup \{j,j\}} e^{\Delta_s A} \pi^Q(dz) \right) \tag{34}$$

The integral over the entire mark space in discrete time corresponds to a summation over all possible states. This allows us to replace the integral with a summation. Furthermore, we have $\Delta_s A = A^j_n - A^k_n$. Hence, we can write the right hand side of (34) as

$$log \left(\sum_{k=0}^{S} e^{-A^k_n} \pi^Q \right) + A^j_n \tag{35}$$

Substituting (35) and (30) into (22) proves (28). □

The form of the intercept term in Proposition 2 is identical to the solution in DSY. Here we show from continuous time how this form arises. The intercept term (28) differs considerably from the intercept derived in CDR. The CDR intercept is obtained in close form and does not depend on $\tilde{\kappa}\tilde{\theta}$ since the authors assume $\tilde{\theta}$ to be a null vector. In our solution, the value of the intercept has to be obtained numerically by recursion. This is an unfortunate byproduct of the dependence on the log term in (28). This is not unduly onerous, since typically, a relatively small number of maturities are used in estimation.

[3]Note that in the discrete time setting $dt = 1$. Hence the interpretation of $\tilde{h}$ (and by assumption $e^{\tilde{h}_0}$) as the probability of regime switching is intuitive.

[4]The singleton is a measure-zero set; $\int_E \pi^Q(dz) = \int_{E \cup \{j,j\}} \pi^Q(dz) = 1$.

3 Estimation

3.1 State space representation

The model characterized in expressions (26), (27) and (28) can be cast in the state space form. It is customary to model interest rates rather than the bond price directly. We can obtain the expression for the yield by taking the log of the bond price $P(s(t), X(t), \tau)$ as given in (7) and dividing it by the time to maturity.

$$R(t,\tau) = -\frac{ln(P(s(t), X(t), \tau))}{\tau} = -\frac{A(\tau, s_t)}{\tau} - \frac{B(\tau)' X_t}{\tau} \tag{36}$$

Typically, a number of yields are available for estimation- each corresponding to a different maturity. We sort the yields by their maturity and stack them into a vector $R(t)$. We define the vector $\dot{A}(s_t)$ and a matrix $\dot{B}$ by similarly stacking $-\frac{A(\tau,s_t)}{\tau}$ and $-\frac{B(\tau)}{\tau}$. Lastly, we formalize the econometric specification of (36) by introducing a vector of errors that follow a multivariate normal distribution. The assumption of normality is typical in these models and is made by CDR, DSY, Diebold and Li (2006) and others. However, we assume that the error term is *conditionally* normal, thereby allowing its variance to depend on the unobserved process s_t. The measurement equation in the state space system is given by

$$R(t) = \dot{A}(s_t) + \dot{B}' X(t) + Z(s_t) e_t. \tag{37}$$

where e_t is a vector of orthogonal error terms and $Z(s_t)Z(s_t)'$ is a variance-covariance matrix.

The transition equation is characterized by the dynamics of the term structure factors X_t, given under the risk neutral measure in expressions (1) and (26). However, estimation takes place under the physical measure. We assume that the dynamics of X_t under this measure are given expression (13). The most notable difference between (1) and (13) is that under the physical measure the speed of mean reversion $\kappa(s_t)$ is regime dependent and that it is not bound by the restriction in (25). This is possible through a particular parameterization of the market price of interest rate risk that we give in (14) and (15).

We obtain the transition equation in discrete time by solving and discretizing the SDE in (13).

$$X_t = (I - e^{-\kappa(s_t)\Delta t})\theta(s_t) + e^{-\kappa(s_t)\Delta t} X_{t-1} + \Phi(s_t)\epsilon_t \tag{38}$$

where I is an identity matrix and

$$\Phi(s_t)\Phi(s_t)' = \int_0^{\Delta t} e^{-\kappa(s_t)s} \Sigma(s_t)\Sigma(s_t)' e^{-\kappa(s_t)'s} ds. \tag{39}$$

If κ is diagonalizable and given that the variance-covariance matrix $\Sigma(s_t)\Sigma(s_t)'$ is constant, we can solve (39) explicitly to yield

$$\Phi(s_t)\Phi(s_t)' = \Lambda\Gamma\Lambda^{-1}, \tag{40}$$

where the $(i,j)^{th}$ element of matrix Γ is given by $\frac{\sigma_{ij}}{\lambda_i\lambda_j}(1 - e^{-\lambda_i\lambda_j\Delta t})$. σ_{ij} is the $(i,j)^{th}$ element of the variance-covariance matrix $\Sigma(s_t)\Sigma(s_t)'$ and Λ is the eigenvector of $\kappa(s_t)$.

3.2 Approximate-ML algorithm for hidden Markov models

Gaussian state space models can be estimated via a Maximum Likelihood (ML) approach in conjunction with a Kalman filter. For an outline of the estimation methodology for a single regime model, please refer to DNS and CDR. Estimation by maximum likelihood is also possible in a multi-regime case but with a the standard KF.

The modification, as outlined by Kim and Nelson (1999C) is as follows. First notice that the measurement and transition equations have different coefficients for each realization of s_t. Hence, given a particular starting value, X_0, the KF algorithm will filter m state variables in the first step (i.e. X_1 will have m possible values from the recursion), where m is the number of regimes despite the fact that there is only one independent observation ($R(t)$) for each point in time t. Therefore at step 1, there are m possible values values of X_1. If this process is repeated, at step 2, there would be m^2 values of X_2 and so on. The algorithm would explode rapidly to m^t possibilities at the t^{th} step. This is not tractable, even for small t.

Kim (1994) proposed an approximation that allows the collapse of the paths in each step. This is achieved by taking expectations of the different paths for each t and the number of X_t is thereby reduced to m. The conditional probability of each regime at time t is calculated using the Hamilton filter (Hamilton (1989)). In the remainder of this section we outline this algorithm as applied to our regime switching arbitrage free NS model.

For notational purposes, we simplify notation of certain elements in expressions (37), (38) and (40) as follows,

$$\beta^i = e^{-\kappa(s_t=i)\Delta t} \tag{41}$$

$$\alpha^i = (I-\beta^i)\theta(s_t=i) \tag{42}$$

$$H^i = \Phi(s_t=i)\Phi(s_t=i)' \tag{43}$$

$$W^i = Z(s_t=i)Z(s_t=i)' \tag{44}$$

$$A^i = \dot{A}(s_t=i) \tag{45}$$

Additionally, Π_t^{ij} is the transition probability from regime i to regime j under the physical measure[5], n is the number of factors and q is the number of available yields. The steps of the approximate KF are as follows.

$$\text{Predicted state } (m\times m\times n)\text{: } X^{ij}_{t|t-1} = \alpha^j + \beta^j X^i_{t-1}$$

$$\text{Predicted covariance } (m\times m\times n\times n)\text{: } \Sigma^{ij}_{t|t-1} = \beta^j \Sigma^i_{(t-1)} \beta^{j'} + H^j$$

$$\text{Innovation residual } (m\times m\times n)\text{: } \eta^{ij}_t = R(t) - A^j - BX^{ij}_{t|t-1}$$

[5]We note that in a general setting, the transition probability may be time-dependent. A simple way to incorporate heterogeneity is to assume that each entry in the transition matrix depends on some independent variables though a logit transformation.

$$\text{Innovation covariance } (m \times m \times q \times q)\text{: } F_t^{ij} = B\Sigma_{t|t-1}^{ij}B' + W^j$$

$$\text{Kalman gain } (m \times m \times q \times q)\text{: } KG_t^{ij} = \Sigma_{t|t-1}^{ij}B(F_t^{ij})^{-1}$$

$$\text{Updated state } (m \times m \times n)\text{: } X_{t|t}^{ij} = X_{t|t-1}^{ij} + KG_t^{ij}\eta_t^{ij}$$

$$\text{Updated covariance } (m \times m \times n \times n)\text{: } \Sigma_{t|t}^{ij} = (1 - KG_t^{ij}B')\Sigma_{t|t-1}^{ij}$$

Each superscript ij refers to a matrix or a vector that corresponds to a transition from regime i to regime j, for each t. Ignoring this notation, the algorithm above is identical to the standard KF.

At each t, as discussed above, the updated states and covariances are collapsed into m vectors or matrices respectively. This is done by weighing each X^{ij} and Σ^{ij} by the probability of transition to a state i from state j via the following approximations

$$X_t^j = \left(\sum_{j=1}^{m} P_t^{ij} X_{t|t}^{ij}\right) / P_t^j,$$

$$\Sigma_t^j = \left(\sum_{j=1}^{m} P_t^{ij}(\Sigma_{t|t}^{ij} + (X_t^j - X_{t|t}^{ij})(X_t^j - X_{t|t}^{ij})')\right) / P_t^j,$$

where

$$P_t^{ij} = \frac{l_t^{ij}}{\sum_{i=1}^{m}\sum_{j=m}^{m} l_t^{ij}},$$

$$l_t^{ij} = N\left(\eta_t^{ij}, F_t^{ij}\right)\Pi_t^{ij}P_{t-1}^i,$$

and

$$P_t^j = \sum_{i=1}^{m} P_t^{ij}.$$

l_t^{ij} can be understood as the likelihood that a transition from i to j took place, given that the data at time t has been observed. P_t^{ij} is the transition probability implied by the likelihood values and P_{t-1}^j is the unconditional probability of state j, which is calculated as the sum of all transitions to that state. Finally, the log-likelihood is calculated as

$$L = \sum_{t=1}^{T}\left[\sum_{i=1}^{m}\sum_{j=1}^{m} N\left(\eta_t^{ij}, F_t^{ij}\right)\Pi_t^{ij}P_{t-1}^i\right]. \tag{46}$$

4 Data

We estimate the Nelson and Siegel regime switching model on US treasury strip curve from Bloomberg. The data consists of strip yields for 3 and 6 month and 1, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25 and 30 year maturities from April 1991 to August 2010 at monthly frequency. This represents 233 time series and 15 cross-sectional observations. The period under consideration spans three recessions and two expansions(according to NBER) and contains diverse term structure outcomes. Figure 2 plots the data and Table 4 provides information on the central moments and persistence of the individual series.

Table 1: Data Summary: This table presents the descriptive statistics and autocorrelation of each of the 15 yields used in this paper. The yields are monthly from April 1991 to August 2010, for a total of 233 time series obsrvations per series.

	Central moments, median, min and max							Autocorrelations	
	Mean	Median	Stdev	Skew	Kurt	Min	Max	Lag 1	Lag 2
3-month	0.0355	0.0403	0.0189	-0.4523	-1.2201	0.0010	0.0625	0.9904	0.9812
6-month	0.0368	0.0420	0.0192	-0.4550	-1.2012	0.0018	0.0654	0.9921	0.9834
1-year	0.0386	0.0440	0.0191	-0.4445	-1.0947	0.0027	0.0728	0.9903	0.9783
2-year	0.0418	0.0453	0.0183	-0.4185	-0.9529	0.0052	0.0769	0.9870	0.9706
3-year	0.0443	0.0463	0.0172	-0.3706	-0.8839	0.0077	0.0778	0.9844	0.9650
4-year	0.0466	0.0481	0.0163	-0.2856	-0.8519	0.0108	0.0785	0.9814	0.9600
5-year	0.0483	0.0491	0.0155	-0.1815	-0.8311	0.0139	0.0799	0.9796	0.9548
7-year	0.0515	0.0511	0.0140	0.0561	-0.8283	0.0197	0.0833	0.9754	0.9480
8-year	0.0527	0.0517	0.0136	0.1301	-0.8121	0.0220	0.0843	0.9750	0.9470
9-year	0.0537	0.0524	0.0134	0.1848	-0.8166	0.0246	0.0851	0.9745	0.9461
10-year	0.0550	0.0533	0.0130	0.2299	-0.7562	0.0261	0.0857	0.9732	0.9429
15-year	0.0580	0.0569	0.0123	0.3312	-0.7094	0.0315	0.0873	0.9725	0.9432
20-year	0.0588	0.0573	0.0121	0.3353	-0.7329	0.0313	0.0870	0.9755	0.9489
25-year	0.0586	0.0573	0.0121	0.3249	-0.7845	0.0297	0.0860	0.9756	0.9497
30-year	0.0578	0.0564	0.0123	0.2956	-0.7993	0.0266	0.0848	0.9751	0.9487

5 DNS and AFNS models

In this section we present the results of two models that will be used as benchmarks in the analysis of our regime switching model. In particular, we estimate the three-factor dynamic Nelson-Siegel model of Diebold and Li (2006) and the arbitrage-free Nelson and Siegel model developed by CDR. There are other models that could have been used for comparison, but the DNS and AFNS models are well documented in the literature and relatively popular in practice. Additionally, they are particularly comparable to our model as they are both three-factor Nelson-Siegel type models and can be estimated via the KF. In what follows we briefly outline the specification of the models and the estimation methodology and results.

The measurement equation for the DNS model is given by

$$R(t) = BX(t) + Ze_t \tag{47}$$

where B is a matrix of Nelson and Siegel factor loadings, that consists of stacked $\frac{-B(\tau)}{\tau}$ vectors in expression (27). Z is a diagonal matrix of standard deviations and e_t is a vector of orthogonalized error terms. The transition equation is given by

$$X_t = (1-K)\mu + KX_{t-1} + \Phi\epsilon_t \tag{48}$$

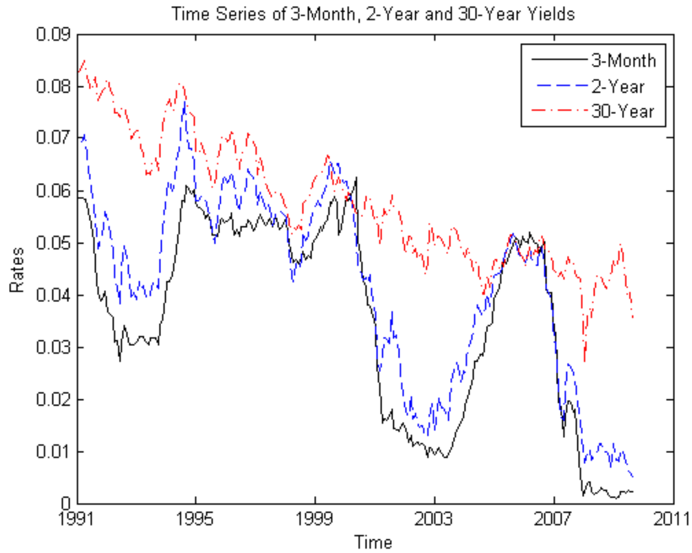

Figure 2: Yield Plot: This figure plots the 3-month, 2-year and 20-year strip yields from April 1991 to August 2010.

where μ is the long run mean of the factors X_t and $\Phi\Phi'$ is the covariance matrix. As before ϵ_t are orthogonal errors terms which are assumed to be uncorrelated with e_t. The matrix $K = \Upsilon E \Upsilon^{-1}$ is assumed to be diagonalizable, where Υ are the eigenvectors with diagonal elements normalized to 1, and where E is a diagonal matrix of eigenvalues. This separation was necessary so that the estimated model dynamics could be restricted to ensure stationary behavior of the factors. The estimated coefficients of the model are presented in Table 2.

Table 2: DNS Model Results: This table presents the results from the dynamic nelson and Siegel model. The λ for this model was estimated to be 0.0509(0.0007). The first two columns of he table give the eigenvalues and eigenvectors of the matrix K respectively. The next column presents the long term mean of the term structure factors and the last column is the square-root of the covariance matrix $\Phi\Phi'$.

E	Υ				Mean	Φ Matrix			
$Diag(E)$		$\Upsilon_{.,1}$	$\Upsilon_{.,2}$	$\Upsilon_{.,3}$	μ		$\Phi_{.,1}$	$\Phi_{.,2}$	$\Phi_{.,3}$
0.9668	$\Upsilon_{1,.}$	1.0000	-0.0609	1.6197	0.0036	$\Phi_{1,.}$	0.0891		
(0.0293)			(0.1387)	(0.6991)	(0.0006)		(0.0025)		
0.8888	$\Upsilon_{2,.}$	-4.3653	1.0000	-0.7009	-0.0010	$\Phi_{2,.}$	-0.0387	0.0091	
(0.0501)		(6.9670)		(0.8698)	(0.0006)		(0.0038)	(0.0044)	
0.9999	$\Upsilon_{3,.}$	-0.4925	-0.6042	1.0000	0.0063	$\Phi_{3,.}$	0.0026	-0.0034	0.0023
(0.0067)		(1.1185)	(0.4823)		(0.0004)		(0.0001)	(0.0003)	(0.0001)

For comparison purposes, the matrix K in this case is given by

$$K = \begin{pmatrix} 1.0023 & 0.0079 & 0.0017 \\ -0.0821 & 0.9379 & 0.0896 \\ 0.0608 & 0.0197 & 0.9153 \end{pmatrix}$$

and

$$\Phi\Phi' = \begin{pmatrix} 0.0094 & -0.0038 & 0.0004 \\ -0.0038 & 0.0016 & -0.0001 \\ 0.0004 & -0.0001 & 0.0001 \end{pmatrix}.$$

The AFNS model is an arbitrage-free extension of the DNS model above. In practice, this implies that the measurement equation includes an intercept term which is dependent on the tenure of the corresponding yield, λ, the decay parameter and the covariance of the interest rate factors. The measurement and transition equations are given by

$$R(t) = A + BX(t) + Ze_t \tag{49}$$

$$X_t = (I - e^{-\kappa\Delta t})\mu + e^{-\kappa\Delta t}X_{t-1} + \Phi\epsilon_t \tag{50}$$

where A is the intercept term,[6] B, Z and e_t are as in (47), κ is a diagonalizable matrix and $\Delta_t = 1/12$ since yields are observed at monthly frequency. The covariance matrix is given by

$$\Phi\Phi' = \Lambda\chi\Lambda^{-1}, \tag{51}$$

where Λ is the eigenvector of κ. χ is given by $\chi_{ij} = \frac{\sigma_{ij}}{\lambda_i\lambda_j}(1 - e^{-\lambda_i\lambda_j\Delta t})$, where σ_{ij} is the $(i,j)^{th}$ element of the instantaneous factor covariance matrix $\Sigma\Sigma'$. The estimates from this model are given in table 5.

Table 3: AFNS Model Results: This table presents the results of the AFNS model. The first two columns respectively present the eigenvalues and eigenvectors of the matrix κ. The next column gives the long term mean of the term structure factors X_t. The rightmost column shows the estimates of the square root of the instantaneous covariance matrix of X_t. The λ for this model is 0.2243(0.0058).

		κ			Mean		Σ Matrix		
λ		$\Lambda_{.,1}$	$\Lambda_{.,2}$	$\Lambda_{.,3}$	μ		$\Sigma_{.,1}$	$\Sigma_{.,2}$	$\Sigma_{.,3}$
0.0012	$\Lambda_{1,.}$	1.0000	-0.3875	-0.1475	0.0818	$\Sigma_{1,.}$	0.0084		
(0.0143)			(0.4409)	(0.1459)	(0.0191)		(0.0002)		
0.2999	$\Lambda_{2,.}$	0.8605	1.0000	0.3503	-0.0212	$\Sigma_{2,.}$	-0.0056	0.0108	
(0.1764)		(0.9913)		(0.2399)	(0.0185)		(0.0008)	(0.0005)	
2.2675	$\Lambda_{3,.}$	0.1061	-0.5607	1.0000	-0.0020	$\Sigma_{3,.}$	0.0008	-0.0002	0.0198
(0.6005)		(0.4557)	(0.2519)		(0.0097)		(0.0018)	(0.0014)	(0.0015)

The matrices $e^{-\kappa\Delta t}$ and $\Phi\Phi'$ are given by,

$$e^{-\kappa\Delta t} = \begin{pmatrix} 0.9857 & 0.0143 & 0.0183 \\ 0.0348 & 0.9648 & -0.0428 \\ 0.0561 & -0.0471 & 0.8526 \end{pmatrix}$$

and

$$\Phi\Phi' = \begin{pmatrix} 4.11 \times 10^{-06} & -2.42 \times 10^{-06} & 4.66 \times 10^{-07} \\ -2.42 \times 10^{-06} & 1.17 \times 10^{-05} & -3.99 \times 10^{-06} \\ 4.66 \times 10^{-07} & -3.99 \times 10^{-06} & 3.75 \times 10^{-05} \end{pmatrix}.$$

[6]The explicit form of A is cumbersome and will not be restated here. We refer the interested reader to CDR (2010).

6 Regime switching model results

In this section, we present the estimation results of the Nelson and Siegel arbitrage free regime switching model and contrast them to the single-regime alternatives presented in the previous section. Before beginning to estimate the model, however, we make several simplifying assumptions. First, we assume that two regimes are sufficient to capture the nonstationary behavior of interest rates. This choice is motivated through theoretical and practical arguments. In particular, the understanding of regimes is often tied to expansions and recessions. Monetary policy responds to macroeconomic phenomena and it is therefore natural to expect expansionary and recessionary monetary regimes. This choice is also motivated by a large number of previous studies, which nearly exclusively assume the existence of only two regimes. Finally, the densities of interest rates and factors in figure 1 are bi-modal. This indicates that a mixture of two normal densities should be sufficient to model interest rate behavior. From a practical perspective, increasing the number of regimes leads to a significant increase in the number of parameters. In term structure models, which typically rely on a large number of parameters (including the latent factors), more than two regimes are untenable.

Second, we assume that the latent factors are independent. This is a natural assumption, established in the literature (Nelson and Siegel (1987); CDR). Furthermore, we have no reason to believe *a priori* that the level, slope and curvature should be correlated. In fact, the factors are often linked to principal components, which are independent by construction. The assumption of independence improves convergence and results in a highly tractable model that can be estimated in practice relatively easily. As a result of independence, the coefficient of the lagged factors $e^{-\kappa(s_t)\Delta t}$ and the volatility matrix $\Phi(s_t)$ in expression (38) are diagonal.

Third, we follow Diebold and Li (2006) and CDR and assume that the measurement covariance matrix is diagonal. This is a somewhat strong assumption because the underlying economic rationale is that three factors are sufficient to capture the correlation structure of yields. Additionally, we assume that the measurement covariance matrix $Z(s_t)Z(s_t)'$, given in expression (37) is regime dependent.[7] This assumption converts expression (28) in proposition 2 to the form

$$A_{n+1}^{j} = -(\tilde{\kappa}\tilde{\theta}(s_t))'B - \frac{1}{2}\sum_{i=1}^{3}[B_i\Sigma_{ii}(s_t)]^2 - log\left(\Sigma(s_t)_{k=0}^{S}\pi^{jk}e^{-A_n^k}\right) \tag{52}$$

which simplifies the estimation by reducing the number of parameters to be estimated.

Finally, we assume that the transition probability under the measure P is also independent of the factors X_t and time.[8] This assumption increases the feasibility of the model and reduces

[7] In Hypothesis 6 of section 6.3 we investigate the validity of this assumption.

[8] A time-dependent market price of risk could have been obtained by assuming a time-dependent transition probability matrix. Each element of the transition probability could be a function of several time varying endogenous or exogenous variables. The factors are ideal candidates for endogenous variables and interest rates (used in DSY) as exogenous variables.

the number of factors to be estimated.

The model estimates conditional on the states are presented in Table 4. The top panel shows the parameters in Regime 1 and the lower panel for Regime 2. The λ for this model is 0.6032 (0.0166) and is highly significant as per the standard error in the parentheses. The negative log likelihood of the model is -19267.2, and in this metric, as in all the others, it dominates the single regime models. The estimated homogeneous transition probabilities are given by expression (53). The superscripts P and Q denote the physical and risk-neutral measures respectively.

Table 4: Arbitrage-free Nelson and Siegel Hidden Markov Model Results: The parameter values for the model under the measure P are given for Regime 1 and 2. κ is the mean reversion parameter. θ is the mean. α^Q is as in equation (42) under the risk-neutral measure Q. Σ is the measurement covariance matrix.

				Regime 1					
	$\kappa_{\cdot,1}$	$\kappa_{\cdot,2}$	$\kappa_{\cdot,3}$	θ	α^Q		$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$
$\kappa_{1,\cdot}$	0.0012			0.0848	0	$\Sigma_{1,\cdot}$	0.0061		
	(0.0480)			(0.0026)			(0.0003)		
$\kappa_{2,\cdot}$		0.2515		-0.0243	-0.0018	$\Sigma_{2,\cdot}$		0.0098	
		(0.1846)		(0.0052)	(0.0004)			(0.0006)	
$\kappa_{3,\cdot}$			1.4119	0.0042	0.0094	$\Sigma_{3,\cdot}$			0.0226
			(0.5338)	(0.0045)	(0.0007)				(0.0015)
				Regime 2					
	$\kappa_{\cdot,1}$	$\kappa_{\cdot,2}$	$\kappa_{\cdot,3}$	θ	α^Q		$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$
$\kappa_{1,\cdot}$	0.1448			0.0397	0	$\Sigma_{1,\cdot}$	0.0060		
	(0.1646)			(0.0278)			(0.0004)		
$\kappa_{2,\cdot}$		1.1807		-0.0499	-0.0012	$\Sigma_{2,\cdot}$		0.0117	
		(0.4321)		(0.0059)	(0.0004)			(0.0011)	
$\kappa_{3,\cdot}$			3.0173	-0.0386	-0.0014	$\Sigma_{3,\cdot}$			0.0410
			(1.1811)	(0.0063)	(0.0009)				(0.0055)

$$\Pi^P = \begin{pmatrix} 0.9679 & 0.0321 \\ (0.0278) & \\ 0.0763 & 0.9237 \\ & (0.0115) \end{pmatrix} \qquad \Pi^Q = \begin{pmatrix} 0.7504 & 0.2496 \\ (0.0409) & \\ 0.0010 & 0.9990 \\ & (0.0202) \end{pmatrix} \tag{53}$$

In order to compare the model to the DNS and AFNS models, we present the matrices β^i and H^i (as specified in expressions (41) and (42)). These are comparable to K, $e^{-\kappa\Delta t}$ and $\Phi\Phi'$ matrices in Section 5. We note that stationarity was imposed on each regime in the hidden Markov model and for both single-regime models.

$$\beta^1 = \begin{pmatrix} 0.9999 & 0 & 0 \\ 0 & 0.9792 & 0 \\ 0 & 0 & 0.8890 \end{pmatrix} \qquad \beta^2 = \begin{pmatrix} 0.9880 & 0 & 0 \\ 0 & 0.9063 & 0 \\ 0 & 0 & 0.7777 \end{pmatrix}$$

$$H^1 = \begin{pmatrix} 3.05 \times 10^{-06} & 0 & 0 \\ 0 & 7.81 \times 10^{-06} & 0 \\ 0 & 0 & 3.79 \times 10^{-06} \end{pmatrix} \qquad H^2 = \begin{pmatrix} 2.97 \times 10^{-06} & 0 & 0 \\ 0 & 1.04 \times 10^{-05} & 0 \\ 0 & 0 & 1.10 \times 10^{-04} \end{pmatrix}$$

Regime 1 is characterized by higher long run factor means. The first element of the θ vector corresponds to the level factor, which in a Nelson-Siegel setting, captures the behavior of the longest maturity yield. The second element is the mean of the slope factor. We have shown in section 2.1 that in DNS models $X_1 + X_2$ (the sum of the level and slope factors) approaches the instantaneous short rate. This relationship is intuitive, since it corresponds to the level of the short rate and a duration adjustment. The long term instantaneous short rate in Regime 1 is 6.05% and -1.02% in Regime 2. The implied negative rate in Regime 2 is a consequence of very low short rates in 2010. Both short rates and long rates are higher in Regime 1 than in Regime 2. The long run slope is higher in regime 2, which is consistent with what has been observed historically. This narrative is collaborated by figure 3, which plots the average model-implied term structure of interest rates, per regime. These estimates are also contrasted to average model-implied term structure in single regime models. The term structure is higher in Regime 1 than in Regime 2. We also observe that the term structure of the DNS and AFNS models are below term structure in Regime 1 and above Regime 2 for maturities up to 25 years.

Factor dynamics in Regime 1 appear to be more persistent (as can be seen from β^1, β^2 matrices), although later we show that the difference in persistence are not significant. The same can be said about the persistence of the states, under both measures. The transition probability Π^P indicates that the probability of transitioning to Regime 2 from Regime 1 is only 3.21%. The probability of a return to Regime 1 is more than twice that (7.63%). This relationship is consistent with the well-documented asymmetry of the business cycle. Expansions should be more pervasive then contractions and this is reflected in the monetary policy response. Our model captures this regularity well. The covariances in the transition equation in Regime 1 are less than or equal to the covariances in Regime 2 as can by seen from H^1 and H^2.

This analysis suggests that Regime 1 is a contractionary monetary policy regime, characterized by high interest rates and low volatility. Regime 2 is an expansionary monetary policy regime, that captures periods of low rates and relatively high volatility. This conclusion is consistent with the fact that Regime 1 is more persistent and that Regime 2 exhibits higher slope of the term structure. The interpretation of the regimes is validated by figure 4 which plots the states over time. Regime 2 transpired between April 1991 and January 1993, from April 2001 to October 2003 and from November 2007 to May 2010. These intervals are associated with loose monetary policy related to combating three recessions in 1991, 2001 and 2008-2009. The 1991 recession, as dated by NBER, ended in March 1991, one month before the start of our sample. The NBER dates for the 2001 recession are from March to November 2001. The most recent slowdown is dated from December 2007 to June 2009. Our model accurately captures all or part of both recessions. Our estimates indicate that loose monetary policy persists long after the recessions it intended to address were over. In other words, monetary policy impulse persisted for many months after recovery began. This may indicate that monetary policy has been highly effective in combating recessions, but that interest rates were kept too low for too long over the last three business cycles or that persistent monetary easing is necessary to sustain economic recovery..

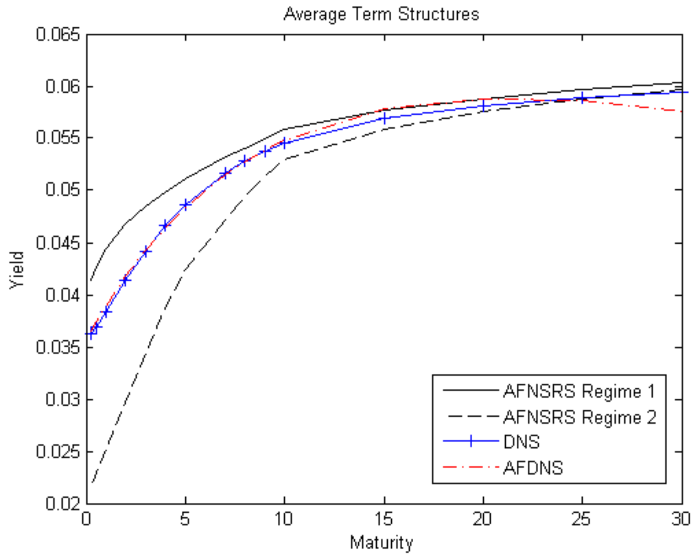

Figure 3: Average Term Structure: The average term structure for the was computed as the mean of the filtered values for each of the 15 yields. For the regime switching model, this was done for each regime.

These results and interpretation are in line with related literature. Bansal and Zhou (2002) and DSY both identify a high and a low interest rate regime, although in both cases, the regime selection seems to be driven by the monetary experiment in the 1980's. Our sample excludes this period and this allows us to focus on recent monetary phenomena. The interpretation of the regimes is also consistent although in Bansal and Zhou (2002) and DSY, the regimes are associated with economic expansions and contractions. We have argued above that in our case it is productive to distinguish between macroeconomic and monetary regimes. Furthermore, the link to monetary policy rather than macroeconomic activity is more natural in this framework, since there is no direct process by which GDP enters the model.

6.1 Market prices of risk

As in any equilibrium model, there exist market prices of risk for each source of randomness. Although we have developed our model under the risk neutral measure, it is still possible, under certain assumptions (outlined in Section 1.2), to infer the market prices associated with the interest rate factors and the state process. We begin by analyzing the market price of regime switching risk.

We note that the assumptions in Section 1.2 imply that the state transitions under the risk neutral measure Q are independent of the term structure factors. Furthermore, as discussed in the beginning of this section, the same is assumed for the transitions under the measure P. Consequently, the market price of regime switching risk is constant and is given as the

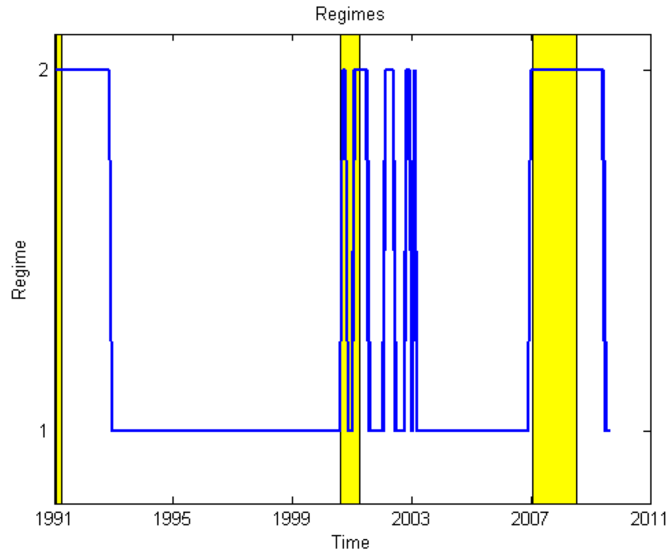

Figure 4: Regimes: This figure shows the evolution of the regimes across time from 1991 to 2010. The economy is said to be in a particular regime if the probability of that regime (P_t^j outlined in section 3.2) is greater than 0.5. The shaded regions of the figure represent the NBER recessions over the sample period.

log ratio of the transition probability under measure P to the transition probability under measure Q,[9]

$$log(\frac{\Pi^P}{\Pi^Q}) = \begin{pmatrix} -0.2546 & 2.0526 \\ -4.3348 & 0.0784 \end{pmatrix},$$

where the $(i,j)^{th}$ element represents the market price of risk of switching from regime i to regime j.

Consistent with intuition, the price of regime switching risk is greatest when transitioning from Regime 1 to Regime 2. In other words, the cost of hedging in a contractionary monetary policy regime is higher than the cost of hedging in an expansionary regime. This implies that investors place a large premium on hedging against economic downturn, when the economy is in a boom state. The second highest price of regime switching is associated with the continuation of Regime 2. This is consistent with the hedge against an economy remaining in a recession, given that it is in one already. This result is also intuitive, since when the marginal utility of consumption is high, investors are less willing to allocate resources towards hedging future consumption shocks.

[9]This is true because in discrete time $\tilde{h}(z,X)$ and $h(z,X)$ in expressions (6) and (20) are the transition probabilities Π^P and Π^Q. This, in combination with the assumption of independence from X_t and with expression (16) yields $log(\frac{\Pi^P}{\Pi^Q}) = -\psi_0$. The resulting expression is consistent with the parameterization of the market price of regime switching risk in DSY.

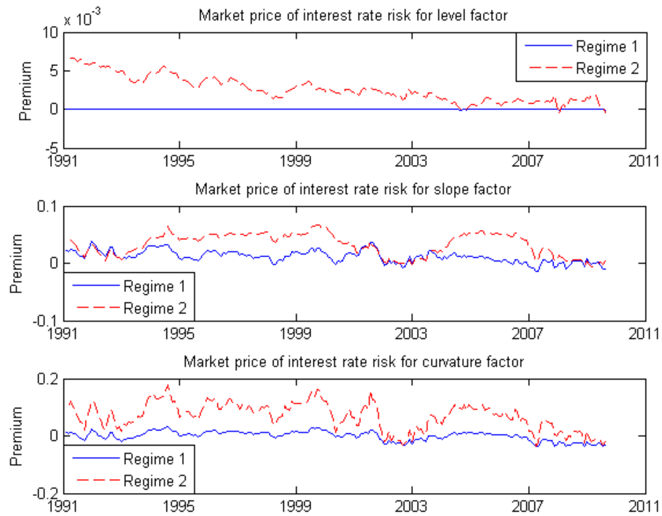

Figure 5: Market Prices of Interest Rate Risk: This figure shows the evolution of the market prices of risk, defined in equation (11). Sub-figure i corresponds to λ_i. λ_1 is the market price of risk of the level, λ_2 is the market price of risk of the slope and λ_3 is the market price of risk of the curvature factor.

The factor risk can be seen through the evolution $\vec{\lambda}$ given in equation (11) over time. The estimated market prices of factor risk, based on expressions (14) and (15), are given in figure 5. The market prices are higher in Regime 2 for all factors. The excess return in a monetary expansion regime is low and the risk (measured by the variance of the factors) is high. The difference between the prices in Regime 2 and Regime 1 can be primarily attributed to the mean reversion coefficient under the real measure ($\kappa(s_t)$).

Economically, a high market price of interest rate risk during recessions is meaningful. The higher risk aversion of investors during recessions leads to bond purchases which drive down bond yields even further than they would be under monetary easing. This increases the price of bonds and causes significant interest rate risk for bond holders.

6.2 Comparison to single regime models

In this section we compare the regime-switching model to the DNS and AFNS models introduced in Section 5. We compare the three models in terms of in-sample fit, but we also contrast the factors loadings, yield adjustment vectors and the factor dynamics.

The overall fit of the regime-switching model is superior, in-sample, to the DNS and AFNS models. Table 5 presents the likelihoods of the estimated models together with the Akaike information criterion (AIC), the Bayesian information criterion (BIC) statistics and the number of parameters in each model. Regime switching models provide a superior description of the dynamics of the term structure even after controlling for a larger number of parameters. We note that a large number of the estimated parameters (15 of the 34 parameters in the single-regime models and 30 of the 57 parameters in the regime-switching model) pertain to the measurement covariance matrix. Note that this superior performance is in spite of the assumption that the factors of the regime switching model are uncorrelated. We make no such assumption about the DNS and AFNS models. This strengthens the argument in favor of the viewing the level, slope and curvature factors as orthogonal.

Table 5: In-sample fit comparison: The negative log-likelihood, number of parameters, AIC and BIC values are given for the regime-switching model and the DNS and AFNS models.

Model	Log Likelihood	Parms	AIC	BIC
DNS	-18625	34	-37182	-37065
AFNS	-18876	34	-37685	-37568
AFNSRS	-19267	57	-38420	-38223

The improvement in the likelihood is also illustrated by the superior cross-sectional fit of the model. Figure 6, plots the cross-sectional fit of the DNS, AFNS and the regime-switching model at four dates.[10] There is a trade-off between fitting short term and long term yields, as indicated by the fit of the DNS model relative to the AFNS model. The DNS model fits the short end of the yield curve well in this sample, but does not accurately capture the shape at the long end of the yield curve. Conversely, the AFDNS model tracks the long yield well, but fails to capture short term yields adequately. This regularity is not specific to the four dates chosen and is a feature of this sample.

The regime-switching model also appears to capture the short term structure well, but is out-performed by the AFNS model at longer horizons. On average however, it does better than the AFNS model with short yields and better than the DNS model with long yields. The model also seems to capture short yields as well as the DNS model. The similarity with the DNS model can also be seen from the graph of the factor loadings (figure 7). The similarity in the loadings is also reflected in λ.

Given the concurrence in the loadings, the difference in the cross-sectional fit of the regime-switching and DNS models is due to the yield adjustment vector $\dot{A}(s_t)$. This term affects the level, slope and curvature of the term structure, conditional on a particular regime. Figure 8, plots the yield adjustment terms for the AFNS model (DNS model does not have an adjustment term) and for each regime of the hidden Markov model. The restrictions imposed in the AFNS model force the adjustment to be a monotonic function of τ.[11] In

[10] The fit for the regime-switching model is based on the probability weighted expectation of the term structure on those dates, where the probability is given by the unconditional probability that the system is in regime i. The conditional expectation of the term structure corresponds to the measurement equation in expression (37), where the factors X_t are taken from the updating step of the KF described in Section 3.

[11] CDR are able to achieve a closed form solution for the yield adjustment term in a single-regime setting by assuming that $\tilde{\theta}$ is zero.

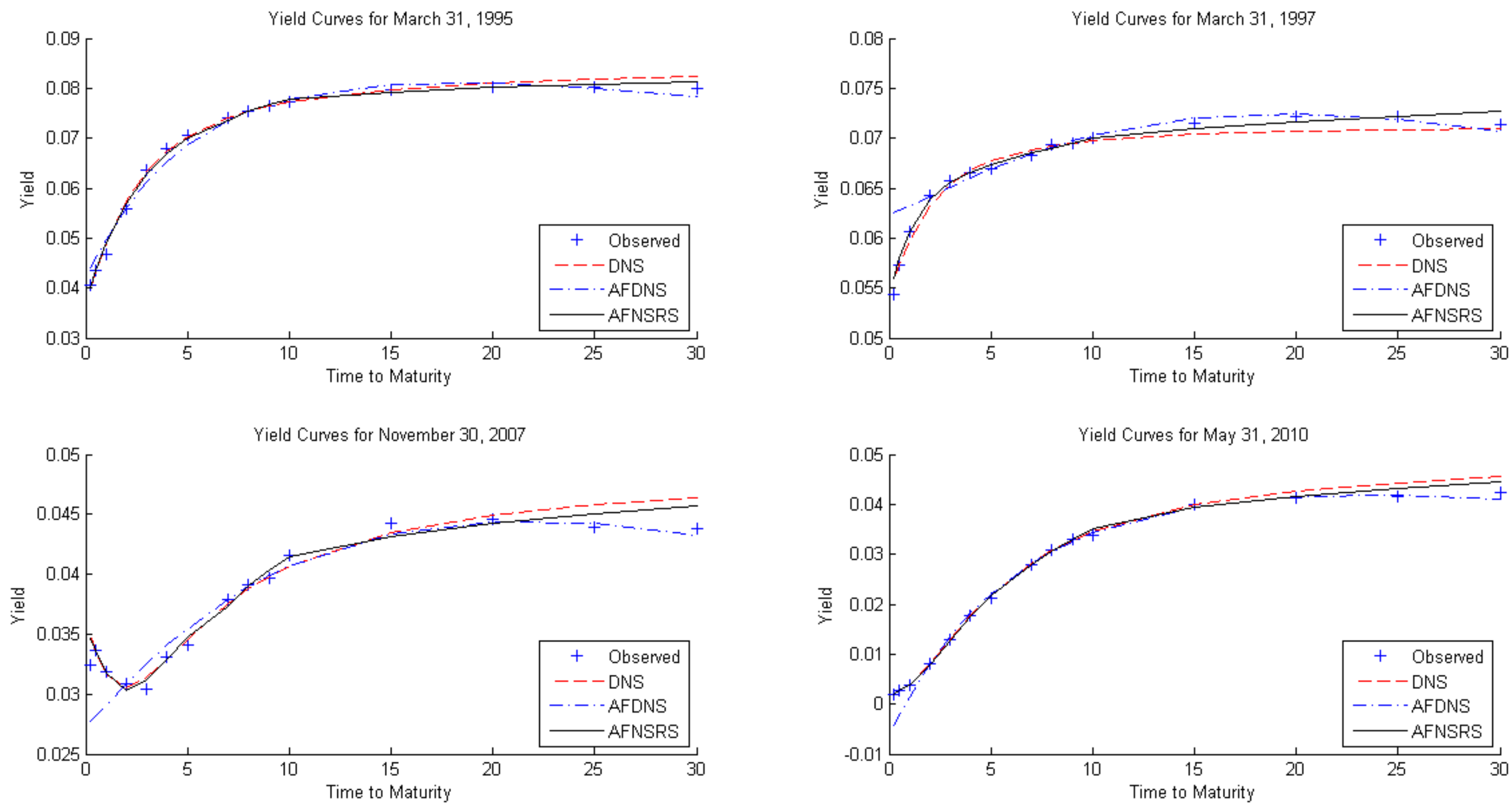

Figure 6: Cross-sectional Fit on 4 Dates: This figure shows the cross-sectional fit of all the estimated models. This is based on the updated factors from the KF on a particular day.

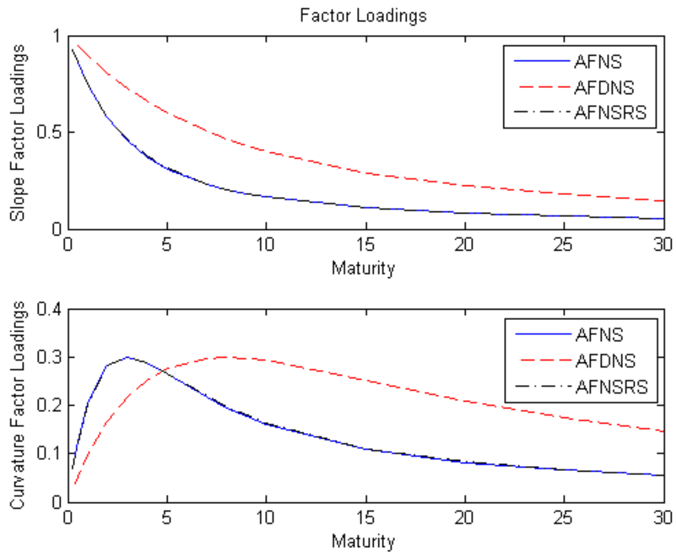

Figure 7: Slope and Curvature Factor Loadings: This figure shows the slope (upper panel) and the curvature (lower panel) loadings for the three models. The DNS and the regime swithcing model are nearly indistinguishable.

contrast,the yield adjustment term in the regime switching model permits several inflection points granting it greater flexibility. This is possible due to the recursive nature of the solution for $A(s_t)$ in (28).

The yield adjustment term for instantaneous yields is zero in both regimes. As maturity increases, the regime-dependent adjustments diverge, but remain related through the logarithmic term in $A(s_t)$. The yield adjustment term in Regime 1 adjusts the short yields downwards flattening the yield curve. It hits the minimum at the five-year maturity and then increases thereafter. In contrast, the adjustment in Regime 2 steepens the term structure at the short end and reaches a maximum at the 10-year maturity. This leads to increased curvature of the yield curve in Regime 2. It is important to note, however, that the yield adjustment term is small relative to the curve implied by the factor loadings and the term structure factors. As a result, the yield adjustment has a limited effect on the cross-sectional shape of the term structure. However, this small adjustment guarantees that the model is arbitrage-free. The fact that the adjustments in the AFNS and regime-switching models are relatively small suggests that the DNS is almost (but not entirely) arbitrage free.

The filtered term structure factors corroborate the fact that the regime-switching model shares certain characteristics with the DNS model. The regime-switching factors are closer to the DNS factors, but are highly correlated with the AFNS model as well. The correlation between the factors in Regime 1 and the DNS factors is particularly high. The AFNS model curvature factor does not conform with the curvature factors of the other two models. Curvature is typically the hardest factor to estimate and appears to be effected by the relatively high yield adjustment term in the AFNS model.

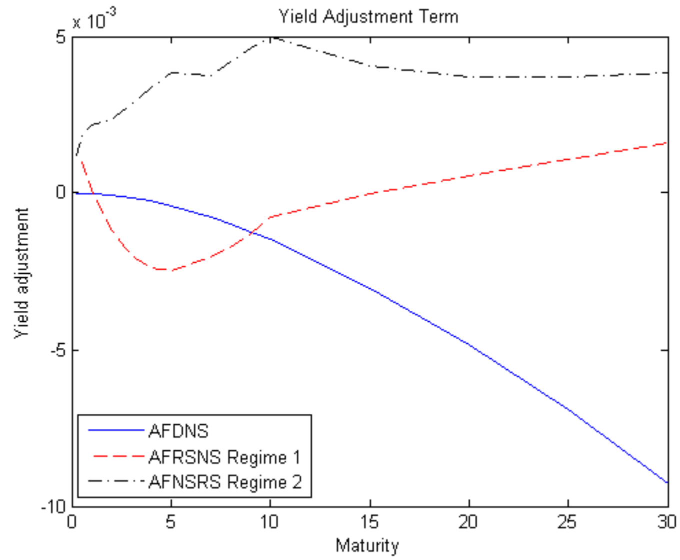

Figure 8: Yield Curve Adjustment Term: The yield adjustment term is given by $A(s_t)$ in expression (37), in the case of the regime switching model. The adjustment is different for each regime, as shown in the figure. The DNS model does not have a yield adjustment term and the close form solution for the adjustment in the AFNS model is given in CDR.

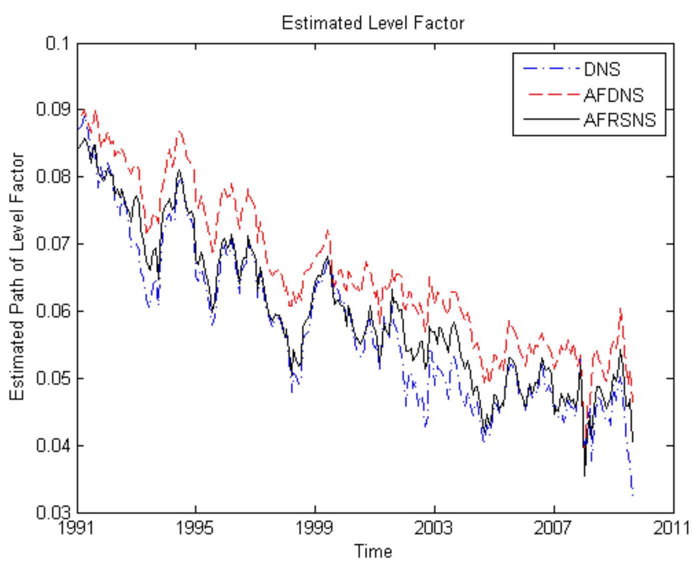

Figure 9: Estimated Level Factors: This figure plots the estimated level factor (X_1) for the estimated models.

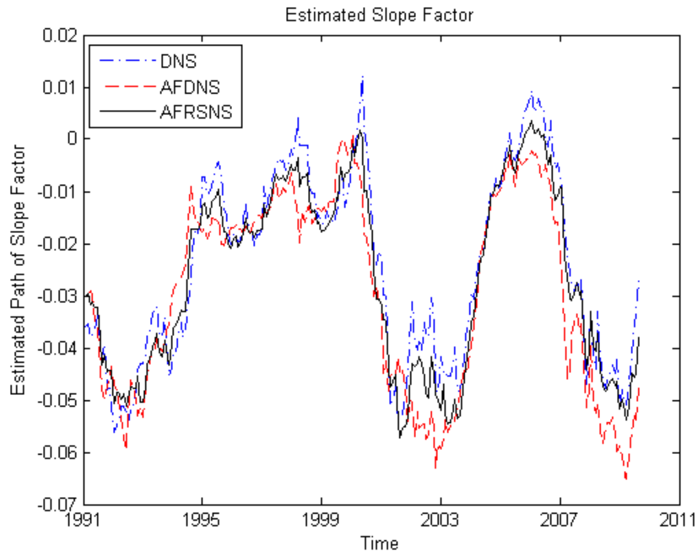

Figure 10: Estimated Slope Factors: This figure plots the estimated slope factor (X_2) for the estimated models.

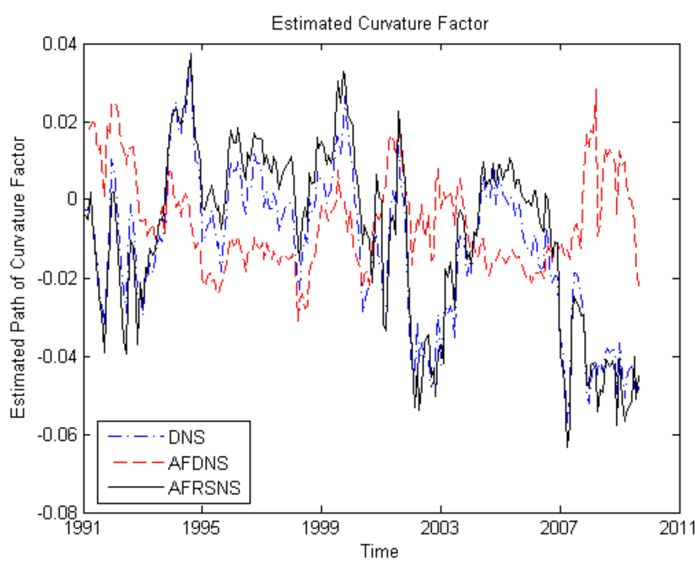

Figure 11: Estimated Curvature Factors: This figure plots the estimated curvature factor (X_3) for the estimated models.

The similarity in the factors is not surprising. The KF updates the factors based on data available as of time t. It is therefore natural to expect that the model performs well in-sample. All models rely on three factors and in each case the loadings are assumed to be regime independent. Additionally, the yield adjustment terms are relatively small. The gains in likelihood for the regime-switching models therefore originate from the specification of the transition equation or from the regime dependence of the measurement covariance matrix. These will be explored next.

6.3 Robustness tests

In the previous discussion and Table 4, we have shown that regimes exist in our sample and that many term structure parameters are regime dependent. However, some questions remain. Are the differences between parameters in Regime 1 and Regime 2 statistically significant? Are the regimes driven by the differences in means, variances or the mean reversion parameter? Is the risk of regime switching priced? Is the measurement covariance matrix regime independent? To answer these questions we formulate several hypotheses and test them using likelihood ratio tests.

Hypothesis 1(a): $\theta_{\text{Regime 1}} = \theta_{\text{Regime 2}}$
Hypothesis 1(b): $\tilde{\theta}_{\text{Regime 1}} = \tilde{\theta}_{\text{regime 2}}$
Hypothesis 2: $\Sigma\Sigma'_{\text{Regime 1}} = \Sigma\Sigma'_{\text{Regime 2}}$
Hypothesis 3: $\kappa_{\text{Regime 1}} = \kappa_{\text{Regime 2}}$
Hypothesis 4: $\tilde{\theta}_{\text{Regime 1}} = \tilde{\theta}_{\text{regime 2}} = \vec{0}$
Hypothesis 5: $\Pi^P = \Pi^Q$
Hypothesis 6: $ZZ'_{\text{Regime 1}} = ZZ'_{\text{Regime 2}}$

Table 6 summarizes the results of the tests. The first column identifies the hypothesis that was tested, followed by the likelihood of the restricted model, the total number of parameters estimated and the difference between the degrees of freedom relative to the unrestricted model. The last column presents the probability of acceptance under the null. The first row corresponds to the unrestricted model.

The findings from the model are consistent with the observation of business cycles. The tests show that we can reject Hypotheses 1(a) and 1(b); that the long term means, both under both measures are significantly different from each other. This is to be expected and evidence of monetary policy in the expansionary regime under both measures. We also reject Hypothesis 2; the transition covariance matrices represented by $\Sigma\Sigma'$ in expression (39) is regime dependent partly explaining the observed heteroskedasticity of the factors in the DNS and AFNS models. We cannot reject Hypothesis 3 at the 5% significance level. This indicates that the mean reversion parameters under the physical measure are not statistically different from each other. Under the risk free measure, $\tilde{\kappa}$ is assumed to be regime invariant. This conclusion is surprising, given the large difference in persistence implied by the parameters in table 4.

Table 6: Hypothesis Test Results: This table summarizes the results of the likelihood ratio tests that were used to test Hypotheses 1 - 6. The first column of the table givens the number and a description of the hypothesis. The next column shows the negative log likelihood value of the model under the specific hypothesis. The remaining columns present respectively the total number of parameters in the model, the number of restrictions and the probability that the null hypothesis is true.

Regime Switching Models	Likelihood	N	Diff DF	Prob
Unrestricted	-19267	57		
Equal θ (Hypothesis 1(a))	-19259	54	3	0.0008
Equal $\tilde{\theta}$ (Hypothesis 1(b))	-19260	55	2	0.0004
Equal variances (Hypothesis 2)	-19258	54	3	0.0002
Equal κ (Hypothesis 3)	-19264	54	3	0.0675
$\tilde{\theta} = 0$ (Hypothesis 4)	-19077	53	4	$<$0.0001
Market price of risk = 0 (Hypothesis 5)	-19236	55	2	$<$0.0001
Equal Measurement covariances (Hypothesis 6)	-18941	42	15	$<$0.0001

We also test specifically that $\tilde{\theta}_{\text{Regime 1}} = \tilde{\theta}_{\text{regime 2}} = \vec{0}$ (Hypothesis 4). We reject the hypothesis and conclude that the long term mean under the risk neutral measure is different from zero in at least one regime.[12] Hypothesis 5 tests whether the market price of risk exists by testing whether the state transition matrices under measures Q and P are equal. We find that Hypothesis 5 is rejected and that in our sample, the risk of regime switching is priced. This conclusion is consistent with DSY, although in their model, the transition probability under the real world measure is heterogeneous.

Finally, we test whether the measurement covariances are identical across regimes (Hypothesis 6). Failure to reject the hypothesis means that the errors in the measurement equation do not change with the regime i.e. that the model captures all regime-dependent behavior in yields. Unfortunately, we reject Hypothesis 6 and conclude that there are residual regime-switching elements that are not explained by the model. Furthermore, it appears that the differences in the covariance matrices are one of the major drivers of the regimes. However we observe that forcing the covariances to be equal does not destroy the regime structure. When viewed together these findings suggest that there is scope for improving this model possibly by the specification of a more robust measurement equation. This highlights the trade-off that exists between models that have robust econometric specifications and economically defensible (arbitrage-free) ones.

[12]Motivated by the need to make the model admissable in $A_0(3)$ sense (see Dai and Singleton (2000)), CDR assume that $\tilde{\theta} = \vec{0}$. However, we do not impose this restriction because the homoskedasticity of the model structure in this paper (i.e. $\vec{b}_i = \vec{0}$) guarantees identifiability. Additionally, because δ_0 and δ_1 are fixed in this model, we consider it unnecessary to restrict the long term mean of the factors.

7 Conclusion

We present an affine, arbitrage-free regime-switching dynamic Nelson-Siegel model that generalizes the dynamic Nelson-Siegel model of Christensen, Diebold and Rudebusch (2010). In the process of deriving this model we show, in detail, what restrictions need to be imposed on the general affine arbitrage-free hidden Markov model in order to obtain tractable specifications. We show that the resulting Nelson and Siegel regime switching model has the level, slope and curvature factors of CDR and the yield-adjustment term derived in DSY.

We propose an approximate maximum likelihood algorithm to estimate regime switching term structure models based on Kim and Nelson (1999C). This is in contrast with other authors who have assumed that the term structure is matched exactly at only a few points. We estimate the model on data from April 1991 to August 2010. This data spans two business cycles and includes two NBER recessions. We exclude the period of the monetary experiment in the 1980's, since this has been studied by a number of regime switching studies and may affect regime recognition in the most recent period.

We identify two regimes: (i) an expansionary monetary policy regime, characterized by low long term mean of interest rates, but high volatility; and (ii) a contractionary monetary policy regime, with high levels of interest rates and low volatilities.Our regime labeling is different from comparable studies, which usually identify expansionary and recessionary macroeconomic regimes. We find that although our regimes capture the NBER recession dates, they persist long after the recession has ended. This suggests that monetary policy is effective in combating recessions, but monetary easing persists too long.

In line with the intuition, we find that the contractionary monetary policy regime is far more persistent than the expansionary monetary policy regime (which coincides with NBER recessions). We find that the market prices of regime switching risk are highest when investor are faced with the threat of transition into Regime 2 - the recessionary regime - and that the highest market price of regime switching risk is associated with the probability of transitioning into Regime 2 from Regime 1. We find that the market price of interest rate risk is consistently higher in Regime 2 compared to Regime 1. During recessionary periods when investor risk aversion increases, the difference between the risk neutral and physical measures increases causing an increase in the market prices of interest rate risk. The dynamics of the market price of regime switching and interest rate risks in a two-regime framework is interesting and we leave this research for future work.

We compare our model to single regime benchmarks, the DNS and AFNS model in particular and find that in-sample, our model outperforms these models, even when we control for the increased number of parameters using the AIC and BIC criteria. We show that the estimated average term structure in our model fits neatly on either side of the other models. We also document that the shape of the yield adjustment term in our model is significantly different to that in CDR. This is due to the fact to that no unnecessary restrictions were placed on the long term mean under the risk neutral measure.

We also test for equivalence in regime switching parameters. In particular, we show that although the long term means and variances of the regimes are regime dependent, the mean reversion term appears regime invariant. This restriction can be imposed in estimating this model in practice, with no significant loss of fit over this period. Consistent with DSY, we find that market price of regime switching risk is priced and that measurement covariance is regime dependent. This suggests that some regime switching features of the yield remain unexplained in our model.

One of the main features of our model is that it is feasible and relatively easy to implement. Our model requires the estimation of 27 term structure parameters (excluding measurement covariance matrix). In contrast DSY (2010) estimate a 50 parameter model, but impose a number of ad hoc (and identifying) restrictions to reduce that to 33 parameters.

The restrictions imposed in this paper are grounded in the NS representation of the term structure. For instance, we assume that our κ has a specific structure under the risk-neutral measure. But this structure leads to a system of ordinary differential equations; the solution of which gives the level, slope and curvature factors of the yield curve. Additionally we assume that the term structure factors are independent. This is a natural assumption in term structure literature as the level, slope and curvature factors are often related to the first three principal components which are, by construction, orthogonal. In estimating our model we do impose several simplifying assumptions. In particular we assume that the price of regime switching risk is constant, or equivalently, that the transition matrix under the real world measure is homogeneous. This ensures that ease of estimation is maintained, but it also circumvents the specification of the heterogeneous transition probability.

In this work, we present a model which is feasible enough so that one-step-ahead forecasting becomes possible in a flexible but easy-to-implement setting. In a follow-up paper we focus on the true test of a model such as the one presented in this paper-out-of-sample performance. Such regime switching models which are able to consider dramatic changes in macroeconomic conditions can be of immense practical importance in a volatile economic environment. An additional topic for investigation would be to make the transition probability matrix heterogeneous and investigate whether the transition probabilities yield insights into the risk aversion parameters of the market. We leave these investigations for future research.

References

Bjork, Kabanov, Runggaldier (1997), "Bond market structure in the presence of marked point processes," Mathematical Finance, Vol. 7, No. 2, 211-239

Christensen, Diebold and Rudebusch (2009), "An arbitrage free generalized Nelson-Siegel term structure model," The Econometrics Journal, Vol. 12, pp 33-64

Christensen, Diebold and Rudebusch (2010), "An arbitrage-free class of Nelson-Siegel term structure model," Journal of Econometrics, Vol. 164, pp 4-20

Bansal, Tauchen and Zhou (2003), "Regime-shifts, risk premiums in the term structure, and the business cycle", Federal Reserve Working Paper

Bansal and Zhou (2002), "Term structure of interest rates with regime shifts", Journal of Finance, Vol. 57, No. 5, 1997-2043

Bernadell, Coche, and Nyholm (2005), Yield Curve Prediction for the Strategic Investor," European Central Bank Working Paper No. 472

Blais (2009), "A Bayesian analysis of affine term structure models," WP, Bank of Canada

Bremaud (1981), "Point Processes and Queues : Martingale Dynamics," Springer Verlag

Diebold and Li (2006), "Forecasting the term structure of government bond yields," Journal of Econometrics, Vol. 130, pp 337-364

Dai, and Singleton (2000). Specification Analysis of Affine Term Structure Models. Journal of Finance LV, pp 1943-1978.

Dai, Singleton and Yang (2007), "Regime shifts in a dynamic term structure model of U.S. treasury bond yields," Review of Financial Studies, Vol. 20, pp 1669-1706

Duffie, Kan (1996), "A yield-factor model of interest rates," Mathematical Finance, Vol. 6, No. 4, 379-406

Gray (1996), "Modelling the conditional distribution of interest rates as a regime switching process", Journal of Financial Economics, Vol. 42, No. 1, 27-62

Joslin, Singleton and Zhu (2011), "A new perspective on Gaussian term structure models," Review of Financial Studies, Forthcoming

Hamilton (1989), "A new approach to the economic analysis of non-stationary time series and the business cycle," Econometrica, Vo. 57, No. 2, 357-384

Hamilton (1990), "Analysis of time series subject to changes in regime", Journal of Econometrics, No. 45, 39-70

Hansen and Poulsen (2000), "A simple regime switching model of term structure", Finance and Stochastics, Vol. 4, 409-429

Heath, Jarrow, Morton (1992), "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation", Econometrica, Vol. 60, No. 1, 77-105.

Hull and White (1993), “One factor interest-rate models and the valuation of interest rate derivative securities”, The Journal of Financial and Quantitative Analysis, Vol. 28, No. 2, 235-254

Kim (1994), “Unobserved component time series models with Markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainty”, Journal of Business and Economic Statistics, Vol. 11, No. 3, 341-349

Kim (1996), “Predicting business cycle phases with indexes of leading and coincident indicators: A multivariate regime-shift approach”, Journal of Monetary Economics, Vol2, No. 2, 1-27

Kim and Nelson (1998), “Business cycle turning points, a new coincident index and tests of duration dependence based on a dynamic factor model with regime switching”, Review of Economics and Statistics, Vol. 80, 188-201

Kim and Nelson (1999A), “Friedman’s plucking model of business fluctuations: tests and estimates of permanent and transitory components”, Journal of Money, Credit and banking, Vol. 31, No. 3, 317-334

Kim and Nelson (1999B), “Has the US economy become more stable? A Bayesian approach based in a Markov-switching model of the business cycle”, Review of Economics and Statistics, Vol. 81, No. 4, 608-6616

Kim and Nelson (1999C), “State-space models with regime switching”, MIT Press

Landen (2000), “Bond pricing in a hidden Markov model of the short rate”, Finance and Stochastics, Vol. 4, 371-389

Lee and Naik (1998), “Yield curve dynamics with discrete shifts in economic regimes: Theory and estimation”, Working Paper

Nelson and Siegel (1987), ”Parsimonious modeling of yield curves,” The Journal of Business, Vol. 60, No. 4, pp 473-489

Protter, P. (2003), Stochastic Integration and Differential Equations”, 2nd Edition, Berling: Springer Verlag

Wu and Zhang (2004), ”Affine regime-switching models for interest rate term structure,” Contemporary Mathematics

Zantedeschi, D., P. Damien, and N.G Polson (2011), Predictive Macro-Finance With Dynamic Partition Models,” Journal of the American Statistical Association, 106, 427-439

Appendix

Marked point processes

Definition: *An E-marked point process is a paired sequence $(T_n, Y_n)_{n\geq 1}$ where (i) T_n is a point process and (ii) Y_n is a sequence of E-valued random variables.*

We may define a counting process on the set $A \in 2^E$ as

$$N_t(A) = \sum_{n\geq 1} \mathbf{1}_{T_n \leq t} \mathbf{1}_{Y_n \in A}. \tag{54}$$

We can thus establish a relationship between the continuous and discrete time version of this notation as $N_t(A) = \mu((0,t], A)$ where μ is as given in (4). Denote $N_t = N_t(E)$ i.e. $\int_E$ in continuous time is equivalent to $\sum_{E,i\neq j}$ in discrete time.

Derivation of the Pricing Kernel

As shown in Bremaud (1981), given any Wiener process (W_t) and a marked point process $(\mu(dt, dz))$ a local martingale (M_t) can be represented as

$$\frac{dM_t}{M_{t-}} = -r_t dt - \lambda_t d\tilde{W}_t - \int_E \psi(z, X_{t-})[\tilde{\mu}(dt, dz) - \tilde{\gamma}(dt, dz)], \tag{55}$$

where λ is the market price of diffusion risk and ψ is the market price of regime-switching risk. From this, we may use the exponential formula (Bremaud, 1981) to obtain the solution to M_t

$$M_t = exp\Big(-\int_0^t (r_s + \frac{1}{2}\lambda\lambda')ds - \frac{1}{2}\int_0^t \lambda' d\tilde{W}_t + \int_0^t \int_E \psi(z, X_{t-})\gamma(ds, dz)\Big) \prod_{n=1}^{N_t} (1 - \psi(T_n, Y_n)). \tag{56}$$

Recognizing that $(1 - \psi(T_n, Y_n)) = exp^{log((1-\psi(T_n,Y_n)))}$ and using the definition of dN_t we obtain the pricing kernel in (10).

The continuous version (without jumps) of the Radon-Nikodym derivative is typically expressed as

$$\frac{dQ}{dP} = \mathcal{E}(-\int_0^t \lambda_s dW_s) \tag{57}$$

where $\mathcal{E}$ denotes the stochastic exponential. We can extend this formula to the case with a jump described by a marked point process by extending the Radon-Nikodym derivative as

$$\frac{dQ}{dP} = \mathcal{E}\Big(-\int_0^t \lambda_s dW_s + \int_o^t \int_E \psi_s[\tilde{\mu}(ds, dz) - \tilde{\gamma}(ds, dz)]\Big). \tag{58}$$

Hence, using the results in Bremaud (1981), we express the Radon-Nikodym derivative for a change of measure in the explicit form

$$\begin{aligned} \frac{dQ}{dP} &= exp\Big(-\frac{1}{2}\int_0^t \lambda\lambda' ds - \int_0^t \lambda_s dW_s \\ &+ \int_0^t \int_E \psi_s \tilde{\gamma}(ds,dz) + \int_0^t \int_E log(1-\psi_s)\tilde{\mu}(ds,dz)\Big), \end{aligned} \tag{59}$$

where in the last expression we have used an argument similar to (56) and (10).

Proof of Proposition 1

Define $f = e^{A(\tau,s_t)+B(\tau,s_t)'X_t}$. By the multivariate Ito's Lemma for semi-martingales (Protter, 2003) $df = \frac{\partial f}{\partial t} + \nabla_X^T f.dX + \frac{1}{2}dX^T.\nabla_X^2 f.dX + [f(t,X_t,s_t,T) - f(t-,X_{t-},s_{t-},T)]$. We can show that the first three terms on the right hand side resolve to

$$f(-\frac{\partial A(\tau,s_t)}{\partial \tau} - X'\frac{\partial B(\tau,s_t)}{\partial \tau})dt + fB'dX + f\frac{1}{2}B'\Sigma(s_t)V(s(t))\Sigma(s_t)'Bdt. \tag{60}$$

Collecting the dt terms together, we have

$$\begin{aligned} \frac{df}{f} = [(-\frac{\partial A(\tau,s_t)}{\partial \tau} - X'\frac{\partial B(\tau,s_t)}{\partial \tau}) + B'[\tilde{\kappa}(\theta(\tilde{s_t}) - X)] + \frac{1}{2}B'\Sigma(s_t)V(s(t))\Sigma(s_t)'B]dt \\ +B'\Sigma(s_t)\sqrt{V(s(t))}dW + \frac{1}{f}[f(t,X_t,s_t,T) - f(t-,X_{t-},s_{t-},T)]. \end{aligned}$$

The last term in this expression pertains to the regime shift. This term can be expressed as $\int_E \Delta_s f\tilde{\mu}(dt,dz)$. The compensator function can be subtracted and added back within the dt term. This gives us

$$\begin{aligned} \frac{df}{f} &= \Big[(-\frac{\partial A(\tau,s_t)}{\partial \tau} - X'\frac{\partial B(\tau,s_t)}{\partial \tau}) + B'[\tilde{\kappa}(\theta(\tilde{s_t}) - X)] + \frac{1}{2}B'\Sigma(s_t)V(s(t))\Sigma(s_t)'B \\ &+ \frac{1}{f}\int_E \nabla_s\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz)\Big]dt + B'\Sigma(s_t)\sqrt{V(s(t))}dW \\ &+ \frac{1}{f}\int_E \Delta_s f(\mu(dt,dz) - \tilde{\gamma}_\mu(dt,dz)). \end{aligned}$$

No arbitrage implies that the instantaneous expected return of all assets equal the short-term risk-free interest rate under the risk-neutral measure. Hence, we have

$$\begin{aligned} &(-\frac{\partial A(\tau,s_t)}{\partial \tau} - X'\frac{\partial B(\tau,s_t)}{\partial \tau}) + B'[\tilde{\kappa}(\theta(\tilde{s_t}) - X)] + \frac{1}{2}B'\Sigma(s_t)V(s(t))\Sigma(s_t)'B \\ &+ \frac{1}{f}\int_E \nabla_s\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz) = r. \end{aligned} \tag{61}$$

Separating the vector and scalar components, we can re-write this equation as

$$\begin{aligned} - & \frac{\partial A(\tau, s_t)}{\partial \tau} + (\tilde{\kappa}\theta(\tilde{s_t}))' B + \frac{1}{2} B'\Sigma(s_t)V(s(t))\Sigma(s_t)'B + \frac{1}{f}\int_E \Delta_s f\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz) \\ - & X'\frac{\partial B(\tau, s_t)}{\partial \tau} - X'\tilde{\kappa}'B = r. \end{aligned} \tag{62}$$

Consider the term $\int_E \Delta_s f\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz)$. This expands to

$$\int_E [e^{A(\tau,s+\zeta)+B(\tau,s+\zeta)'X} - e^{A(\tau,s)+B(\tau,s)'X}]\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz).$$

f does not depend on the mark space, E. Hence, we bring this factor outside the integral and write

$$\begin{aligned} & \int_E [e^{A(\tau,s+\zeta)+B(\tau,s+\zeta)'X} - e^{A(\tau,s)+B(\tau,s)'X}]\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz) \\ = \;& f\int_E [e^{\Delta_s A(\tau,s_t)+\Delta_s B'X)} - 1]\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz). \end{aligned}$$

Substituting into equation (62), we obtain

$$\begin{aligned} - & \frac{\partial A(\tau, s_t)}{\partial \tau} + [\tilde{\kappa}(\theta(\tilde{s_t})]' B + \frac{1}{2} B'\Sigma(s_t)V(s(t))\Sigma(s_t)'B \\ + & \int_E [e^{\Delta_s A(\tau,s_t)+\Delta_s B'X)} - 1]\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz) - X'\frac{\partial B(\tau, s_t)}{\partial \tau} - X'\tilde{\kappa}'B = r. \end{aligned} \tag{63}$$

Consider the term $\int_E [e^{\Delta_s A(\tau,s_t)+\Delta_s B'X)} - 1]\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz)$. Substituting from (6) we have

$$\begin{aligned} & \int_E [e^{\Delta_s A(\tau,s_t)+\Delta_s B'X)} - 1]\tilde{h}\mathbf{I}\{s=i\}\epsilon_z(dz) \\ = & \int_E [e^{\Delta_s A(\tau,s_t)+\Delta_s B'X)} - 1]e^{\tilde{h}_0(z,X)+\tilde{h}_1(z,X)'X}\mathbf{I}\{s=i\}\epsilon_z(dz) \\ = & \int_E [e^{[\Delta_s A(\tau,s_t)+\tilde{h}_0(z,X)]+[\Delta_s B+\tilde{h}_1(z,X)]'X} - e^{\tilde{h}_0(z,X)+\tilde{h}_1(z,X)'X}]\mathbf{I}\{s=i\}\epsilon_z(dz). \end{aligned}$$

From the Taylor series expansion $[e^{[\Delta_s B+\tilde{h}_1(z,X)]'X} \approx 1 + [\Delta_s B + \tilde{h}_1(z,X)]'X$ and similarly, $e^{\tilde{h}_1(z,X)'X} \approx 1 + \tilde{h}_1(z,X)'X$. Substituting into (63) yields

$$\begin{aligned} - & \frac{\partial A(\tau, s_t)}{\partial \tau} + [\tilde{\kappa}(\theta(\tilde{s_t})]' B + \frac{1}{2} B'\Sigma(s_t)V(s(t))\Sigma(s_t)'B \\ + & \int_E [e^{\Delta_s A(\tau,s_t)} - 1]e^{\tilde{h}_0} + [e^{\Delta_s A(\tau,s_t)+\tilde{h}_0}(\Delta_s B + \tilde{h}_1) - e^{\tilde{h}_0}\tilde{h}_1]'X\mathbf{I}\{s=i\}\epsilon_z(dz) \\ - & X'\frac{\partial B(\tau, s_t)}{\partial \tau} - X'\tilde{\kappa}'B = r. \end{aligned}$$

The current state is known, and so we set $\mathbf{I}\{s = i\} = 1$. Since the integral is taken over the entire mark space, we can also set $\epsilon_z = 1$.

$$
\begin{aligned}
- \quad & \frac{\partial A(\tau, s_t)}{\partial \tau} + [\tilde{\kappa}(\tilde{\theta(s_t)})]' B + \frac{1}{2} B' \Sigma(s_t) V(s(t)) \Sigma(s_t)' B \\
+ \quad & \int_E [e^{\Delta_s A(\tau, s_t)} - 1] e^{\tilde{h}_0}(dz) + \int_E [e^{\Delta_s A(\tau, s_t) + \tilde{h}_0} (\Delta_s B + \tilde{h}_1) - e^{\tilde{h}_0} \tilde{h}_1]' X(dz) \\
- \quad & X' \frac{\partial B(\tau, s_t)}{\partial \tau} - X' \tilde{\kappa}' B = r
\end{aligned}
$$

Substituting for $V(s(t))$ from (2)

$$
\begin{aligned}
- \quad & \frac{\partial A(\tau, s_t)}{\partial \tau} + [\tilde{\kappa}(\tilde{\theta(s_t)})]' B + \frac{1}{2} \sum_{i=1}^{n} [B' \Sigma(s_t)]_i^2 a_i(s_t)] + \frac{1}{2} \sum_{i=1}^{n} [B' \Sigma(s_t)]_i^2 b_i(s_t) X_i(t)] \\
+ \quad & \int_E [e^{\Delta_s A(\tau, s_t)} - 1] e^{\tilde{h}_0}(dz) + \int_E [e^{\Delta_s A(\tau, s_t) + \tilde{h}_0} (\Delta_s B + \tilde{h}_1) - e^{\tilde{h}_0} \tilde{h}_1]' X(dz) \\
- \quad & X' \frac{\partial B(\tau, s_t)}{\partial \tau} - X' \tilde{\kappa}' B = r
\end{aligned}
$$

Using (3) and separating the vector factors δ_1 of $X(t)$ from the scalar δ_0 we have (8) and (9). $\square$