Birkbeck Centre for Applied Macroeconomics

ISSN 1745-8587

Birkbeck
UNIVERSITY OF LONDON

BCAM 1403

Estimating and Forecasting the Yield Curve Using a Markov Switching Dynamic Nelson and Siegel Model

Constantino Hevia
World Bank

Martin Gonzalez-Rozada
Universidad Torcuato Di Tella

Martin Sola
Birkbeck, University of London

Fabio Spagnolo
Brunel University

July 2014

BCAM

Estimating and Forecasting the Yield Curve Using a Markov Switching Dynamic Nelson and Siegel Model

Constantino Hevia[a,c,*], Martin Gonzalez-Rozada[c],
Martin Sola[b,c], and Fabio Spagnolo[d]
[a]World Bank, Washington DC, USA
[b]School of Economics, Mathematics & Statistics, Birkbeck College, London, UK
[c]Department of Economics, Universidad Torcuato Di Tella, Buenos Aires, Argentina
[d]Department of Economics and Finance, Brunel University, Uxbridge, UK

This draft: January 2014; First draft: July 2011

Abstract

We estimate versions of the Nelson-Siegel model of the yield curve of U.S. government bonds using a Markov switching latent variable model that allows for discrete changes in the stochastic process followed by the interest rates. Our modeling approach is motivated by evidence suggesting the existence of breaks in the behavior of the U.S. yield curve that depend, for example, on whether the economy is in a recession or a boom, or on the stance of monetary policy. Our model is parsimonious, relatively easy to estimate, and flexible enough to match the changing shapes of the yield curve over time. We also derive the discrete time non-arbitrage restrictions for the Markov switching model. We compare the forecasting performance of these models with that of the standard dynamic Nelson and Siegel model and an extension that allows the decay rate parameter to be time-varying. We show that some parameterizations of our model with regime shifts outperform the single regime Nelson and Siegel model and other standard empirical models of the yield curve.

Key words: Yield Curve; Term structure of interest rates, Markov regime switching; Maximum likelihood; Risk premium.

JEL classification: C13; C22; E43.

*We thank the editor and two anonymous referees for very useful comments. Constantino Hevia (chevia@utdt.edu) is the corresponding author. Address: Universidad Torcuato Di Tella, Av. Figueroa Alcorta 7350 (C1428BCW), Ciudad de Buenos Aires, Argentina. Phone: 5411 51697300 Fax: 5411 51697348

1 Introduction

Recent developments in the modeling of government bond yields assume that a handful of possibly unobserved factors determine the evolution of the entire yield curve. Broadly speaking, the literature on the yield curve evolved into two related branches: the class of arbitrage-free affine term structure models (e.g. Piazzesi, 2010) and the class of dynamic Nelson and Siegel (1987) models, as proposed by Diebold and Li (2006). Both classes of models assume that observed yields are an affine function of the factors, thereby reducing the variability of the entire yield curve to the variability of a few factors. The two approaches differ in the construction of the factor loadings and, possibly as well, in the interpretation of the factors. While in arbitrage-free models the factor loadings are derived from imposing lack of arbitrage across bonds of different maturities—using an appropriate log-linear stochastic discount factor—the class of Nelson and Siegel models impose a parsimonious parametric structure to the loading on the factors. These loadings are a function of a single parameter, which we denote by λ, usually referred to as the "exponential decay rate parameter."[1]

Within the dynamic Nelson and Siegel framework (DNS), there are three unobserved factors that evolve as a vector autoregression of order one. While widely used by practitioners, this framework has two potential shortcomings: first, the model does not rule out arbitrage opportunities across bonds of different maturities, and second, the crucial exponential decay rate parameter seems to change over time. As a response to these shortcomings, the literature evolved in different ways: Christensen et al., (2011) derive arbitrage-free conditions for the DNS model and evaluate to what extent they improve the forecasting ability of the model, while Koopman et al., (2010) model the exponential decay rate parameter λ as a fourth unobserved component, thereby affecting the loadings on the other three factors.[2]

To our knowledge, it is not possible to account for both extensions simultaneously. Therefore, the purpose of this paper is twofold. First, to evaluate whether the changes in the yield

[1] Work in the arbitrage-free tradition includes, among others, Knez et al., (1994); Duffie and Khan (1996), Dai and Singleton (2000), and Ang and Piazzesi (2003). Work in the Nelson and Siegel tradition includes Diebold et al., (2006); Yu and Zivot (2006); and Bianchi et al., (2009).

[2] Fitting a Nelson and Siegel model using arbitrage-free data, however, could presumably produce a fitted yield curve that is approximately arbitrage-free (see the findings in Coroneo et al., 2011).

curve driven by changes in the key parameter λ can be captured using a two regime Markov switching model to avoid potential overfitting of the data. And second, to what extent it is possible to incorporate non-arbitrage restrictions within the Markov switching framework and, at the same time, account for changes in the shape of the yield curve over time.[3] In all cases, we conduct a forecasting exercise to evaluate the relative merits of allowing for variation in the parameter λ and of imposing the non-arbitrage restrictions to the Markov switching model. For comparison purposes, we also present a version of the four factor DNS model of Koopman et al., (2010).

To obtain the arbitrage-free representation of the Markov switching DNS model (MS-DNS), we extend Niu and Zeng (2012) who provide a discrete time derivation of the (single regime) arbitrage-free DNS model of Christensen et al., (2011). Importantly, we found that imposing non-arbitrage restrictions to the MS-DNS model requires the parameter λ to be constant across regimes. Therefore, the arbitrage-free MS-DNS model can only have switching in the measurement equation through an additional regime-specific constant implied by the absence of arbitrage—of course, the model also allows for switching in the state equation.[4]

All the models that we consider have a state-space representation, where a state equation determines the evolution of a set of unobserved state variables, and a measurement equation relates the observed yields to the state variables. A Markov switching structure can be added to the state equation, to the measurement equation, or to both. Our results strongly suggest that the Markov switching parameterization needs to be present in the measurement equation.[5] This can be achieved either by allowing the parameter λ in the measurement equation to switch across regimes, or by fixing the parameter λ while imposing non-arbitrage restrictions. The latter specification introduces an additional maturity-specific constant that

[3] Using rolling windows, Coroneo et al. (2013) report that the mean squared errors of the forecasts of excess holding bond returns increase dramatically at the beginning of recessions. This result is consistent with the view that the yield curve may be subject to structural shifts.

[4] Xiang and Zhu (2013) propose a MS-DNS model with switching only in the state equation. Bandara and Munclinger (2012) derive arbitrage-free restrictions in a continuous time version of the MS-DNS model. They assume that there is no switching in the parameter λ and do not discuss if this is a parameterization choice or a necessary condition for the absence of arbitrage. Furthermore, they do not perform forecasting exercises, which is one of the main objetives of this paper.

[5] The reason is that the single regime DNS model is quite successful in modeling the three unobserved components. Our estimated models with only switching in the unobserved components do not improve the fit of the model relative to the single regime model.

switches across regimes. The proposed Markov switching models, with and without non-arbitrage restrictions, are parsimonious and relatively easy to estimate. This simplicity is accomplished by evaluating the likelihood function using an approximate non-linear filter that collapses a growing mixtures of densities to a single density, dramatically reducing the dimensionality of the estimation problem.

Based on US zero-coupon data, all of our estimated models present significant evidence of regime shifts. Our results suggest that the conventional stylized facts of the yield curve are roughly associated with booms or with periods of active monetary policy as identified by Bikbov and Chernov (2013). In contrast, the characteristics of the yield curve during recessions are rather different. The models that we propose seem to not only successfully characterize the data under scrutiny but also, and more importantly, to have a good forecasting performance. Some of the models with a Markov switching structure have a better forecasting performance than standard empirical models of the yield curve. The forecasting results are particularly noteworthy because one of the perceived weaknesses of nonlinear models is their relatively poor out-of-sample performance.

We compare the forecasting performance of the proposed models to that of the single regime DNS model by computing mean squared errors, tests of equal forecast accuracy, and the proportion of times that each model attains the lowest forecast error. We found that imposing non-arbitrage restrictions to the Markov switching model seems to produce improved forecasts relative to the single regime DNS model only at short and medium horizons. In addition, some parameterizations of the Markov switching models that allow for shifts in the parameter λ—and hence, do not impose arbitrage restrictions—have better forecasting performance than the single regime DNS model. On the other hand, the four factor DNS model in which λ is treated as a fourth unobserved factor outperforms all the other models in terms of fit. Interestingly, we found that this model performs very well in terms of forecasting but only at the shortest forecast horizon (1 month ahead). Yet, overfitting seems to be a problem since the forecasting performance of the model at medium and long horizons is rather poor relative to the other models. Overall, the paper shows that several models have better forecasting performance than the single regime three-factor model and that which model is preferred depends on the particular forecast horizon that is considered.

The paper is organized as follows. Section 2 presents the different extensions of the dynamic Nelson and Siegel model that we use throughout the paper. Section 3 describes the econometric model and an approximate filtering algorithm used to evaluate the likelihood function of the nonlinear models. In Section 4 we apply the models using U.S. data on government bond yields and assess the out-of-sample performance of the models. Section 5 concludes.

2 Models of the yield curve

This section describes different extensions of the Nelson and Siegel (1987) model that we use to parameterize the yield curve. The basic framework is the dynamic version of the Nelson and Siegel model developed by Diebold and Li (2006), which consists of a parsimonious model of the yield curve of the form

$$R_t(\tau) = \beta_{1t} + \beta_{2t}\frac{1-e^{-\lambda\tau}}{\lambda\tau} + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \varepsilon_t(\tau), \tag{1}$$

where t denotes time; $R_t(\tau)$ is the yield for a zero-coupon bond that matures in τ months; $\boldsymbol{\beta}_t = \{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ is a vector of unobserved latent factors that evolve as a first order vector autoregression; λ is a parameter associated with the exponential rate of decay of the factor loadings at different maturities; and $\varepsilon_t(\tau)$ is a measurement error distributed as $N(0, Q)$, where Q is a covariance matrix with dimension equal to the number of observed yields.[6] As usual, measurement errors are added to avoid the inherent stochastic singularity of the model.[7] Following Diebold et al., (2006), we estimate the parameters of this model—namely, λ and the parameters of the process for the latent factors—using maximum likelihood, where the likelihood function is evaluated using the Kalman filter.

[6] Throughout the paper, $N(\mu, \Omega)$ denotes a Normal distribution with mean μ and covariance matrix Ω.

[7] One of the main insights obtained from Diebold and Li (2006) is that β_{1t} is a long term factor associated with the level of interest rates, β_{2t} is a short term factor associated with the slope of the yield curve, and β_{3t} is a medium term factor associated with the curvature of the yield curve.

2.1 The Markov switching dynamic Nelson and Siegel model

The MS-DNS model postulates that there is an unobservable random variables x_t taking the values 0 or 1 that indexes the two different "regimes" in which the economy could be at time t. The variable x_t evolves over time according to a time-homogeneous Markov chain with transition probabilities

$$p_{00} = \Pr\left(x_{t+1} = 0 | x_t = 0\right) \text{ and } p_{11} = \Pr\left(x_{t+1} = 1 | x_t = 1\right). \tag{2}$$

The yield of a zero-coupon bond that mature in τ months is now given by

$$R_t\left(\tau\right) = \left(1 - x_t\right) R_{0t}\left(\tau\right) + x_t R_{1t}\left(\tau\right),$$

where

$$R_{it}(\tau) = \beta_{1t} + \beta_{2t}\frac{1 - e^{-\lambda_i \tau}}{\lambda_i \tau} + \beta_{3t}\left(\frac{1 - e^{-\lambda_i \tau}}{\lambda_i \tau} - e^{-\lambda_i \tau}\right) + \varepsilon_{it}\left(\tau\right) \tag{3}$$

is the yield conditional on regime $i = 0, 1$. In addition, $\boldsymbol{\varepsilon}_{it}$ is a vector of measurement errors with dimension equal to the number of observed yields distributed as $N\left(0, Q_i\right)$, where Q_i is a state-dependent covariance matrix for $i = 0, 1$. The distribution of the dynamic latent factors $\boldsymbol{\beta}_t = \{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ conditional on x_t is determined by the autoregressive process

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_i + F_i \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_{it}, \tag{4}$$

where $i = 0, 1$ denote the regime, $\boldsymbol{\mu}_i = \left(\mu_i^1, \mu_i^2, \mu_i^3\right)'$, F_i is a 3×3 matrix, and $\boldsymbol{\eta}_{it}$ is a 3×1 innovation normally distributed with mean zero and covariance matrix H_i. We parameterize $\lambda_i = \lambda_0(1-x_t)+\lambda_1 x_t$, $Q_i = Q_0\left(1 - x_t\right)+Q_1 x_t$, $\boldsymbol{\mu}_i = \boldsymbol{\mu}_0\left(1 - x_t\right)+\boldsymbol{\mu}_1 x_t$, $F_i = F_0\left(1 - x_t\right)+F_1 x_t$, and $H_i = H_0\left(1 - x_t\right) + H_1 x_t$.

The key feature of this model is that the yield curve depends on a variable that can be interpreted as capturing discrete changes in economic conditions. For example, this framework is able to capture that the slope of the yield curve is different at the different phases of the business cycle. This parameterization is intended to capture those effects.

2.1.1 The Markov switching DNS model with non-arbitrage restrictions

In this subsection we derive an arbitrage-free version of the MS-DNS model by extending the results in Niu and Zeng (2012) to a Markov switching framework. These authors provide a derivation in discrete time of the arbitrage-free DNS model obtained by Christensen et al., (2011) in continuous time. The arbitrage-free DNS model has a structure similar to that of pure affine models, but with restrictions on the values of the parameters of the key recursions associated with the class of affine term structure models.

As we show in the Appendix, the arbitrage-free MS-DNS model conditional on regime $i = 0, 1$ can be expressed as

$$R_{it}(\tau) = -\frac{A_i(\tau)}{\tau} + \beta_{1t} + \beta_{2t}\frac{1-e^{-\lambda\tau}}{\lambda\tau} + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \varepsilon_{it}(\tau), \tag{5}$$

where the constant $A_i(\tau)$ depends on the maturity of the bond and is determined recursively, under the risk-neutral measure, by the following pair of difference equations

$$\begin{aligned} A_i(\tau) &= p_{ii}\left(A_i(\tau-1) + \frac{1}{2}B'(\tau-1)H_iB(\tau-1) + B'(\tau-1)\zeta_i\right) \\ &\quad + p_{ij}\left(A_j(\tau-1) + \frac{1}{2}B'(\tau-1)H_jB(\tau-1) + B'(\tau-1)\zeta_j\right), \end{aligned}$$

$$B(\tau) = -\delta' + \Phi' B(\tau-1),$$

for $i, j = 0, 1$ and $i \neq j$; where δ is a 3×1 vector and Φ is a 3×3 matrix. The initial conditions for the recursion are $A_i(0) = 0$ and $B_i(0) = 1$. Following Dai and Singleton (2000) and Niu and Zeng (2012), for identification of the model we set $\zeta_i = (\psi_i, 0, 0)'$, where ψ_0 and ψ_1 are free parameters to be estimated along with the other parameters of the model. The Appendix shows the unique values that δ and Φ must take for the model to be consistent with the lack of arbitrage opportunities.

There are a number of points that are worth mentioning. As we discuss in the Appendix, the arbitrage-free MS-DNS model can only be derived by restricting the parameter λ to be the same in both regimes. Therefore, this result and equation (5) imply that there could be switching in the measurement equation only because the constants $A_i(\tau)$ are allowed to switch—through shifts in the parameter ψ_i. In addition, even though the arbitrage-free model (5) has more parameters than the MS-DNS model (ψ_0 and ψ_1), there is no restriction on these parameters that delivers the MS-DNS model as a special case. This follows from the observation that $A_i(\tau)$ cannot be made equal to zero for any choice of ψ_i. The property that the models are non-nested was already discussed by Christensen et al., (2011) in the context of the single regime model.

2.2 A dynamic Nelson and Siegel model with continuously time-varying λ

The last model that we consider is the four factor DNS model proposed by Koopman et al., (2010). This model allows for changes in the parameter λ_t as an additional unobserved component within the single regime Nelson and Siegel model. The observed yield $R_t(\tau)$ is now assumed to be given by

$$R_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) + \varepsilon_t(\tau), \tag{6}$$

where the unobserved factors $(\beta_{1t}, \beta_{2t}, \beta_{3t}, \log\lambda_t)$ evolve as a stable first order vector autoregression. As discussed below, the additional latent factor λ_t gives the model substantially more flexibility to match the different shapes of the yield curve over time, leading to a markedly improved in-sample fit relative to the baseline DNS model

Because λ_t enters nonlinearly in the measurement equation (6), we follow Koopman et al., (2010) and linearize the model around the long-run values $(\beta_1^*, \beta_2^*, \beta_3^*, \log\lambda^*)$ implied by the stochastic process followed by the latent variables. Once the observation equation is linearized, we use the Kalman filter to perform the maximum likelihood estimation of the parameters. Linearizing equation (6) gives

$$\begin{aligned} R_t(\tau) \approx\ & \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda^*\tau}}{\lambda^*\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda^*\tau}}{\lambda^*\tau} - e^{-\lambda^*\tau}\right) \\ & + \left[(\beta_2^* + \beta_3^*)\left(\frac{e^{-\lambda^*\tau}(1+\lambda^*\tau)-1}{\lambda^*\tau}\right) + \beta_3^*\tau\lambda^* e^{-\lambda^*\tau}\right](\log\lambda_t - \log\lambda^*) + \varepsilon_t(\tau). \end{aligned}$$

3 The econometric model

In this section we present an econometric model that accounts for the existence of different regimes when estimating the yield curve.[8] The model postulates the existence of an unobserved discrete variable, $x_t \in \{0, 1, .., K\}$, which indexes the current regime and follows a Markov chain with transition probabilities $p_{ij} = \Pr(x_t = j | x_{t-1} = i)$ for $i, j = 0, 1, ..., K$. At time $t = 1$, the probability of x_1 is given by $\Pr(x_1)$. We consider the following conditional linear Gaussian model where, for any $t \geq 1$ and regime x_t, the observation and state equations are given by

$$y_t = \Lambda_{x_t} f_t + \varepsilon_{x_t t} \quad (7)$$

$$f_t = \mu_{x_t} + A_{x_t} f_{t-1} + \eta_{x_t t}. \quad (8)$$

Here, $y_t \in \Re^m$ is a vector of observed variables, $f_t \in \Re^n$ is a vector of unobserved continuous state variables, $\varepsilon_{x_t t} \in \Re^m$ is normally distributed with mean zero and $m \times m$ covariance matrix Q_{x_t}; $\mu_{x_t} \in \Re^n$; A_{x_t} is an $n \times n$ matrix; and $\eta_{x_t t} \in \Re^n$ is normally distributed with mean zero and $n \times n$ covariance matrix H_{x_t}. Moreover, $\eta_{x_t t}$ and $\varepsilon_{x_t t}$ are independent of each other at all leads, lags, contemporaneously for different x_t, and independent of f_0, where f_0 is Gaussian with mean $\widehat{f}_0$ and $n \times n$ covariance matrix V_0.

3.1 Approximate filtering and evaluation of the likelihood function

Given a vector of parameters $\boldsymbol{\theta}$ and a sample $Y^T = \{y_1, y_2, ..., y_T\}$, we evaluate the log-likelihood function using the prediction-error decomposition formula

$$\ell\left(\boldsymbol{\theta}; Y^T\right) = \sum_{t=1}^{T} \log \Pr\left(y_t | Y^{t-1}\right),$$

where $Y^{t-1} = \{y_1, y_2, ..., y_{t-1}\}$ denotes the history of observations up to time $t-1$. The probabilities $\Pr\left(y_t | Y^{t-1}\right)$ are obtained as a by-product of a recursive Bayesian filter used

[8]Different names of this model are: *Multi-process class-II model* (Harrison and Stevens, 1976), *Dynamic linear model with Markov-switching* (Kim, 1994), and *State-space models with Markov-switching* (Kim and Nelson, 1999).

to estimate the distribution of the latent variables f_t and x_t conditional on Y^{t-1}. As it is well known, Bayesian filtering with Markov switching implies that posterior distributions are mixtures of prior distributions that grow exponentially with time. We operationalize the filter by collapsing the posterior mixture distribution of the unobserved state to a single distribution at each time t.

Given filtered probabilities $\Pr\left(f_{t-1}|Y^{t-1}\right)$ and $\Pr\left(x_{t-1}|Y^{t-1}\right)$, we begin by computing the posterior densities $\Pr\left(f_t|Y^t\right)$ and $\Pr\left(x_t|Y^t\right)$, and the contribution to the likelihood function $\Pr\left(y_t|Y^{t-1}\right)$. To that end, suppose that the filtered probability $\Pr\left(f_{t-1}|Y^{t-1}\right)$ is Gaussian,

$$\Pr\left(f_{t-1}|Y^{t-1}\right) = N(\widehat{f}_{t-1|t-1}, V_{t-1|t-1}). \tag{9}$$

The vector $\left\{f_{t-1}|Y^{t-1}\right\}$ is Gaussian by assumption at $t = 1$ and by our approximating formula at any other $t > 1$.

3.1.1 The prediction step

Model (8) implies that the prediction probability $\Pr\left(f_t|Y^{t-1}, x_t = j\right)$ is Gaussian, as it is an affine function of two Gaussian random variables, $\left\{f_{t-1}|Y^{t-1}\right\}$ and η_{jt}. Thus, $\Pr\left(f_t|Y^{t-1}, x_t = j\right) = N(\widehat{f}^j_{t|t-1}, V^j_{t|t-1})$, where

$$\widehat{f}^j_{t|t-1} = \mu_j + A_j \widehat{f}_{t-1|t-1} \text{ and } V^j_{t|t-1} = A_j V_{t-1|t-1} A_j' + H_j.$$

Likewise, equation (7) implies that $\Pr\left(y_t|Y^{t-1}, x_t = j\right) = N(\widehat{y}^j_{t|t-1}, \Omega^j_{t|t-1})$, where

$$\widehat{y}^j_{t|t-1} = \Lambda_j \widehat{f}^j_{t|t-1} \text{ and } \Omega^j_{t|t-1} = \Lambda_j V^j_{t|t-1} \Lambda_j' + Q_j.$$

It then follows that the contribution to the likelihood function at time t, $\Pr\left(y_t|Y^{t-1}\right)$, is a mixture of K Gaussian variables,

$$\begin{aligned} \Pr\left(y_t|Y^{t-1}\right) &= \sum_{j=0}^{K} \Pr\left(x_t = j|Y^{t-1}\right) \Pr\left(y_t|Y^{t-1}, x_t = j\right) \\ &= \sum_{j=0}^{K} \left(\sum_{i=0}^{K} p_{ij} \Pr\left(x_{t-1} = i|Y^{t-1}\right) \right) \Pr\left(y_t|Y^{t-1}, x_t = j\right), \end{aligned} \tag{10}$$

where the second equality uses Bayes' law and $\Pr\left(x_t = j|x_{t-1} = i, Y^{t-1}\right) = p_{ij}$.

3.1.2 The updating step

Given the observation y_t, we use Bayes' law to update the probabilities $\Pr\left(x_t|Y^t\right)$ and $\Pr\left(f_t|Y^t\right)$. In particular,

$$\Pr\left(x_t = j|Y^t\right) = \frac{\Pr\left(y_t|Y^{t-1}, x_t = j\right) \sum_{i=0}^{K} p_{ij} \Pr\left(x_{t-1} = i|Y^{t-1}\right)}{\Pr\left(y_t|Y^{t-1}\right)}$$

and

$$\Pr\left(f_t|Y^t, x_t = j\right) = \frac{\Pr\left(y_t|f_t, Y^{t-1}, x_t = j\right) \Pr\left(f_t|Y^{t-1}, x_t = j\right)}{\Pr\left(y_t|Y^{t-1}, x_t = j\right)}.$$

Using standard arguments, it is easy to show that $\Pr\left(f_t|Y^t, x_t = j\right) = N(\hat{f}^j_{t|t}, V^j_{t|t})$, where the mean and covariance matrix are given by

$$\begin{aligned} \widehat{f}^j_{t|t} &= \widehat{f}^j_{t|t-1} + V^j_{t|t-1}\Lambda'_j \left(\Omega^j_{t|t-1}\right)^{-1} \left(y_t - \Lambda_j \widehat{f}^j_{t|t-1}\right), \\ V^j_{t|t} &= V^j_{t|t-1} - V^j_{t|t-1}\Lambda'_j \left(\Omega^j_{t|t-1}\right)^{-1} \Lambda_j V^j_{t|t-1}. \end{aligned}$$

A direct corollary of this observation is that $\Pr\left(f_t|Y^t\right)$ is a mixture of $K+1$ Gaussian variables,

$$\Pr\left(f_t|Y^t\right) = \sum_{j=0}^{K} \Pr\left(x_t = j|Y^t\right) \Pr\left(f_t|Y^t, x_t = j\right).$$

3.1.3 Collapsing the posterior probability $\Pr\left(f_t|Y^t\right)$

So far we showed that, if the prior probability $\Pr\left(f_{t-1}|Y^{t-1}\right)$ is Gaussian, the posterior probability $\Pr\left(f_t|Y^t\right)$ is a mixture of $K+1$ Gaussian distributions. We operationalize the recursive evaluation of the filter by collapsing $\Pr\left(f_t|Y^t\right)$ to a single Gaussian distribution. In particular, the best approximating Gaussian distribution under the Kullback-Leibler pseudo-distance has the mean and covariance matrix of the Gaussian mixture (West and Harrison, 1997). Simple algebra shows that these means and covariances are given by

$$\begin{aligned}\widehat{f}_{t|t} &= \sum_{j=0}^{K} \Pr\left(x_t = j|Y^t\right) \widehat{f}^{j}_{t|t}, \text{ and} \\ V_{t|t} &= \sum_{j=0}^{K} \Pr\left(x_t = j|Y^t\right) \left(V^{j}_{t|t} + \left(\widehat{f}_{t|t} - \widehat{f}^{j}_{t|t}\right)\left(\widehat{f}_{t|t} - \widehat{f}^{j}_{t|t}\right)'\right)\end{aligned}$$

This assumption closes the approximate recursive Bayesian filter. Note that, even though the filtered probability of the state f_t is collapsed to a single Gaussian, the contribution to the likelihood function $\Pr\left(y_t|Y^{t-1}\right)$ is always a Gaussian mixture with $K+1$ components.

3.2 Forecasting

In this section we discuss an algorithm to compute optimal forecasts using the Markov switching model (7)-(8). As in the filtering step, forecast distributions are Gaussian mixtures that grow exponentially with the forecast horizon. However, because our longest forecast horizon is only 12 periods ahead, we are able to keep track of the growing Gaussian mixture.

We start forecasting at some time t using the filtered probabilities $\Pr\left(x_t|Y^t\right)$ and $\Pr\left(f_t|Y^t\right)$ obtained from the approximate Bayesian filter. Consider first forecasting future regime probabilities x_{t+h} at some horizon $h > 0$. Given the Markovian structure,

$$\Pr\left(x_{t+h} = j|Y^t\right) = \sum_{i=0}^{K} p^{(h)}_{ij} \Pr\left(x_t = i|Y^t\right).$$

where $p^{(h)}_{ij}$, the probability of moving from state i to state j in h periods, is equal to the (i, j) element of the matrix P^h.

Consider now the one-step ahead density $\Pr\left(f_{t+1}|Y^t, x_{t+1} = i_1\right)$. Equation (8) implies that

$$\Pr\left(f_{t+1}|Y^t, x_{t+1} = i_1\right) = N\left(\widehat{f}^{i_1}_{t+1|t}, V^{i_1}_{t+1|t}\right)$$

where $\widehat{f}_{t+1|t}^{i_1} = \mu_{i_1} + A_{i_1} \widehat{f}_{t|t}$, and $V_{t+1|t}^{i_1} = A_{i_1} V_{t|t} A'_{i_1} + H_{i_1}$. Integrating out the regimes gives the marginal probability

$$\Pr\left(f_{t+1}|Y^t\right) = \sum_{i_1=0}^{K} \Pr\left(f_{t+1}|Y^t, x_{t+1} = i_1\right) \Pr\left(x_{t+1} = i_1|Y^t\right).$$

Similarly, equation (7) implies

$$\Pr\left(y_{t+1}|Y^t, x_{t+1} = i_1\right) = N\left(\widehat{y}_{t+1|t}^{i_1}, \Omega_{t+1|t}^{i_1}\right)$$

where $\widehat{y}_{t+1|t}^{i_1} = \Lambda_{i_1} \widehat{f}_{t+1|t}^{i_1}$, and $\Omega_{t+1|t}^{i_1} = \Lambda_{i_1} V_{t+1|t}^{i_1} \Lambda'_{i_1} + Q_{i_1}$. Integrating over future regimes gives the forecast density

$$\Pr\left(y_{t+1}|Y^t\right) = \sum_{i_1=0}^{K} \Pr\left(y_{t+1}|Y^t, x_{t+1} = i_1\right) \Pr\left(x_{t+1} = i_1|Y^t\right).$$

Repeating the previous argument, equations (7) and (8) imply that the conditional h-period ahead forecast densities satisfy

$$\begin{aligned} \Pr\left(f_{t+h}|Y^t, x_{t+1} = i_1, x_{t+2} = i_2, ..., x_{t+h} = i_h\right) &= N\left(\widehat{f}_{t+h|t}^{i_1,i_2...,i_h}, V_{t+h|t}^{i_1,i_2...,i_h}\right) \\ \Pr\left(y_{t+h}|Y^t, x_{t+1} = i_1, x_{t+2} = i_2, ..., x_{t+h} = i_h\right) &= N\left(\widehat{y}_{t+h|t}^{i_1,i_2...,i_h}, \Omega_{t+h|t}^{i_1,i_2...,i_h}\right) \end{aligned}$$

where $\widehat{f}_{t+h|t}^{i_1,i_2...,i_h} = \mu_{i_h} + A_{i_h} \widehat{f}_{t+h-1|t}^{i_1,i_2...,i_{h-1}}$, $V_{t+h|t}^{i_1,i_2...,i_h} = A_{i_h} V_{t+h-1|t}^{i_1,i_2...,i_{h-1}} A'_{i_h} + H_{i_h}$, $\widehat{y}_{t+\tau|t}^{i_1,i_2...,i_h} = \Lambda_{i_h} \widehat{f}_{t+h|t}^{i_1,i_2...,i_h}$, and $\Omega_{t+\tau|t}^{i_1,i_2...,i_h} = \Lambda_{i_h} V_{t+h|t}^{i_1,i_2...,i_h} \Lambda'_{i_h} + Q_{i_h}$. Integrating over future regimes gives the forecast densities

$$\begin{aligned} \Pr\left(f_{t+h}|Y^t\right) &= \sum_{i_1,...,i_h=0}^{K} \Pr\left(f_{t+h}|Y^t, x_{t+1} = i_1, ..., x_{t+h} = i_h\right) \Pr\left(x_{t+h} = i_h|Y^t\right) \\ \Pr\left(y_{t+h}|Y^t\right) &= \sum_{i_1,...,i_h=0}^{K} \Pr\left(y_{t+h}|Y^t, x_{t+1} = i_1, ..., x_{t+h} = i_h\right) \Pr\left(x_{t+h} = i_h|Y^t\right) \end{aligned}$$

These prediction densities are used to compute the forecasts $E\left(y_{t+h}|Y^t\right)$.

Note that the h-period ahead forecast densities are a mixture of $(K+1)^h$ Gaussian variables. With $K = 1$ and $h = 12$ months, this is a mixture with 4096 components. While large, this mixture is still manageable using a standard laptop computer.

4 Empirical results

We examine U.S. Treasury yields of fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. The yields are derived from bid-ask average price quotes, from January 1972 through December 2000, as constructed by Diebold and Li (2006).[9]

There are many possible parameterizations of the Markov switching model specified by equations (3) and (4). We estimate special cases that constrain some parameters to be the same across regimes, along with the arbitrage-free version of the model. We also estimate the single regime model, and the linearized version of the four factor single regime model. In all cases, we evaluate the in-sample fit and out-of-sample forecasting performance of the models. To simplify the exposition, we do not report the estimates of those models that are clearly outperformed in terms of these criteria.[10] Since all the reported models are non-nested, we use information criteria for in-sample comparisons.[11]

Tables 1 and 2 show the estimation results of the different models. When the model allows for switching, the parameters corresponding to regimes 0 and 1 are shown in the first and second column, respectively. Model 1 in Table 1 corresponds to the baseline estimation without switching; Model 2 only allows for switching in the parameter λ and assumes diagonal F and H matrices; Model 3 differs form Model 2 in allowing also for switching in all the parameters of the state equation. Model 4 in Table 2 imposes the non-arbitrage restrictions to Model 3 but, as we explained above, the parameter λ is not allowed to switch. Finally, Model 5 is an extension of the Diebold and Li that allows for time-varying λ as an additional unobserved component.

[9] We use these particular data and sample period to compare our results to those of Diebold and Li.

[10] Christensen et al., (2011) note that it is common to find parameterizations of a yield curve model that are not rejected in-sample but that have very poor out-of-sample forecasting performance. This problem of overfitting is particularly important in non-linear models (e.g. Driffill et al. 2009). Given the large number of possible parametrizations that are special cases of the general model, we only report the two best restricted models. Results on the omitted models are available on request.

[11] Information criteria have been found to be useful in selecting among different regime dependent models (Psaradakis and Spagnolo, 2003 and 2006). Note, however, that comparing switching models with their single regime counterpart is problematic due to the usual nuisance parameter problem.

[Tables 1 and 2 about here]

Consider first the switching Models 2 and 3. In both cases, the estimated parameter λ is very different between regimes. The top panel in Figure 1 shows the yields for different maturities and NBER recessions, presented in shaded areas. A simple inspection of the figure shows that recessions are preceded by (and begin in) periods where the yield curve is relatively flat; afterwards, interest rates drop and the yield curve becomes steeper. The lower panel displays the smoothed probability of regime 0 of Model 3, $\Pr(x_t = 0|Y^T; \widehat{\boldsymbol{\theta}})$, where $\widehat{\boldsymbol{\theta}}$ is the estimated vector of parameters.[12] From the separation of the regimes, it can be inferred that the configuration of parameters corresponding to regime 0 is associated with periods of relatively flat yield curves. The estimated transition probabilities imply that the two regimes are persistent. For example, in Model 3, the economy spends about 56 percent of the time in regime 0 and 44 percent of the time in regime 1. Moreover, the expected number of months that the economy stays in regime 0 (regime 1) conditional on being in regime 0 (regime 1) is 14.5 months (11.6 months). Clearly, these probabilities do not coincide with the NBER dating of booms and recessions. However, the estimated probabilities of regime 1 tend to coincide with periods labeled as "active monetary policy" by Bikbov and Chernov (2013).

[Figure 1 about here]

We plot the regime specific loadings on factors β_{2t} and β_{3t} corresponding to Model 3 in the top panel of Figure 2; the lower panel of the figure displays the factor loadings of the linear Model 1. At short maturities, the loadings in regime 0 give comparatively more weight to factor β_{3t} and less weight to factor β_{2t} relative to those in regime 1. The estimated λ in regime 0 implies that the loading on factor β_{3t} is maximized at a maturity of 13 months while the estimated λ in regime 1 implies that the aforementioned loading is maximized at a maturity of about 30 months. This means that changes in the factor β_{3t} affect mostly short term yields in regime 0 but longer term yields in regime 1. On the other hand, the loadings in regime 1 always give more weight to changes in the factor β_{2t} than those in regime 0. To

[12] This panel also shows the smoothed probabilities for Model 4. Those results are commented below.

understand this result, recall that the model separates periods with flat yield curves (regime 0) from periods with steeper yield curves (regime 1). Consistent with this observation, the slope component β_{2t} is more relevant the steeper is the yield curve. Finally, the estimated volatilities of the state equation, measured by the diagonal elements of the matrix H, are largest for the curvature factor β_{3t} and lowest for the level factor β_{1t}.

[Figure 2 about here]

In Model 3 we observe different dynamics of the unobserved factors in each regime, as factor β_{2t} is more persistent in regime 0 than in regime 1. In addition, the drifts components $\boldsymbol{\mu}_i$ are very different between regimes, although many of them are not statistically significant. The remaining estimates are similar to those in Model 1. In terms of fit, both the Akaike and Schwarz information criteria (AIC and BIC) select Model 3 over Model 1 and Model 2.

In the single regime model, Diebold and Li (2006) interpret the factors β_{2t} and β_{3t} as associated with the slope and curvature of the yield curve, defined, respectively, as $R_t(120) - R_t(3)$ and $2R_t(24) - R_t(3) - R_t(120)$. Given the estimated values of $\lambda_0 = 0.13$ and $\lambda_1 = 0.05$, the slope and curvature of the yield curve in regime 0 satisfy

$$\begin{aligned} R_{0t}(120) - R_{0t}(3) &= -0.77\beta_{2t} - 0.08\beta_{3t} \\ 2R_{0t}(24) - R_{0t}(3) - R_{0t}(120) &= -0.27\beta_{2t} + 0.32\beta_{3t}. \end{aligned}$$

while, in regime 1, they are given by

$$\begin{aligned} R_{1t}(120) - R_{1t}(3) &= -0.77\beta_{2t} + 0.08\beta_{3t} \\ 2R_{1t}(24) - R_{1t}(3) - R_{1t}(120) &= 0.05\beta_{2t} + 0.34\beta_{3t}. \end{aligned}$$

Note that, in regime 1, the slope and curvature of the yield curve are indeed mostly associated with β_{2t} and β_{3t}, respectively. Yet, this association is less clear in regime 0. In particular, while the slope is still mostly affected by β_{2t}, the curvature is almost equally sensitive to variations in β_{2t} and β_{3t}. We computed the correlation between the estimated unobserved components and the level, slope, and curvature of the yield curve as in Diebold and Li (2006) and found that the correlation between β_{1t} and the level of yields, and between the β_{2t} and the slope is about 0.98 in both cases. On the other hand, and consistent with the observation made above, the correlation between the factor β_{3t} and the curvature is smaller (about 0.8).

In Table 3 and Figure 3 we show regime specific statistics computed by: i) dividing the sample between booms and recessions according to NBER dates (top panel) and ii) separating regimes using a dummy variable that equals 1 if $\Pr(x_t = 0|Y^T; \widehat{\boldsymbol{\theta}}) > 0.5$ and zero otherwise (bottom panel).[13] As noted above, recessions start in periods with flat yield curves; in addition, Model 3 assigns a high probability to regime 0 (associated with flat yield curves) during recessions. For this reason, we find the following similarities between the regimes separated by the model and the stylized facts in booms and recessions displayed in Table 3. First, average yields are higher in regime 0 than in regime 1; second, except at short maturities, the average yield curve is flatter in regime 0 than in regime 1; third, the average yield curve in regime 1 has a similar shape to that over the entire sample; fourth, the volatility of yields decreases sharply as maturity increases in regime 0, but it is flatter in regime 1 (although also decreasing with maturity); and finally, yields are more persistent in regime 1 than in regime 0 (recall that yields are also more persistent in booms). This difference in persistence, however, is more clear at short maturities.

Model 4 in Table 2 reports the estimates of the arbitrage-free MS-DNS model described in subsection 2.1.1.[14] As explained above, the switching in the measurement equation comes through a constant term that differs across maturities. In this model, the separation of regimes is very different from that in Model 3, and regime 1 is more closely related to NBER recessions (Figure 1).[15] Consistent with this observation, the expected time of staying in regime 1 is much lower than in regime 0. In addition, the persistence of the three unobserved components

[13]In computing these statistics, we only use sub-periods with six or more consecutive observations in a given regime.

[14]Since the model has an affine structure, we transformed the yields from percentage to rates and then reescaled the estimated parameters to make them comparable with the estimates from the other models.

[15]Notice that in model 4 the switching in the measurement equation arises because the constants $A_i(\tau)$ are allowed to switch—through shifts in the parameter ψ_i, which in turn affects the level of the yield curve under the risk neutral.measurement. This parameter that governs the regime specific price of the risk appears to be driven by booms and recessions.

is higher in regime 0 than in regime 1.

Finally, Model 5 reports the estimates of the single regime model that includes λ_t as an additional unobserved component (see equation (6)). For the diagonal parametrization, Models 2 and 5 only differ in the process assumed for λ. In both cases, λ is an unobserved component but in Model 2 λ can take only two values, while in Model 5 it can take a continuum of values. Interestingly, the estimates of the other parameters of the models are similar, suggesting that Model 2 successfully captures the variability in λ. Yet, as a result of the additional flexibility in the process assumed for λ, Model 5 outperforms all the other models in terms of fit, as reflected by the AIC and BIC criteria. This result is expected since, as shown in Koopman et al., (2010) using cross-sectional regressions, the parameter λ varies substantially over time.

4.1 Out-of-sample forecasts

In this section we evaluate the accuracy of the out-of-sample forecasts of the empirical models discussed earlier. In particular, we are interested in assessing the relative merits in terms of forecasting ability of either including non-arbitrage restrictions, as in Model 4, or allowing for time variation in the decay rate parameter λ, as in Models 2, 3, and 5. In Models 2 and 3, λ can take two possible values according to the Markov switching structure, while in Model 5 λ can take a continuum of values as an additional unobserved component in a single regime model. In summary, we found that the relative merits of the different model depend not only on the forecast horizon, but also on the maturity that is forecasted.

We compare the forecasting accuracy of Models 2 through 5 to that of the linear Model 1. We use Model 1 as the baseline for comparisons because: i) this model outperforms other standard forecasting models (random walks, VARs, etc.) and ii) we use the same data that Diebold and Li (2006) use in their forecasting exercises. This allows us to focus on the relative merits of the different models for a given sample of data. We compare the out-of sample forecasts based on a series of recursive forecasts beginning in 1994:1 and extending through 2000:12 (84 sample points), for all the available maturities, and for forecast horizons of h=1, 3, 6, and 12 months ahead.

To compare the forecast accuracy of the models, we compute the mean squared errors (MSE) of the forecasts, including the proportion of times that each model achieves the smallest squared forecast error over the 84 sample points. We also test for equal predictive accuracy using a modified version of the Diebold and Mariano (1995) test, proposed by Harvey et al., (1997).[16] Tables 4 to 7 report the results of the forecasting exercise. A first inspection of the tables shows that the linear Model 1 is generally outperformed by the nonlinear alternatives, with the only exception at long maturities for the 12 months ahead forecasts, implying that all the extensions considered in this paper are valuable in this respect.

The one month ahead forecasts, displayed in Table 4 show that Models 4 and 5 have the smallest MSEs for all but one maturity. This finding suggests that including non-arbitrage restrictions and allowing for time variation in the decay rate parameter λ is important for short horizon predictability. On the right panel of the table we report the proportion of times that each model achieves the smallest MSE over the 84 forecast periods calculated for each individual maturity. Contrary to the previous finding, Model 5 is outperformed by the alternative specifications most of the times. In particular, Model 5 outperforms the other models by this criterion for only 2 of the 17 maturities. In contrast, Models 2 and 3, that have a relativelly high MSE, achieve the smallest squared forecast errors most of the times. This is probably due to the fact that in some sub-periods Models 2 and 3 perform very poorly, probably as a results of a misclassification of states.

When we look at longer forecasting horizons, results change considerably. For example, for the 3 months ahead forecasts, Models 2 and 4 achieve the smallest MSE while Model 5 performs rather poorly. This finding suggests that the gains in terms of fit and short horizon predictability, driven by the greater flexibility in λ associated with Model 5, are lost for longer forecasting horizons. The good forecasting performance of Model 2 suggests that allowing for changes in λ is important for out-of-sample forecasting, but allowing for too much flexibility in λ, as in Model 5, could result in overfitting of the data. In other words, a parsimonious

[16] The purpose of the modification is to overcome the problem of over-sized of the original test in small and moderate samples (particularly acute for longer forecast horizons). The modified statistic is $S^* = \{[n + 1 - 2h + n^{-1}h(h-1)]/n\}^{1/2}S$, where n is the number of forecasts, h the forecast horizon and S the original Diebold and Mariano statistic. We compute the standard error of the differences in forecasts using the Newey-West estimator with the automatic lag-length selection of Andrews (1991).

two state Markov switching parameterization seems to be more successful for this data set.

The previous pattern is accentuated when we consider the 6 months ahead forecast horizon. We find that Model 2 achieves the smallest MSE for 15 maturities while Model 4 does it only for 2 maturities. This result suggests that imposing the non-arbitrage restrictions is relatively more important for the shortest forecast horizons. For example, when we consider the 12 months ahead forecasts, Model 4 is always outperformed. For this horizon, Models 2 and 3 achieve the smallest MSE for the shortest maturities, while Model 1 does it for the longer maturities. On the other hand, while Model 3 does not have the smallest MSE across maturities, it achieves the smallest squared errors in the majority of the forecast dates and for most maturities.

We also use tests of equal forecast accuracy that are designed to examine whether the MSE of two alternative non-nested models are significantly different from each other. The comparison is made between the linear Model 1 and the nonlinear alternatives using the modified Diebold and Mariano (1995) statistic proposed by Harvey et al., (1997). In line with the previous findings based on the MSE criterion, Models 4 and 5 outperform Model 1 for 1 month ahead forecasts, in particular for short and medium maturities. For the 3 months ahead forecasts, while Models 4 and 5 are significantly better than Model 1 at short maturities, Model 2 is significantly better than Model 1 for maturities between 9 and 21 months. The pattern is similar for the 6 months ahead forecast horizon, with the difference that Model 2 is significantly better than Model 1 for all maturities but one between 3 and 30 months. Finally, for the 12 months horizon, the nonlinear models are rarely significantly better than Model 1. Moreover, for maturities longer than 5 years, the switching model that imposes non-arbitrage restrictions (Model 4) forecasts significantly worse than Model 1.

Summarizing, we find that Model 5, with the best in-sample fit, has only good forecasting performance for the shortest forecast horizons. The ability of the model to match in-sample swings in the decay rate parameter λ_t leads to the usual problem of overfitting, reflected in the poor forecasting performance of the model at longer horizons. In addition, we find that imposing non-arbitrage restrictions in the Markov switching model is beneficial mostly for short forecast horizons. On the other hand, the Markov switching models with state dependent decay parameter λ have a relatively good forecasting performance for medium and long horizons. Overall, these results show that several nonlinear models have better forecasting performance than the single regime three factor model. These findings are particularly noteworthy because one of the major weaknesses of many existing nonlinear models is their relatively poor out-of-sample performance.

5 Conclusions

In this paper we propose several Markov switching extensions of the Diebold and Li (2006) dynamic Nelson and Siegel model and evaluate their merits relative to other extensions proposed in the literature. The extensions are motivated by the observation that the shape of the yield curve seems to change over time in ways that may be captured by a Markov switching framework. For example, the statistical properties of the yield curve depend on particular partitions of the sample, such as booms and recessions, or active and passive monetary policy regimes. Along this line, we document that the yield curve is substantially less persistent and flatter during recessions than in booms for the sample under consideration. In addition, we also observe that the yield curve is relatively flat at the beginning of recessions, interest rates drop afterwards, and, as the economy recovers, yields spread out and the yield curve becomes steeper.

We consider models that impose non-arbitrage restrictions and models that allow for time variation in the exponential decay rate parameter λ that governs the shape of the yield curve. We also derive a discrete-time version of the non-arbitrage restrictions associated with the Markov switching model. We show that, to be consistent with the dynamic Nelson and Siegel framework, the parameterization of the model cannot allow for switching in the decay rate parameter. However, the associated measurement equations include switching in a constant specific for each maturity. The proposed Markov switching models, with and without non-arbitrage restrictions, are parsimonious and relatively easy to estimate. This simplicity is accomplished by an approximate non-linear filter that collapses a growing mixture of densities to a single density, dramatically reducing the dimensionality of the estimation problem.

We compare the Markov switching models with the standard three-factor dynamic Nelson and Siegel model and with another single regime model that treats the decay rate parameter as a continuously time-varying unobserved component. We assess the importance, in terms of fit and forecasting performance, of each of these extensions: namely, the importance of allowing for changes in the decay rate parameter λ and of imposing non-arbitrage restrictions.

The merits of the different models are assessed in terms of their forecasting performance. The single-regime model, that treats λ as a continuously time-varying factor, performs very well in terms of fit and forecasting performance only at the shortest forecasting horizon. Yet, the substantial gains of this model in terms of fit are obtained against losses in mid and long horizons forecasts. Within the class of Markov switching models, imposing non-arbitrage restrictions is important only at short and medium horizons. On the other hand, models with switching in the decay rate parameter have a relatively good forecasting performance at medium and long horizon. Overall the paper shows that several models have better forecasting performance than the single regime dynamic Nelson and Siegel model. Which model is preferred, however, depends on the particular forecasting horizon.

References

Andrews, D. W. K. 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817-858.

Ang A, Piazzesi M. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50, 745-787.

Bandara W, Munclinger R. 2012. Term structure modeling with structural breaks: a simple arbitrage-free approach. The George Washington University Working Paper.

Banzal R, Zhou H. 2002. Term structure of interest rates with regime shifts. *Journal of Finance* 57, 1997-2043.

Bianchi F, Mumtaz H, Surico P. 2009. The great moderation of the term structure of UK interest rates. *Journal of Monetary Economics* 56, 856-871.

Bikbov R, Chernov M. 2013. Monetary policy regimes and the term structure of interest rates. *Journal of Econometrics* 174, 27-43.

Christensen J.H.E, Diebold F. X, Rudebusch G. D. 2011. The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics* 164, 4-20.

Coronoe L, Giannone D, Modugno M. 2013. Unspanned macroeconomic factors in the yield curve. Working paper.

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

Dai Q, Singleton K. 2000. Specification analysis of affine term structure models. *Journal of Finance* 55, 1943-78.

Diebold F.X, Li C. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337-64.

Diebold F.X, Mariano R.S. 1995. Comparing predictive accuracy. *Journal of Business, Economics, and Statistics* 13, 253-263.

Diebold F.X, Rudebusch G.D, Aruoba S.B. 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics* 131, 309-38.

Driffill J, Kenc T, Sola M, Spagnolo F. 2009. The effects of different parameterizations of Markov-Switching in a CIR model of bond pricing. *Studies in Nonlinear Dynamics & Econometrics* 13, 1-24.

Duffie D, Kan R. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6, 379-406.

Harrison P.J, Stevens C.F. 1976. Bayesian forecasting. *Journal of the Royal Statistical Society, Series B* 38, 207-247.

Harvey D. I, Leybourne S.J, Newbold P. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13, 281-291.

Kim C.J, Nelson C.R. 1999. *State-Space models with regime switching.* Cambridge, Massachusetts: MIT Press.

Kim C.J. 1994. Dynamic linear models with Markov-Switching. *Journal of Econometrics* 60, 1-22.

Knez P, Litterman R, Scheinkman J. 1994. Exploration into factors explaining money market returns. *Journal of Finance* 49, 1861-82.

Koopman S.J, Mallee M.I.P, van der Wel M. 2010. Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters. *Journal of Business and Economic Statistics* 28, 329-343.

Nelson C.R, Siegel A.F. 1987. Parsimonious modeling of yield curves. *Journal of Business* 60, 473-89.

Niu L, Zeng G. 2012. The discrete-time framework of arbitrage-free Nelson-Siegel class of term structure models. WISE Working Paper.

Piazzesi M. 2010. *Affine term structure models.* Handbook of Financial Econometrics, Elsevier.

Psaradakis Z, Spagnolo N. 2003. On the determination of number of regimes in Markov-switching autoregressive models. *Journal of Time Series Analysis* 24, 237- 252

Psaradakis Z, Spagnolo N. 2006. Joint determination of the state dimension and autoregressive order for models with Markov regime switching. *Journal of Time Series Analysis* 27, 753- 766.

West M, Harrison J. 1997. *Bayesian forecasting and dynamic models.* Springer-Verlag, New York.

Xiang J, Zhu X. 2013. A Regime-Switching Nelson-Siegel Terms Structure Model and Interest Rate Forecasts. *Journal of Financial Econometrics* 11, 522-555.

Yu W, Zivot E. 2011. Forecasting the term structures of treasury and corporate yields: dynamic Nelson-Siegel models evaluation. *International Journal of Forecasting* 27, 579-591.

Appendix: MS-DNS model with no-arbitrage restrictions

We consider an affine arbitrage-free model in discrete time following the set up in Ang and Piazessi (2003). The dynamics of the state variables Z_t follows, under the physical measure, the process

$$Z_t = \mu_{x_t} + F_{x_t} Z_{t-1} + \eta_{x_t}$$

where $x_t \in \{0, 1\}$ is the Markov regime and $\eta_{x_t} \sim N(0, H_{x_t})$. Payoffs in period $t+1$ are discounted using the following pricing kernel $M_{t+1}(x_{t+1}) = \exp\left(-r_t - \frac{1}{2}\Lambda'_{x_{t+1}}\Lambda_{x_{t+1}} - \Lambda'_{x_{t+1}}\eta_{x_{t+1}}\right)$, where $\Lambda_{x_t} = \Lambda^0_{x_t} + \Lambda^1_{x_t} Z_t$ are the time-varying market price of risk associated with the sources of uncertainty $\eta_{x_{t+1}}$, $r_t = \delta_{x_t} Z_t$ is the short rate equation that is assumed to be a function of Z_t, and δ_{x_t} is a state-dependent 1×3 vector to be defined.

The price of a zero-coupon bond at time t with maturity τ in regime $x_t = i$, denoted by $P^\tau_t(i)$, can be computed recursively according to

$$P^\tau_t(i) = \sum_{j=0}^{1} p_{ij} E_t\left[M_{t+1}(j) P^{\tau-1}_{t+1}(j) | i\right].$$

Using a guess and verify strategy, one can show that bond prices are given by $P^\tau_t(i) = \exp(A_i(\tau) + B'_i(\tau) Z_t)$, where, for $i \neq j$, the coefficients $A_i(\tau)$ and $B_i(\tau)$ follow the difference equations

$$\begin{aligned} A_i(\tau) &= p_{ii}(A_i(\tau-1) + \frac{B'_i(\tau-1) H_i B_i(\tau-1)}{2} + B'_i(\tau-1)(\mu_i - H_i\Lambda^0_i)) \\ &+ p_{ij}(A_j(\tau-1) + \frac{B'_j(\tau-1) H_j B_j(\tau-1)}{2} + B'_j(\tau-1)(\mu_j - H_j\Lambda^0_j)) \end{aligned}$$

$$B_i(\tau) = p_{ii}(-\delta'_i + (F_i - H_i\Lambda^1_i)' B_i(\tau-1)) + p_{ij}(-\delta'_i + (F_j - H_j\Lambda^1_j)' B_j(\tau-1)),$$

with initial conditions $A_i(0) = 0$ and $B_i(0) = [0, 0, 0]'$.

We construct an arbitrage-free MS-DNS model by extending Niu and Zeng (2012) to a Markov switching framework. As argued below, the arbitrage-free MS-DNS model can only be derived if the matrix $B_i(\tau)$ is regime independent, $B_i(\tau) = B(\tau)$.

Under the risk-neutral measure and imposing $B_i(\tau) = B(\tau)$, $\delta_i = \delta$, and $\Phi_i = \Phi$ we obtain

$$\begin{aligned} A_i(\tau) &= p_{ii}(A_i(\tau-1) + \frac{B'(\tau-1) H_i B(\tau-1)}{2} + B'(\tau-1)\zeta_i) \\ &+ p_{ij}(A_j(\tau-1) + \frac{B'(\tau-1) H_j B(\tau-1)}{2} + B'(\tau-1)\zeta_j) \end{aligned} \quad \text{(A1)}$$

$$B(\tau) = -\delta' + \Phi' B(\tau-1)), \quad \text{(A2)}$$

where $\zeta_i = \mu_i - H_i\Lambda^0_i$ and $\Phi = F_i - H_i\Lambda^1_i$. Under this assumption, we can find the underlying parameters of equations (A1) and (A2) such that the affine model has the same factor loadings of the Nelson and Siegel model. Clearly, the underlying parameters should be a function of λ.

By substitution, it is easy to show that the yields for the arbitrage-free MS-DNS model can be expressed as

$$-\frac{A_i(\tau)+B(\tau)'Z_t}{\tau} = -\frac{A_i(\tau)}{\tau} + \beta_{1t} + \beta_{2t}\frac{1-e^{-\lambda\tau}}{\lambda\tau} + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right),$$

with $Z_t = [\beta_{1t}, \beta_{2t}, \beta_{3t}]'$, $r_t = R_i(1) = \delta Z_t$, $\delta = [1, \frac{1-e^{-\lambda}}{\lambda}, \left(\frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda}\right)] = -B(1)'$, $B(\tau)' = [-\tau, -\frac{1-e^{-\lambda\tau}}{\lambda}, -\left(\frac{1-e^{-\lambda\tau}}{\lambda} - \tau e^{-\lambda\tau}\right)]$, and

$$\Phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & 0 \\ 0 & \lambda e^{-\lambda} & e^{-\lambda} \end{bmatrix}.^{17}$$

As in Christensen et al., (2011), the arbitrage-free MS-DNS model augments the Nelson and Siegel model with an additional state-dependent term $-A_i(\tau)/\tau$. Equations (A1) and (A2) imply that $A_i(\tau)$ cannot be zero which implies that the MS-DNS is incompatible with the arbitrage-free conditions. Moreover, the model does not impose any restriction on the parameters of the physical measure. In particular, given the estimated values for F_i and μ_i, we can recover the price of risk parameters Λ_i^0 and Λ_i^1 by solving $\Lambda_i^1 = H_i^{-1}(F_i - \Phi)$ and $\Lambda_i^0 = H_i^{-1}(\mu_i - \zeta_i)$. To identify the model, we follow Dai and Singleton (2000) in setting the constant of the affine short rate equation equal to zero and letting $\zeta_i = [\psi_i, 0, 0]'$ for $i = 0, 1$, to be estimated as free parameters along with the other parameters of the model. The arbitrage-free MS-DNS model has two additional parameters compared with the MS-DNS.

Derivation of equations (A1) and (A2).

The price of a zero-coupon bond with maturity τ in regime $x_t = i$ can be computed recursively from

$$P_t^{\tau}(i) = \sum_{j=0}^{1} p_{ij} E_t\left[M_{t+1}(j)P_{t+1}^{\tau-1}(j)|i\right].$$

Then, under the informational assumptions of Banzal and Zhou (2002),

$$\begin{aligned} E_t\left[M_{t+1}(0)P_{t+1}^{\tau-1}(0)|i\right] &= E_t\left[e^{\left(-\delta_i Z_t - \frac{1}{2}\Lambda'_{x_{t+1}=0}\Lambda_{x_{t+1}=0} - \Lambda'_{x_{t+1}=0}\eta_{x_{t+1}=0}\right)} e^{A_0(\tau-1)+B_0(\tau-1)'Z_{t+1}}\right] \\ E_t\left[M_{t+1}(1)P_{t+1}^{\tau-1}(1)|i\right] &= E_t\left[e^{\left(-\delta_i Z_t - \frac{1}{2}\Lambda'_{x_{t+1}=1}\Lambda_{x_{t+1}=1} - \Lambda'_{x_{t+1}=1}\eta_{x_{t+1}=1}\right)} e^{A_1(\tau-1)+B_1(\tau-1)'Z_{t+1}}\right] \end{aligned}$$

which allows us to express the pricing equation conditional on $x_t = i$ as

$$\begin{aligned} e^{A_i(\tau)+B_i'(\tau)Z_t} &= p_{ii} e^{\left(-\delta_i Z_t - \frac{1}{2}\Lambda'_{x_{t+1}=i}\Lambda_{x_{t+1}=i} + A_i(\tau-1)\right)} E_t\left[e^{-\Lambda'_{x_{t+1}=i}\eta_{x_{t+1}} + B_i(\tau-1)'Z_{t+1}}\right] \\ &\quad + p_{ij} e^{\left(-\delta_i Z_t - \frac{1}{2}\Lambda'_{x_{t+1}=j}\Lambda_{x_{t+1}=j} + A_j(\tau-1)\right)} E_t\left[e^{-\Lambda'_{x_{t+1}=j}\eta_{x_{t+1}} + B_j(\tau-1)'Z_{t+1}}\right] \end{aligned}$$

[17] To prove that B cannot be state contingent, we may start by assuming that the matrix B is state dependent. It then follows that the matrices Φ_i would depend on τ, which is inconsistent with the arbitrage-free model since the price of risk Λ_i would also depend on τ. More details are available upon request.

or

$$
\begin{aligned}
e^{A_i(\tau)+B_i'(\tau)Z_t} &= p_{ii}e^{(-\delta_i Z_t+A_i(\tau-1))}e^{B_i(\tau-1)'(\mu_i+F_iZ_t)}e^{\frac{B_i'(\tau-1)H_iB_i(\tau-1)}{2}}e^{-\Lambda'_{x_{t+1}=i}H_iB_i(\tau-1)} \\
&+p_{ij}e^{(-\delta_i Z_t+A_j(\tau-1))}e^{B_j(\tau-1)'(\mu_j+F_jZ_t)}e^{\frac{B_j'(\tau-1)H_jB_j(\tau-1)}{2}}e^{-\Lambda'_{x_{t+1}=j}H_jB_j(\tau-1)}
\end{aligned}
$$

Using the approximation $e^x \simeq 1+x$, imposing $B_i(\tau) = B(\tau)$, $\delta_i = \delta$, and substituting $\Lambda_{x_t=i} = \Lambda_i^0 + \Lambda_i^1 Z_t$ it follows that

$$
\begin{aligned}
A_i(\tau) &= p_{ii}(A_i(\tau-1) + \frac{B'(\tau-1)\,H_iB\,(\tau-1)}{2} + B'(\tau-1)\,(\mu_i - H_i\Lambda_i^0)) \\
&+p_{ij}(A_j(\tau-1) + \frac{B'(\tau-1)\,H_jB\,(\tau-1)}{2} + B'(\tau-1)\,(\mu_j - H_j\Lambda_j^0)),
\end{aligned}
$$

$$
B\,(\tau) = p_{ii}(-\delta' + (F_i - H_i\Lambda_i^1)'B\,(\tau-1)) + p_{ij}(-\delta' + (F_j - H_j\Lambda_j^1)'B\,(\tau-1)).
$$

Using that $\zeta_i = \mu_i - H_i\Lambda_i^0$ and $\Phi = F_i - H_i\Lambda_i^1$ we obtain the expressions in equations (A1) and (A2).

Table 1: Yield Estimation Results

	Model 1	Model 2		Model 3	
λ	0.0777	0.1253	0.0473	0.1274	0.0530
	(0.0021)	(0.0030)	(0.0015)	(0.0035)	(0.0019)
μ_1	0.0675	0.0855		0.0462	0.1225
	(0.0691)	(0.0697)		(0.0839)	(0.1416)
μ_2	0.1887	-0.0794		0.1102	-0.6930
	(0.1386)	(0.0423)		(0.0404)	(0.0870)
μ_3	-0.2220	-0.0310		0.0165	-0.0895
	(0.2015)	(0.0464)		(0.0745)	(0.0817)
$F(1,1)$	0.9957	0.9902		0.9950	0.9840
	(0.0177)	(0.0080)		(0.0097)	(0.0161)
$F(1,2)$	0.0285				
	(0.0211)				
$F(1,3)$	-0.0222				
	(0.0232)				
$F(2,1)$	-0.0306				
	(0.0260)				
$F(2,2)$	0.9389	0.9539		0.9958	0.7776
	(0.0309)	(0.0162)		(0.0067)	(0.0276)
$F(2,3)$	0.0393				
	(0.0084)				
$F(3,1)$	0.0242				
	(0.0113)				
$F(3,2)$	0.0229				
	(0.0305)				
$F(3,3)$	0.8438	0.7482		0.8276	0.7500
	(0.0186)	(0.0342)		(0.0513)	(0.0519)
$H(1,1)$	0.0947	0.1010		0.1096	
	(0.0084)	(0.0087)		(0.0093)	
$H(1,2)$	-0.0140				
	(0.0113)				
$H(1,3)$	0.0438				
	(0.0186)				
$H(2,1)$	-0.0140				
	(0.0113)				
$H(2,2)$	0.3822	0.3808		0.3390	
	(0.0305)	(0.0304)		(0.0260)	
$H(2,3)$	0.0094				
	(0.0344)				
$H(3,1)$	0.0438				
	(0.0186)				
$H(3,2)$	0.0094				
	(0.0344)				
$H(3,3)$	0.8007	1.0119		0.9669	
	(0.0813)	(0.0950)		(0.0933)	
p_{00}		0.9578		0.9311	
		(0.0097)		(0.0171)	
p_{11}		0.9003		0.9137	
		(0.0246)		(0.0223)	
Log likelihood	3173.5	3311.66		3343.81	
AIC	-6303.0	-6597.3		-6649.6	
BIC	-6218.3	-6547.2		-6576.4	

Table 2: Yield Estimation Results (Continuation)

	Model 4		Model 5
λ	0.0888		
	(0.0126)		
μ_1	0.0117	1.0345	-0.1700
	(0.0643)	(0.3105)	(0.0547)
μ_2	0.0851	-0.7380	0.0754
	(0.0399)	(0.1603)	(0.0707)
μ_3	0.1476	1.0202	0.0092
	(0.0902)	(0.3001)	(0.0066)
μ_λ			-0.1099
			(0.0536)
$F(1,1)$	0.99	0.8463	0.9908
	(0.0001)	(0.0313)	(0.0080)
$F(2,2)$	0.99	0.4796	0.9758
	(0.0158)	(0.0649)	(0.0115)
$F(3,3)$	0.8921	0.5317	0.8455
	(0.0390)	(0.0792)	(0.0313)
$F(4,4)$			0.9339
			(0.0212)
$H(1,1)$	0.09		0.1027
	(0.08)		(0.0090)
$H(2,2)$	0.23		0.3955
	(0.13)		(0.0333)
$H(3,3)$	0.86		0.8993
	(0.17)		(0.0891)
$H(4,4)$			0.4257
			(0.3978)
ϕ_1	0.9488		
	(0.4121)		
ϕ_2	0.4516		
	(0.5262)		
p_{00}	0.9488		
	(0.4121)		
p_{11}	0.4516		
	(0.5262)		
Log likelihood	3467.36		3620.35
AIC	-6904.7		-7210.7
BIC	-6846.9		-7152.9

Table 3: Descriptive statistics of level, slope, and curvature

A. Data	All sample			Recessions			Booms		
	Mean	Std. dev.	$\hat{\rho}(1)$	Mean	Std. dev.	$\hat{\rho}(1)$	Mean	Std. dev.	$\hat{\rho}(1)$
Level	8.14	2.17	0.98	9.91	0.71	0.59	7.87	0.88	0.81
Slope	1.29	1.46	0.93	0.66	1.33	0.49	1.39	1.31	0.83
Curvature	0.12	0.72	0.79	0.42	0.76	0.33	0.08	0.60	0.62
B. Model 3				Regime 0			Regime 1		
				Mean	Std. dev.	$\hat{\rho}(1)$	Mean	Std. dev.	$\hat{\rho}(1)$
Level				8.65	0.61	0.63	7.79	0.57	0.65
Slope				0.74	0.81	0.64	1.99	0.49	0.60
Curvature				0.33	0.51	0.49	-0.01	0.34	0.46

Note: $\hat{\rho}(1)$ denotes the first order sample autocorrelation.

Table 4: Forecasting 1 month ahead

	MSE: 1 month ahead					MSE percentage of times				
τ	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
3	0.0396	0.0282^{x}	0.0408	0.0284^{x}	**0.0270***	29.8	27.4	11.9	15.5	15.5
6	0.0397	0.0341	0.0431	**0.0319***	0.0344^{x}	25.0	31.0	20.2	13.1	10.7
9	0.0487	**0.0410***	0.0506	0.0440	0.0456$^{\diamond}$	6.0	38.1	26.2	21.4	8.3
12	0.0534	0.0506	0.0631°	**0.0468**	0.0495	9.5	31.0	19.0	31.0	9.5
15	0.0617	0.0617	0.0742°	**0.0523**^{x}	0.0558$^{\diamond}$	8.3	22.6	20.2	39.3	9.5
18	0.0658	0.0644	0.0762°	**0.0575**^{x}	0.0601$^{\diamond}$	10.7	21.4	26.2	34.5	7.1
21	0.0720	0.0692	0.0803$^{\triangleright}$	**0.0639**$^{\diamond}$	0.0662$^{\diamond}$	15.5	20.2	27.4	33.3	3.6
24	0.0772	0.0731$^{\diamond}$	0.0820	**0.0708**	0.0722	13.1	25.0	29.8	29.8	2.4
30	0.0746	0.0719	0.0792	**0.0681**$^{\diamond}$	0.0698^{x}	16.7	28.6	32.1	22.6	0.0
36	0.0734	0.0719	0.0764	**0.0670**^{x}	0.0688^{x}	13.1	27.4	31.0	20.2	8.3
48	0.0755	0.0745	0.0773	**0.0694**^{x}	0.0709^{x}	13.1	16.7	34.5	25.0	10.7
60	0.0798	0.0784	0.0783	**0.0746**$^{\diamond}$	0.0746^{x}	21.4	14.3	29.8	25.0	9.5
72	0.0720	0.0716	0.0723	0.0709	**0.0692**	19.0	14.3	25.0	34.5	7.1
84	0.0752	0.0739	0.0751	0.0746	**0.0720**	16.7	16.7	19.0	44.0	3.6
96	0.0702	0.0686	0.0705	0.0684	**0.0654**	17.9	16.7	16.7	41.7	7.1
108	0.0714	0.0694	0.0712	0.0707	**0.0671**	22.6	15.5	19.0	17.9	25.0
120	0.0720	**0.0715**	0.0727	0.0728	0.0740	28.6	14.3	16.7	8.3	32.1

Note: $\diamond$, x and $*$ are 10, 5 and 1% significance levels for the one sided modified Diebold and Mariano test, respectively. The null hypothesis is that the non-linear model outperforms the linear one. $\triangleright$ and $\circ$ are 10 and 5% significance levels for the linear model outperforming the non-linear one, respectively.

Table 5: Forecasting 3 months ahead

	MSE: 3 months ahead					MSE percentage of times				
τ	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
3	0.1428	0.0932	0.1547	**0.0872**^{x}	0.0943^x	20.2	34.5	15.5	13.1	16.7
6	0.1726	0.1345	0.1861	**0.1294***	0.1428*	21.4	31.0	25.0	15.5	7.1
9	0.2053	**0.1607**^{x}	0.2131	0.1687^x	0.1788*	10.7	29.8	23.8	27.4	8.3
12	0.2185	0.1909	0.2498	**0.1838**$^\diamond$	0.1943$^\diamond$	9.5	27.4	26.2	32.1	4.8
15	0.2430	0.2230	0.2830	**0.2069**	0.2173	11.9	20.2	27.4	34.5	6.0
18	0.2616	0.2352$^\diamond$	0.2899	**0.2261**	0.2339	15.5	23.8	23.8	33.3	3.6
21	0.2808	0.2488^x	0.2986	**0.2456**	0.2514	15.5	23.8	21.4	34.5	4.8
24	0.2948	**0.2585**	0.3019	0.2636	0.2660	16.7	22.6	22.6	34.5	3.6
30	0.2873	**0.2569**	0.2919	0.2583	0.2604	17.9	22.6	22.6	34.5	2.4
36	0.2848	0.2592	0.2857	**0.2584**	0.2603	20.2	21.4	25.0	32.1	1.2
48	0.2810	0.2606	0.2790	**0.2586**	0.2593	23.8	22.6	20.2	31.0	2.4
60	0.2877	**0.2697**	0.2800	0.2755	0.2698	25.0	21.4	23.8	25.0	4.8
72	0.2623	**0.2496**	0.2593	0.2621	0.2511	27.4	25.0	13.1	28.6	6.0
84	0.2655	**0.2499**	0.2613	0.2714	0.2554	23.8	28.6	11.9	31.0	4.8
96	0.2446	**0.2283**	0.2393	0.2513	0.2340	25.0	25.0	13.1	35.7	1.2
108	0.2459	**0.2276**	0.2385	0.2543	0.2380	25.0	22.6	14.3	26.2	11.9
120	0.2442	**0.2288**	0.2386	0.2524	0.2464	31.0	21.4	13.1	11.9	22.6

Note: see note in Table 4.

Table 6: Forecasting 6 months ahead

	MSE: 6 months ahead					MSE percentage of times				
τ	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
3	0.3803	0.2138$^\diamond$	0.3253	**0.2092***	0.2526^x	19.0	36.9	16.7	20.2	7.1
6	0.4159	**0.2827**$^\diamond$	0.3727	0.2880*	0.3320*	17.9	38.1	19.0	17.9	7.1
9	0.4518	**0.3184**^{x}	0.4027	0.3494^x	0.3822	8.3	32.1	22.6	32.1	4.8
12	0.4675	**0.3740**$^\diamond$	0.4668	0.3855$^\diamond$	0.4122$^\diamond$	14.3	32.1	20.2	28.6	4.8
15	0.4886	**0.4139**	0.5080	**0.4139**	0.4371	15.5	29.8	21.4	28.6	4.8
18	0.5207	**0.4381**$^\diamond$	0.5212	0.4523	0.4677	14.3	28.6	22.6	27.4	7.1
21	0.5521	**0.4637**$^\diamond$	0.5383	0.4901	0.4983	15.5	28.6	22.6	27.4	6.0
24	0.5790	**0.4860**$^\diamond$	0.5498	0.5266	0.5264	16.7	26.2	23.8	27.4	6.0
30	0.5695	**0.4912**$^\diamond$	0.5412	0.5263	0.5224	20.2	27.4	23.8	27.4	1.2
36	0.5710	**0.5052**	0.5434	0.5381	0.5310	21.4	27.4	23.8	26.2	1.2
48	0.5715	**0.5224**	0.5481	0.5558	0.5428	25.0	28.6	20.2	23.8	2.4
60	0.6022	**0.5633**	0.5806	0.6152	0.5878	28.6	23.8	20.2	23.8	3.6
72	0.5561	**0.5302**	0.5458	0.5935	0.5556	31.0	22.6	17.9	26.2	2.4
84	0.5579	**0.5333**	0.5514	0.6132	0.5645	32.1	25.0	14.3	28.6	0.0
96	0.5242	**0.4981**	0.5166	0.5832	0.5322	31.0	25.0	15.5	26.2	2.4
108	0.5255	**0.4938**	0.5137	0.5846	0.5380	33.3	26.2	13.1	22.6	4.8
120	0.5379	**0.5070**	0.5269	0.5905	0.5627	36.9	23.8	16.7	11.9	10.7

Note: see note in Table 4.

Table 7: Forecasting 12 months ahead

	MSE: 12 months ahead					MSE percentage of times				
τ	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
3	1.0213	**0.6069**	0.7638	0.6789$^{\diamond}$	0.7989$^{\diamond}$	25.0	19.0	40.5	10.7	4.8
6	0.9995	**0.7087**	0.8321	0.8052$^{\diamond}$	0.8713	26.2	17.8	42.9	11.9	1.2
9	0.9976	**0.7603**	0.8592	0.8943	0.9129	16.7	20.2	42.9	17.9	2.4
12	0.9502	**0.8129**	0.9058	0.9166	0.9074	15.5	22.6	39.3	20.2	2.4
15	0.9344	**0.8601**	0.9486	0.9395	0.9175	17.9	20.2	39.3	19.0	3.6
18	0.9647	**0.9018**	0.9688	0.9961	0.9573	20.2	22.6	35.7	19.0	2.4
21	0.9950	**0.9389**	0.9866	1.0444	0.9936	20.2	23.8	38.1	14.3	3.6
24	1.0338	**0.9802**	1.0078	1.1017	1.0387	19.0	23.8	38.1	15.5	3.6
30	1.0244	**0.9993**	1.0033	1.1070	1.0441	19.0	22.6	38.1	14.3	6.0
36	1.0305	1.0254	**1.0130**	1.1269	1.0646	23.8	19.0	36.9	14.3	6.0
48	1.0320	1.0524	**1.0257**	1.1566	1.0911	23.8	20.2	38.1	15.5	2.4
60	**1.0985**	1.1362	1.1005	1.2788°	1.1886	29.8	20.2	33.3	13.1	3.6
72	**1.0296**	1.0842	1.0492	1.2512°	1.1414	29.8	25.0	27.4	14.3	3.6
84	**1.0123**	1.0678	1.0400	1.2687°	1.1382$^{\triangleright}$	33.3	27.4	22.6	15.5	1.2
96	**0.9703**	1.0137	0.9918	1.2287°	1.0937	38.1	27.4	17.9	11.9	4.8
108	**0.9811**	1.0087	0.9930	1.2281°	1.1049	38.1	27.4	17.9	10.7	6.0
120	**1.0405**	1.0613	1.0516	1.2599°	1.1764	38.1	28.6	20.2	8.3	4.8

Note: see note in Table 4.

Figure 1: Yields, NBER recessions, and smoothed probability of regime 0

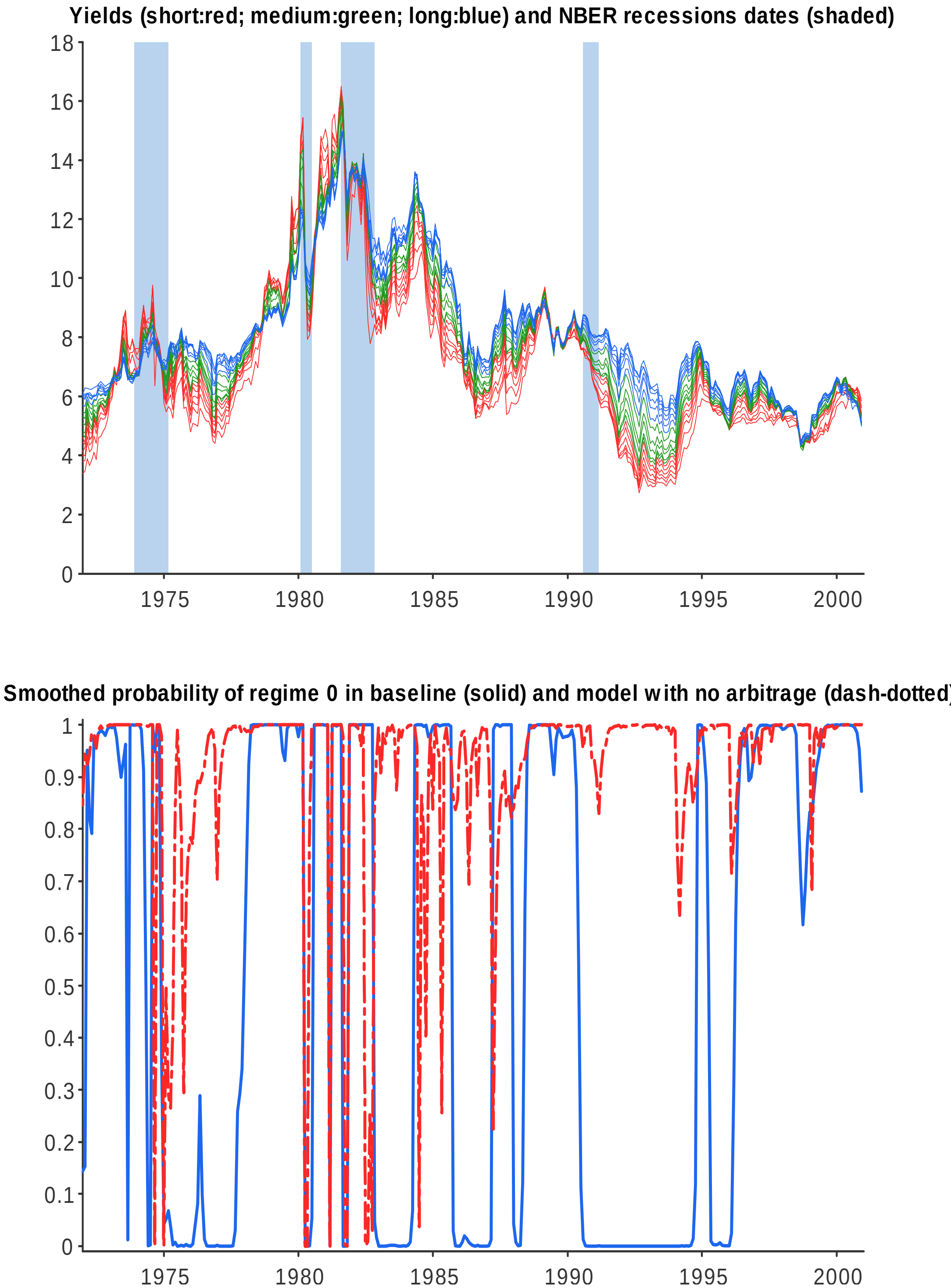

Figure 2: Factor loadings

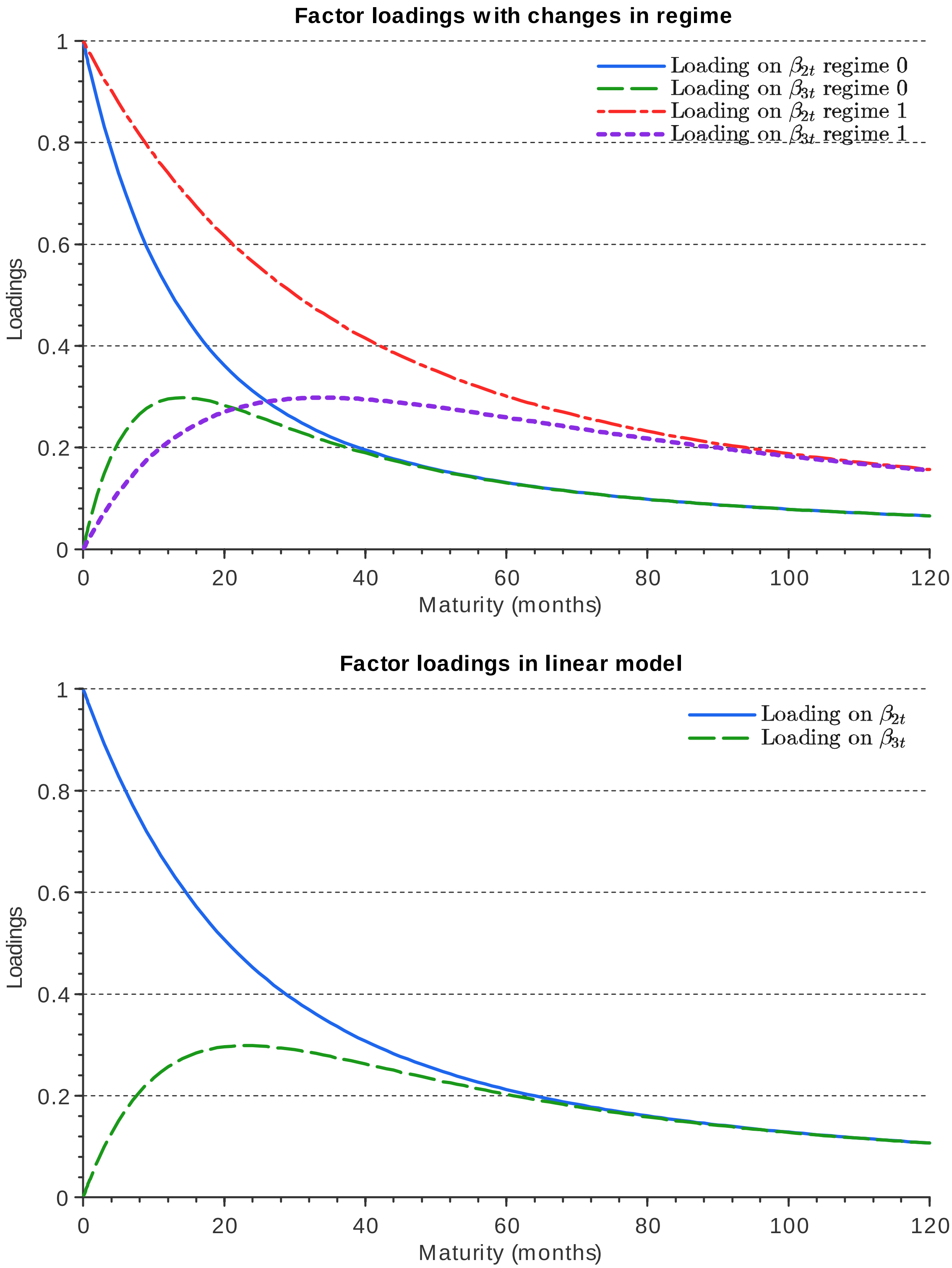

Figure 3: Stylized facts in booms and recessions and in baseline model