

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

^aWorld Bank, Washington DC, USA

^bSchool of Economics, Mathematics & Statistics, Birkbeck College, London, UK

^cDepartment of Economics, Universidad Torcuato Di Tella, Buenos Aires, Argentina

^dDepartment of Economics and Finance, Brunel University, Uxbridge, UK

July 2012

Abstract

In this paper we estimate the yield curve of U.S. government bonds using a Markov switching latent variable model. We show how measures such as the level, slope, and curvature of the yield curve are affected by business cycle conditions. We present a switching latent model which not only seem to capture this features in sample but also performs well out of sample.

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

JEL classification: C13; C22; E43.

**E-mail addresses:* chevia@utdt.edu (C. Hevia), mrozada@utdt.edu (M. Gonzalez-Rozada), m.sola@bbk.ac.uk (M. Sola), and fabio.spagnolo@brunel.ac.uk (F. Spagnolo). The views expressed in this paper are those of the authors, and do not necessarily reflect those of the World Bank, their Boards of Directors, or the countries they represent. Constantino Hevia (chevia@utdt.edu) is the corresponding author. Address: Universidad Torcuato Di Tella, Saenz Valiente 1010, Buenos Aires, Argentina. Phone: 5411 5169 7300. Fax: 5411 5169 7348

1 Introduction

Understanding the dynamics and cross section of bond yields is important for issues as disparate as macroeconomic forecasting, monetary policy, debt policy, portfolio risk management, among others. Current research on the yield curve assumes that a handful of unobserved factors, often given labels such as “level,” “slope,” and “curvature,” describe the entire term structure of interest rates (Knez, Litterman, and Scheinkman, 1994; Duffie and Kan, 1996; Dai and Singleton, 2000; Diebold and Li, 2006). Broadly speaking, the literature evolved into two related but not identical branches: the class of arbitrage-free affine term structure models (Piazzesi, 2010) and the class of dynamic Nelson and Siegel (1987) models as proposed by Diebold and Li (2006). For the most part, the literature has assumed that latent factors follow a stochastic linear evolution equation. This paper contributes to the literature on yield modeling by extending the three factor model of Diebold and Li (2006) with a fourth latent variable that evolves according to a two state Markov chain. We call this model the Markov switching dynamic Nelson and Siegel (MSDNS) model. This extension makes the model nonlinear and is intended to capture the economic conditions, such as the phases of the business cycle, that typically affect the shape and other characteristics of the yield curve.

Diebold and Li (2006) argue that a good model of the yield curve dynamics should be able to reproduce the historical stylized facts concerning the average shape of the yield curve, the variety of shapes assumed at different times, the strong persistence of yields, and the weak persistence of spreads. In this paper, we argue that the behavior of the yield curve differs in the different stages of the business cycle and, therefore, that this has to be taken into account when considering what can be thought of as the stylized facts. Moreover, once it is accepted that the yield curve changes over the business cycle, it is then natural to use the MSDNS framework to model the evolution of the yield curve. This extension is important because the estimated model can then be used for forecasting interest rates, portfolio risk management, and so forth.

Diebold and Li (2006) model the yield curve using a linear state-space model where bond yields are linear functions of three latent variables that are interpreted as level, slope, and curvature factors. Latent variables evolve as a vector autoregression of order one and observed yields are a linear function of them. Koopman, Mallee, and Van der Wel (2010) extend Diebold and Li (2006) and allow time variation not only in the level, slope, and curvature factors but also in the loading

on these factors. Their model includes a fourth unobserved component that affects the loading on the other three factors. Moreover, they also account for possible changes in the volatility of the yield curve by including a GARCH term multiplying the covariance matrix of the state-space representation of their model. The model we propose in this paper also allows for possible changes in factor loadings and in the volatility of the yield curve, although through a different channel: the realized value of the unobserved Markov chain determines the parameters of the state equation describing the evolution of the latent variables and the parameters of the observation equation relating latent variables to observed yields.

The interaction between the yield curve and macroeconomic factors has also received attention in recent years. Ang and Piazzesi (2003) and Diebold, Rudebusch, and Aruoba (2006) developed factor models of the yield curve that explicitly incorporate macroeconomic factors into the equations that describe the evolution of the latent variables. They show that this extended framework is useful to explain the main characteristics of the yield curve for U.S. government bonds. Our modeling strategy is flexible enough to account for (at least some of) the macroeconomic conditions that affect the yield curve without the need of directly including those variables.

One possible criticism of the Nelson and Siegel approach is that, in general, it does not rule out arbitrage opportunities. To address this issue, Christensen, Diebold, and Rudebusch (2011) derive arbitrage free conditions for the Nelson and Siegel model and evaluate their potential improvement in the forecasting of the yield curve. In our empirical work, however, we follow the vast majority of the literature and do not impose the no-arbitrage representation of the Nelson and Siegel model.¹ In this respect, it is important to note that if actual bond yields do not admit arbitrage opportunities (as one would expect given the highly developed and liquid market for U.S. government bonds), fitting a Nelson and Siegel model to that arbitrage-free data will presumably generate a fitted yield curve that is approximately arbitrage-free. As Diebold and Rudebusch (2012) put it,

“...a model may admit arbitrage yet provide a good approximation to a much more complicated (perhaps arbitrage-free) reality and hence forecast well. Moreover, if reality is arbitrage free, and if a model provides a very good description of reality, then imposition of no-arbitrage would presumably have little effect. An accurate model would be

¹This is, for example, the approach followed by Diebold and Li (2006), Diebold, Rudebusch, and Aruoba (2006), and Koopman, Mallee, and Van der Wel (2010).

at least approximately arbitrage-free, even if freedom from arbitrage were not explicitly imposed.”

This statement is supported by the empirical findings of Coroneo, Nyholm, and Vidova-Koleva (2011). Using U.S. Treasury zero-coupon yield data, the authors argue that the Nelson and Siegel model is compatible with the no-arbitrage constraints by showing that: i) the hypothesis that the Nelson and Siegel factor loadings fulfill the no-arbitrage constraints cannot be rejected (i.e. the factor loadings from the model are not statistically different from those derived from an arbitrage-free model), and ii) the Nelson and Siegel model performs as well as its no-arbitrage counterpart in out-of-sample forecasting. On the basis of these results, the authors argue that it is not necessary to impose explicit no-arbitrage constraints on the Nelson–Siegel factor loadings when such a model is estimated on arbitrage-free data, as it is likely to be the case for U.S. bond yields (at least for the sample period used). To corroborate these findings, their Monte Carlo results show that, if the no-arbitrage restrictions do not hold in the data, then the tested hypothesis is rejected. Of course, extending our modeling framework to include the no-arbitrage restriction is not only possible but also desirable. Given the arguments outlined before, however, we leave this extension for future research as it would be outside the scope of the present paper.

Based on US zero-coupon data, all our estimated switching models present significant evidence of regime shifts, with the estimated transition probabilities implying that the regimes are highly persistent. Most importantly, our empirical results show that the conventional stylized facts of the yield curve are typically associated with the boom periods. Furthermore, our proposed models seem to not only successfully characterize the data under scrutiny but also, and perhaps more importantly, to have a good forecasting performance. The forecasting results are particularly noteworthy because one of the perceived weaknesses of nonlinear models is their relatively poor out-of-sample performance. We also discuss the economic relevance of our model for portfolio risk management by showing that ignoring the regime specific nature of the model gives an incorrect coverage of the risk associated with changes in interest rates.

The paper is organized as follows. In Section 2, we present some stylized facts that motivate the use of our model. Section 3 presents the basic Markov switching framework for modeling the yield curve and its economic relevance using an example of portfolio risk management. Section

4 describes the econometric model and an approximate filtering algorithm used to evaluate the likelihood function of the nonlinear model. In Section 5 we apply the model using U.S. data on government bond yields and in Section 6 we assess the out-of-sample performance of the model. Section 7 concludes.

2 Stylized facts of the yield curve: Are booms and recessions important?

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, using the un-smoothed Fama and Bliss (1987) approach, as described in Diebold and Li (2006).² We also use NBER dates to separate booms from recessions. A recession is defined as starting the month after the economy reaches a peak of economic activity and ending the month when activity reaches its trough. During the remaining periods, the economy is said to be in a boom. With these definitions, our sample period contains four recessions that last about one year on average, and five booms that last about five years on average.

Figure 1 displays a number of statistics over the entire sample and separating recession from boom episodes. The top panel displays average yields across maturities. Three things follow from this plot. First, while the yield curve is on average about 35 percent greater in recessions than in booms, the difference decreases as yield maturity increases. Second, except at short maturities, the average yield curve is flatter in recessions than in booms. And third, the average yield curve over the entire sample has a similar shape to that during booms. The bottom left panel of the figure shows the standard deviation of yields across maturities. We compute the standard deviation of yields in each boom (recession) episode and then average over all boom (recession) episodes. (We do the same below with the autocorrelation.) Yields are more volatile in recessions than in booms at short maturities but the converse holds at long maturities. In both cases, however, long rates are less volatile than short rates. Also, the volatility of yields ignoring the state of the economy is substantially greater than that during both booms and recessions.³ Finally, the lower right panel of

²We obtained the data from Professor Diebold's website.

³The reason for this difference is that taking the standard deviation over the entire sample ignores the discrete changes in the level of yields that occur when the economy moves back and forth between booms and recessions.

Figure 1 displays the persistence of yields, defined as their first order sample autocorrelation. It is apparent that yields are substantially more persistent in booms than in recessions and that, within regimes, persistence is similar across maturities. In addition, persistence in booms and recessions is substantially smaller than the estimated persistence ignoring changes in economic activity: while the latter is about 0.98 for all maturities, the cross sectional average persistence is 0.81 in booms and 0.64 in recessions. This finding suggests that the high persistence of yields documented in the literature (e.g. Diebold and Li, 2006) is, to a large extent, driven by discrete changes in the level of interest rates as the economy transits between booms and recessions. Because booms and (to a lesser extent) recessions last several months, these persistent discrete changes in bond yields increase their unconditional autocorrelation.

Figure 1 about here

In panel A of Table 1 we report some statistics on measures of level, slope, and curvature for booms, recessions, and the entire sample. As in Diebold and Li (2006), we define the level as the 10 year yield, the slope as the spread between the 10 year and the 3 month yields, and the curvature as the spread between the 2 year and 3 month yields minus the spread between the 10 year and 2 year yields. The slope in booms is over 100 percent greater than in recessions (1.39 versus 0.66) and is somewhat larger than the average slope in all sample (1.29). In addition, while the volatility of the slope is similar in booms and recessions, the persistence of the slope is also substantially larger in booms than in recessions. As in the case of yields, the discrete changes in the slope between booms and recessions induce more persistence to the slope over the entire sample. Finally, the curvature of the yield curve is over four times greater in recessions than in booms, but is twice as persistent in booms than in recessions. Note, as Figure 1 suggests, that the much greater curvature of the yield curve in recessions vis-a-vis booms is driven by the large spread between the 15 month and 3 month yields. For yields greater than 15 months, the yield curve during recessions is virtually flat.

Table 1 about here

In sum, the above findings suggest that the dynamic properties of the yield curve differ markedly between booms and recessions. A model that does not allow for discrete changes in some variable

These breaks in the level of interest rates show up as a higher overall yield volatility.

(like our Markov regime indicator variable) is likely to run into difficulties when trying to match the different shapes of the yield curve over the business cycle. The model we propose below is flexible enough to be able to account, at least in principle, for the discrete changing patterns of the yield curve.

3 Modeling the yield curve

Practitioners use various functional forms to model the yield curve. One of the most popular models is that of Nelson and Siegel (1987) and its dynamic version proposed by Diebold and Li (2006). Diebold and Li consider a parametric model of the yield curve at some time t of the form

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

where $R_t(\tau)$ represents the zero-coupon yield on a bond that matures in τ months, $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ are time-varying factors, and λ is a constant associated with the exponential decay rate of the factor loadings at different bond maturities.⁴ Even though the four parameters of model (1) could be estimated using non-linear least squares, the usual approach is to fix λ at some value and to estimate the factors $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ using ordinary least squares at each time t .

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. In particular, using Diebold and Li's preferred value of $\lambda = 0.0609$, it follows that $R_t(\infty) = \beta_{1t}$, $R_t(120) - R_t(3) = -0.78\beta_{2t} + 0.06\beta_{3t}$, and $2R_t(24) - R_t(3) - R_t(120) = -0.0004\beta_{2t} + 0.37\beta_{3t}$. That is, the level of the yield curve is determined by β_{1t} , the slope is closely related to β_{2t} , and the curvature is almost completely determined by β_{3t} . However, this interpretation of the latent factors, in particular that of β_{3t} , depends on the value of λ . The relation between the level of interest rates and β_{1t} holds for any value of λ . The relation between the slope of the yield curve and β_{2t} is also robust: the equality $R_t(\infty) - R_t(0) = -\beta_{2t}$ holds for any value of λ , and we can think of $R_t(\infty) - R_t(0)$ as an alternative definition of slope of the yield curve (Diebold and Li, 2006). The relation between the curvature of the yield curve and β_{3t} , however, is less robust and

⁴Koopman, Mallee, and Van der Wel (2010) also allow the parameter λ to be time-varying.

depends on the value of λ ; for some values of λ , like those we estimate below, the curvature factor is a non-trivial linear combination of β_{2t} and β_{3t} .

Independently of the actual interpretation of the latent factors, in this paper we propose to model the yield curve according to

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

where x_t is a regime-indicator variable taking the values 0 or 1,

$$R_{jt}(\tau) = \beta_{1t} + \beta_{2t} \frac{1 - e^{-\lambda_j \tau}}{\lambda_j \tau} + \beta_{3t} \left(\frac{1 - e^{-\lambda_j \tau}}{\lambda_j \tau} - e^{-\lambda_j \tau} \right) \quad (2)$$

is the yield curve conditional on regime $j = 0, 1$, and $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ are dynamic latent factors whose probability distribution may also depend on the regime-indicator variable x_t . We also assume that nature selects regime at time t with a probability that depends on what regime nature was in at time $t - 1$, so that $\{x_t\}$ form a time-homogeneous, first-order Markov chain with state space $\{0, 1\}$ and transition probabilities

$$p_{00} = \Pr(x_{t+1} = 0 | x_t = 0) \text{ and } p_{11} = \Pr(x_{t+1} = 1 | x_t = 1). \quad (3)$$

The key feature of the model is that the yield curve changes by a variable that can be interpreted as capturing business cycle conditions. For example, as it is explained above, it is a fact that the slope of the yield curve is different over the cycle. This parameterization is intended to capture those effects.⁵

3.1 Economic relevance of the proposed model: portfolio risk management

The Nelson and Siegel model is popular for portfolio risk management because it collapses the risk associated with any change in the yield curve into the risk associated with only three factors. Therefore, a portfolio of coupon paying bonds can be immunized of the capital loss and reinvesting risks associated with changes in the yield curve by using only three bonds/portfolios. In partic-

⁵Svensson (1994) extends the Nelson and Siegel model by adding to (1) a term of the form $\beta_{4t} \left(1 - e^{-\hat{\lambda}\tau} / \hat{\lambda}\tau - e^{-\hat{\lambda}\tau} \right)$ that is intended to give more flexibility to the model. The idea of model (2) is to give more flexibility to the Nelson and Siegel model in the same spirit of the extended Svensson's model.

ular, an arbitrary bond (or portfolio of bonds) can be hedged for capital and reinvesting risk by constructing a new portfolio containing the original portfolio and three additional bonds. For each dollar invested in the original portfolio, the extended portfolio invest ϕ_t^z dollars in bond $z = 1, 2, 3$. Because the yield curve is assumed to be generated by model (1), the change in the price of any bond due to arbitrary changes in the entire yield curve are a consequence of movements in just three latent factors. In Appendix A we show that hedging the extended portfolio amounts to choose a vector, $(\phi_t^1, \phi_t^2, \phi_t^3)'$ that makes the change in the value of the extended portfolio equal to zero for any possible (small) change in the three latent factors.

In contrast, the dynamic Nelson and Siegel model with Markov switching has four sources of risk: the risk associated with changes in the latent factors $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ conditional on the Markov regime x_t , and the risk associated with regime switches. It is then natural to conjecture that to hedge the four sources of risks, it is now necessary to construct an extended portfolio with four additional bonds. In Appendix A we show that, indeed, one can construct a portfolio that hedges all (small) changes associated with the three risk factors $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ and the discrete risk associated with changes in regime. It then follows that ignoring the regime specific nature of the model and using only three additional bonds for hedging purposes gives an incorrect coverage of the risk associated with changes in interest rates.

4 The econometric model

In this section we present an econometric model which accounts for the existence of different regimes when estimating the yield curve.⁶ The model postulates the existence of an unobserved discrete variable, $x_t \in \{0, 1, \dots, K\}$. This variable indexes the current regime and follows a Markov chain with transition probabilities $p_{ij} = \Pr(x_t = j | x_{t-1} = i)$ for $i, j = 0, 1, \dots, K$, where $p_{ij} \geq 0$ and $\sum_{j=0}^K p_{ij} = 1$ for all i . 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

⁶This model has different names in the literature: *Multi-process model* (Harrison and Stevens, 1976; West and Harrison, 1997), *Dynamic linear model with Markov-switching* (Kim, 1994), *Switching Kalman filters* (Murphy, 1998), *State-space models with Markov-switching* (Kim and Nelson, 1999), and *Discrete time Markov jump linear system* (Costa, Fragoso, and Marques, 2005),

and state equations are given by

$$y_t = \Lambda_{x_t} f_t + \varepsilon_{x_t t} \quad (4)$$

$$f_t = \mu_{x_t} + A_{x_t} f_{t-1} + \eta_{x_t t}. \quad (5)$$

Here, $y_t \in \mathbb{R}^m$ is a vector of observed variables, $f_t \in \mathbb{R}^n$ is a vector of unobserved continuous state variables, $\varepsilon_{x_t t} \in \mathbb{R}^m$ is normally distributed with mean zero and $m \times m$ covariance matrix Q_{x_t} ; $\mu_{x_t} \in \mathbb{R}^n$; A_{x_t} is an $n \times n$ matrix; and $\eta_{x_t t} \in \mathbb{R}^n$ 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 \hat{f}_0 and $n \times n$ covariance matrix V_0 .

Our model of the yield curve is a special case of the model presented above with $K = 1$ and $n = 3$. In particular, given a regime $x_t = j$, let $\tau = \tau_1, \tau_2, \dots, \tau_m$ denote m different observed maturities of zero-coupon yields, $R_t = (R_t(\tau_1), R_t(\tau_2), \dots, R_t(\tau_m))'$ the vector of observed variables, and $f_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$ the vector of unobserved continuous state variables. The observation equation (4) is given by

$$\begin{pmatrix} R_t(\tau_1) \\ R_t(\tau_2) \\ \vdots \\ R_t(\tau_m) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda_j \tau_1}}{\lambda_j \tau_1} & \frac{1-e^{-\lambda_j \tau_1}}{\lambda_j \tau_1} - e^{-\lambda_j \tau_1} \\ 1 & \frac{1-e^{-\lambda_j \tau_2}}{\lambda_j \tau_2} & \frac{1-e^{-\lambda_j \tau_2}}{\lambda_j \tau_2} - e^{-\lambda_j \tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_j \tau_m}}{\lambda_j \tau_m} & \frac{1-e^{-\lambda_j \tau_m}}{\lambda_j \tau_m} - e^{-\lambda_j \tau_m} \end{pmatrix} \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{jt}^1 \\ \varepsilon_{jt}^2 \\ \vdots \\ \varepsilon_{jt}^m \end{pmatrix} \quad (6)$$

and the state equation (5), by

$$\begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} = \begin{pmatrix} \mu_j^1 \\ \mu_j^2 \\ \mu_j^3 \end{pmatrix} + \begin{pmatrix} A_j(1,1) & A_j(1,2) & A_j(1,3) \\ A_j(2,1) & A_j(2,2) & A_j(2,3) \\ A_j(3,1) & A_j(3,2) & A_j(3,3) \end{pmatrix} \begin{pmatrix} \beta_{1t-1} \\ \beta_{2t-1} \\ \beta_{3t-1} \end{pmatrix} + \begin{pmatrix} \eta_{jt}^1 \\ \eta_{jt}^2 \\ \eta_{jt}^3 \end{pmatrix}. \quad (7)$$

We parameterize $\lambda_j = \lambda_0(1 - x_t) + \lambda_1 x_t$, $A_j(w, z) = A_0(w, z)(1 - x_t) + A_1(w, z)x_t$ for $\{w, z\} = 1, 2, 3$, and $\mu_j^i = \mu_0^i(1 - x_t) + \mu_1^i x_t$ for $i = 1, 2, 3$.

4.1 Approximate filtering and likelihood evaluation

This section summarizes the issues involved in estimating a state space model with Markov switching and explains the approach followed in this paper.

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

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

where $Y^{t-1} = \{y_1, y_2, \dots, y_{t-1}\}$ denotes the history of observations up to time $t-1$. Hence, evaluating the log-likelihood function requires computing the prediction probabilities $\Pr(y_t | Y^{t-1})$. These probabilities are obtained as a by-product of a recursive Bayesian filtering algorithm used to estimate the distribution of the latent variables f_t and x_t conditional on Y^{t-1} . As it is well known, however, Bayesian filtering with Markov switching implies that posterior distributions are mixtures of prior distributions that grow exponentially with time. This makes exact filtering intractable. To operationalize the likelihood evaluation, we use an approximate Bayesian filter that collapses the growing mixture distributions to a single distribution at each time t .

Consider the recursive evaluation of the filter at time t . Given filtered probabilities $\Pr(f_{t-1} | Y^{t-1})$ and $\Pr(x_{t-1} | Y^{t-1})$, our objective is to compute the posterior densities $\Pr(f_t | Y^t)$ and $\Pr(x_t | Y^t)$, and the contribution to the likelihood function $\Pr(y_t | Y^{t-1})$. To that end, suppose that the filtered probability $\Pr(f_{t-1} | Y^{t-1})$ is Gaussian,

$$\Pr(f_{t-1} | Y^{t-1}) = N(\hat{f}_{t-1|t-1}, V_{t-1|t-1}), \quad (8)$$

where, here and throughout the paper, $N(\mu, \Omega)$ denotes a Gaussian distribution with mean μ and covariance matrix Ω . The vector $\{f_{t-1} | Y^{t-1}\}$ is Gaussian by assumption at $t = 1$ and by our approximating formula at any other $t > 1$.

4.1.1 The prediction step

The conditional Gaussian model (5) implies that the prediction probability $\Pr(f_t | Y^{t-1}, x_t = j)$ is also Gaussian, as it is an affine function of two Gaussian variables, $\{f_{t-1} | Y^{t-1}\}$ and η_{jt} . Thus,

$\Pr(f_t|Y^{t-1}, x_t = j) = N(\hat{f}_{t|t-1}^j, V_{t|t-1}^j)$, where

$$\begin{aligned}\hat{f}_{t|t-1}^j &= \mu_j + A_j \hat{f}_{t-1|t-1} \text{ and} \\ V_{t|t-1}^j &= A_j V_{t-1|t-1} A_j' + H_j.\end{aligned}$$

Likewise, equation (4) implies that $\Pr(y_t|Y^{t-1}, x_t = j) = N(\hat{y}_{t|t-1}^j, \Omega_{t|t-1}^j)$, where

$$\begin{aligned}\hat{y}_{t|t-1}^j &= \Lambda_j \hat{f}_{t|t-1}^j \text{ and} \\ \Omega_{t|t-1}^j &= \Lambda_j V_{t|t-1}^j \Lambda_j' + Q_j.\end{aligned}$$

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

$$\begin{aligned}\Pr(y_t|Y^{t-1}) &= \sum_{j=0}^K \Pr(x_t = j|Y^{t-1}) \Pr(y_t|Y^{t-1}, x_t = j) \\ &= \sum_{j=0}^K \left(\sum_{i=0}^K p_{ij} \Pr(x_{t-1} = i|Y^{t-1}) \right) \Pr(y_t|Y^{t-1}, x_t = j),\end{aligned}\tag{9}$$

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

4.1.2 The updating step

We use Bayes's law to update the probabilities $\Pr(x_t|Y^t)$ and $\Pr(f_t|Y^t)$ given the observation of y_t . In particular,

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

and

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

The next proposition, proved in Appendix B, shows that $\Pr(f_t|Y^t, x_t = j)$ is Gaussian.

Proposition 1 : $\Pr(f_t|Y^t, x_t = j)$ is Gaussian with mean and covariance matrix given by

$$\begin{aligned}\widehat{f}_{t|t}^j &= \widehat{f}_{t|t-1}^j + V_{t|t-1}^j \Lambda_j' \left(\Omega_{t|t-1}^j \right)^{-1} \left(y_t - \Lambda_j \widehat{f}_{t|t-1}^j \right), \\ V_{t|t}^j &= V_{t|t-1}^j - V_{t|t-1}^j \Lambda_j' \left(\Omega_{t|t-1}^j \right)^{-1} \Lambda_j V_{t|t-1}^j.\end{aligned}$$

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

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

4.1.3 Collapsing the posterior probability $\Pr(f_t|Y^t)$

So far we showed that, if the prior probability $\Pr(f_{t-1}|Y^{t-1})$ is Gaussian, the posterior probability $\Pr(f_t|Y^t)$ is a mixture of $K + 1$ Gaussians. We make the recursive evaluation of the filter operational by collapsing $\Pr(f_t|Y^t)$ 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, page 453). Simple algebra shows that these means and covariances are given by

$$\widehat{f}_{t|t} = \sum_{j=0}^K \Pr(x_t = j|Y^t) \widehat{f}_{t|t}^j \quad (10)$$

and

$$V_{t|t} = \sum_{j=0}^K \Pr(x_t = j|Y^t) \left(V_{t|t}^j + \left(\widehat{f}_{t|t} - \widehat{f}_{t|t}^j \right) \left(\widehat{f}_{t|t} - \widehat{f}_{t|t}^j \right)' \right). \quad (11)$$

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

Before moving to the next section, we mention that there are other ways of approximating the filtering problem. For example, Harrison and Stevens (1976) and Kim (1994) start by assuming that the filtered probability $\Pr(f_{t-1}|Y^{t-1})$ is a Gaussian mixture with $K + 1$ components. With this assumption, the updated probabilities and the contribution to the likelihood function are Gaussian mixtures with $(K + 1)^2$ components. The algorithm is closed by collapsing the $(K + 1)^2$ -component

Gaussian mixture $\Pr(f_t|Y^t)$ to a mixture with $K + 1$ components. We experimented with this approach but found numerical instabilities when estimating the model of the yield curve. West and Harrison (1997) and Murphy (1998) discuss alternative collapsing schemes.

4.2 Forecasting

In this section we discuss an algorithm to compute optimal forecasts using the Markov switching model (4)-(5). 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 time t imposing the assumption that the filtered probability $\Pr(f_t|Y^t)$ is Gaussian with mean $\hat{f}_{t|t}$ and covariance $V_{t|t}$. Consider first forecasting future regime probabilities x_{t+h} . Given the Markovian structure,

$$\Pr(x_{t+h} = j|Y^t) = \sum_{i=0}^K p_{ij}^{(h)} \Pr(x_t = i|Y^t). \quad (12)$$

where $p_{ij}^{(h)}$, 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 $\Pr(f_{t+1}|Y^t, x_{t+1} = i_1)$. Equation (5) implies that

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

where

$$\begin{aligned} \hat{f}_{t+1|t}^{i_1} &= \mu_{i_1} + A_{i_1} \hat{f}_{t|t}, \text{ and} \\ V_{t+1|t}^{i_1} &= A_{i_1} V_{t|t} A_{i_1}' + H_{i_1}. \end{aligned} \quad (13)$$

Integrating out the regimes gives the marginal probability

$$\Pr(f_{t+1}|Y^t) = \sum_{i_1=0}^K \Pr(f_{t+1}|Y^t, x_{t+1} = i_1) \Pr(x_{t+1} = i_1|Y^t). \quad (14)$$

Similarly, equation (4) implies

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

where

$$\begin{aligned}\hat{y}_{t+1|t}^{i_1} &= \Lambda_{i_1} \hat{f}_{t+1|t}^{i_1}, \text{ and} \\ \Omega_{t+1|t}^{i_1} &= \Lambda_{i_1} V_{t+1|t}^{i_1} \Lambda_{i_1}' + Q_{i_1}.\end{aligned}\tag{15}$$

Integrating over future regimes gives the forecast density

$$\Pr(y_{t+1}|Y^t) = \sum_{i_1=0}^K \Pr(y_{t+1}|Y^t, x_{t+1} = i_1) \Pr(x_{t+1} = i_1|Y^t).\tag{16}$$

Iterating forward, equations (4) and (5) imply that the conditional h-period ahead forecast densities satisfy

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

where

$$\begin{aligned}\hat{f}_{t+h|t}^{i_1, i_2, \dots, i_h} &= \mu_{i_h} + A_{i_h} \hat{f}_{t+h-1|t}^{i_1, i_2, \dots, i_{h-1}}, \\ V_{t+h|t}^{i_1, i_2, \dots, i_h} &= A_{i_h} V_{t+h-1|t}^{i_1, i_2, \dots, i_{h-1}} A_{i_h}' + H_{i_h},\end{aligned}\tag{17}$$

$$\begin{aligned}\hat{y}_{t+\tau|t}^{i_1, i_2, \dots, i_h} &= \Lambda_{i_h} \hat{f}_{t+h|t}^{i_1, i_2, \dots, i_h}, \text{ and} \\ \Omega_{t+\tau|t}^{i_1, i_2, \dots, i_h} &= \Lambda_{i_h} V_{t+h|t}^{i_1, i_2, \dots, i_h} \Lambda_{i_h}' + Q_{i_h}.\end{aligned}\tag{18}$$

Finally, integrating over future regimes gives the forecast densities

$$\Pr(f_{t+h}|Y^t) = \sum_{i_1, \dots, i_h=0}^K \Pr(f_{t+h}|Y^t, x_{t+1} = i_1, \dots, x_{t+h} = i_h) \Pr(x_{t+h} = i_h|Y^t)\tag{19}$$

$$\Pr(y_{t+h}|Y^t) = \sum_{i_1, \dots, i_h=0}^K \Pr(y_{t+h}|Y^t, x_{t+1} = i_1, \dots, x_{t+h} = i_h) \Pr(x_{t+h} = i_h|Y^t). \quad (20)$$

Note that both 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.

We summarize the forecast algorithm:

1. Initialize the algorithm using the filtered probabilities $\Pr(x_t|Y^t)$ and the collapsed distribution $\Pr(f_t|Y^t) \approx N(\hat{f}_{t|t}, V_{t|t})$, both obtained from the approximate filter.
2. Compute the forecast probabilities using (12).
3. Forecasting at $t + 1$:
 - (a) For each possible realization of the regime $x_{t+1} = i_1$, compute the $K + 1$ arrays $\hat{f}_{t+1|t}^{i_1}, V_{t+1|t}^{i_1}$ using (13). Next, marginalize the regimes and compute $\Pr(f_{t+1}|Y^t)$ using (14).
 - (b) For each possible realization of the regime $x_{t+1} = i_1$, compute the $K + 1$ arrays $\hat{y}_{t+1|t}^{i_1}, \Omega_{t+1|t}^{i_1}$ using (15). Next, marginalize the regimes and compute $\Pr(y_{t+1}|Y^t)$ using (16).
 - (c) Compute the forecast $E(y_{t+1}|Y^t)$ using the prediction density $\Pr(y_{t+1}|Y^t)$.
4. Forecasting at $t + h$:
 - (a) For each $x_{t+1} = i_1, x_{t+2} = i_2, \dots, x_{t+h} = i_h$ compute iteratively the $(K + 1)^h$ arrays $\hat{f}_{t+h|t}^{i_1, i_2, \dots, i_h}, V_{t+h|t}^{i_1, i_2, \dots, i_h}$ using (17) and the results of the $h - 1$ forecast horizon. Next, marginalize the regimes and compute $\Pr(f_{t+h}|Y^t)$ using (19).
 - (b) For each $x_{t+1} = i_1, x_{t+2} = i_2, \dots, x_{t+h} = i_h$ compute the $(K + 1)^h$ arrays $\hat{y}_{t+h|t}^{i_1, i_2, \dots, i_h}$ and $\Omega_{t+h|t}^{i_1, i_2, \dots, i_h}$ using (18). Next, marginalize the regimes and compute $\Pr(y_{t+h}|Y^t)$ using (20).
 - (c) Compute the forecast $E(y_{t+h}|Y^t)$ using the prediction density $\Pr(y_{t+h}|Y^t)$.

5 Estimating the yield curve using a Markov switching latent variable model

We estimate a regime switching model that explicitly takes into account the changing patterns of the yield curve over the cycle. The model forms a state-space system with a Markov switching first order vector autoregression equation summarizing the dynamics of the latent state variables, and a switching linear measurement equation relating the observed yields with the state vector. For a given parameter configuration, we use the approximate filtering procedure described above to evaluate the prediction-error decomposition of the likelihood function. To initialize the filter, we set the mean and covariance matrix of the latent variables using their regime-specific long-run mean and covariance matrix, and set the initial regime probabilities using the associated steady state probabilities. The model specified in equations (6) and (7) allow for switching in all the parameters of interest. We estimate special cases of the general model that constrain some parameters to be the same between regimes and evaluate their fitting and forecasting performance. For comparison purposes, we also estimate the single regime model and recover the estimates in Diebold, Rudebusch, and Aruoba (2006).

Christensen, Diebold, and Rudebusch (2011) note that it is common to find that a parametrization of a yield curve model that is not rejected in sample has a 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 high number of possible parametrizations that are special cases of the general model, we only report a subset of all possible restricted models. We do not report models that are outperformed in terms of in sample fitting or out of sample forecasting performance using the measures described below.⁷ Given that the results for the most general model are not presented (for the reasons discussed above) we chose to compare the models using information criteria instead of likelihood ratio tests. Information criteria have been found to be useful in selecting among different regime dependent models in sample Psaradakis and Spagnolo (2003 and 2006)—note, however, that comparing the switching models with their single regime alternative is problematic due to the usual nuisance parameter problem described in Davies (1977 and 1987).

Maximizing the approximate likelihood function is a non-trivial task: standard quasi-Newton

⁷Results on the omitted models are available on request.

methods often fails because the algorithm tries to evaluate the likelihood function at points where it is not well defined. To solve this problem, we maximize the likelihood using a series of optimization algorithms. We start with a robust but slow algorithm and progressively move toward faster but less robust algorithms. In particular, we first maximize the likelihood function using a simulated annealing algorithm (Kirkpatrick, Gelatt, and Vech, 1983) with initial temperature set to 1 and final temperature set to 10^{-5} .⁸ Once the simulated annealing algorithm has converged, we use the estimated parameters to initialize a Nelder-Mead derivative-free direct search algorithm. The algorithm is stopped when the norm of estimated parameters from one iteration to the next is less than 10^{-4} or when the function changes by less than 10^{-4} . Next, we use quasi-Newton methods with the BFGS algorithm to update the Hessian and a cubic spline line search procedure. The convergence criterion for the change in the norm of the parameters is 10^{-6} .⁹ We compute asymptotic standard errors using the delta method and the inverse of the empirical Hessian.

5.1 Empirical results

We now consider the results from estimating the switching state-space model of the type specified in equations (6) and (7). Tables 2 and 3 show the results. In all switching estimations, regimes 0 and 1 are shown in the first and second columns, respectively. Model 1 in Table 2 corresponds to the baseline estimation without switching. Models 2 and 3 in Table 2 show the estimation of the model switching the parameter λ with the full parameter matrices A and H in model 2 and with diagonal A and H in model 3. Model 4 in Table 3 shows the estimation of the model switching not only λ but also μ , with full parameter matrices A and H . Lastly, model 5 reports results switching the parameters λ , μ , and A , with A diagonal and (common) covariance matrix H diagonal.

Tables 2 and 3 about here

All the switching models show significant evidence of regime shifts for some of the parameters. We select models 4 and 5 as our preferred specifications based on the Akaike and Bayesian information criteria. Our estimates suggest that the parameters λ_i ($i = 0, 1$) are very different between

⁸Simulated annealing is a random search optimization method with many appealing properties: it is robust to different starting values; it is known to avoid local optima; it does not require derivative information; and is able to optimize functions where other local optimization methods fail. While robust, the algorithm tends to be very slow near the optimum (Goffe, Ferrier, and Rogers 1994).

⁹Both local search algorithms are those of the MATLAB optimization package.

regimes. We plot the regime specific loadings on factors β_{2t} and β_{3t} corresponding to model 5 in the top panel of Figure 2. 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 estimate of λ in regime 0 (about 0.13) implies that the loading on factor β_{3t} is maximized at a maturity of 13 months while the estimate of λ in regime 1 (about 0.05) implies that the aforementioned loading is maximized around a maturity of 30 months. In other words, (equal) 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 β_{2t} factor than those in regime 0. Note, also, that the estimate of λ in the linear model 1—which replicates the results in Diebold, Rudebusch, and Aruoba (2006)—is always between the estimates of λ_0 and λ_1 of the models with regime shifts. (We plot the factor loadings of the linear model 1 in the lower panel of Figure 2.)

Figure 2 about here

Model 4 suggests highly persistent dynamics of β_{1t} , β_{2t} , and β_{3t} , with estimated eigenvalues of 0.98, 0.92, and 0.75. When allowing the transition matrix to switch, in model 5, we observe somewhat different dynamics in each regime as factor β_{2t} is more persistent in regime 0 than in regime 1. In model 4, the drift for the factors β_{1t} and β_{2t} is positive in both regimes, and the drift for the factor β_{3t} is negative and statistically significant in both regimes, but it is twice as large, in absolute value, in regime 1. In model 5, the drifts are very different across regimes, although they are not statistically significant. Finally, in both models (and in linear model 1) we find the same pattern for the estimates of the covariance matrix H : the volatility of the transition shocks, measured by the estimated diagonal elements of H , increases as we move from factors β_{1t} to β_{2t} to β_{3t} .

Diebold and Li (2006) interpret the latent variables β_{2t} and β_{3t} as directly associated with the slope and curvature of the yield curve, respectively. For values of λ different from those proposed by Diebold and Li, this interpretation is less obvious. In particular, in regime 0 ($\lambda_0 = 0.13$), the slope and curvature of the yield curve satisfy

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

While the slope of the yield curve is still mostly affected by the factor β_{2t} , the curvature is almost equally sensitive to fluctuations in β_{2t} and β_{3t} .¹⁰

Figure 3 shows the estimated evolution of the three latent factors and the actual level, slope, and curvature of the yield curve.¹¹ Following Diebold and Li (2006), we plot β_{1t} , $-\beta_{2t}$, and $0.3\beta_{3t}$ and compare these factors with the actual level, slope, and curvature of the yield curve. The correlation between the factor β_{1t} and the level of yields, and between the factor β_{2t} and the curvature is large, of 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, of about 0.8. In any case, the three factors are still sufficiently correlated with the measures of level, slope, and curvature and, therefore, we keep Diebold and Li's interpretation of the latent factors and call β_{1t} the level factor, β_{2t} the slope factor, and β_{3t} the curvature factor.

The estimated transition probabilities imply that the two regimes are persistent. For example, in model 5, the economy spends on average 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).¹² The bottom right panel of Figure 3 displays the filtered probability of regime 0, given by $\Pr(x_t = 0|Y^t; \hat{\theta})$ where $\hat{\theta}$ is an estimator of the unknown parameters in (6) and (7). Even though these probabilities do not coincide with the NBER dates of the business cycles peaks and troughs, we still find some robust similarities between NBER recessions and regime 0, and between NBER booms and regime 1.¹³

We reproduce in Figure 4 and panel B of Table 1 the statistics computed in Figure 1 and panel A of Table 1, but separating regimes using a dummy variable that equals 1 if $\Pr(x_t = 0|Y^t; \hat{\theta}) > 0.5$ and zero otherwise. In computing these statistics, we only use subperiods with six or more consecutive observations in a given regime. Consider first the average yield curves. As in the

¹⁰Factors in regime 1 ($\lambda = 0.05$) are consistent with Diebold and Li's interpretation in that the slope is mostly associated with β_{2t} and the curvature, with β_{3t} . Indeed,

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

¹¹We report the mean of the collapsed posterior density of β_{1t} , β_{2t} , and β_{3t} computed using equation (10).

¹²The first statistic is computed using the invariant distribution of regimes. The second statistic is computed as $1/(1 - p_{ii})$ where $i = 0, 1$ denotes the current regime.

¹³Interestingly, regime 1 seems to be capturing, at least in part, deviations of the yield curvature from the factor β_{3t} : the correlation between the filtered probability of regime 1 and the absolute value of the difference between the curvature and $0.3\beta_{3t}$ is 0.26.

comparison between recessions and booms, we observe the following similarities: (1) average yields are greater in regime 0 than in regime 1 and the difference between yields decreases as maturity increases; (2) except at short maturities, the average yield curve is flatter in regime 0 than in regime 1; and (3) the yield curve in regime 1 has a similar shape to the average yield curve over the entire sample. Second, the volatility of yields decreases sharply as maturity increases in regime 0 but is substantially flatter in regime 1. And third, we still observe, as in the comparison between booms and recessions, that yields are more persistent in regime 1 than in regime 0. This difference in persistence, however, is more clear at short maturities.

Some of the statistics in panel B of Table 1 are also consistent the view that regime 0 resembles recessions and regime 1 resembles booms. For example, the average level, slope, and curvature of the yield curve are greater in regime 0 than in regime 1, and the curvature is substantially more volatile in regime 0 than in regime 1. Of course, the proposed analogy between regimes and recessions/booms is far from perfect. This is to be expected as, during the sample period, the economy spent roughly the same time in regimes 0 and 1, but only 46 months out of 348 months in recession according to the NBER dates. In any case, however, many of the characteristics of the yield curve in recessions are captured by regime 0 and those in booms, by regime 1.

Figure 4 about here

In Figure 5 we compare actual and fitted yield curves for selected dates using the linear model and switching models 4 and 5. There are two things to note. First, on the proposed dates, the models with switching matches the yield curve better than the linear model. We also computed the in-sample mean absolute errors and root-mean squared errors of fitted yields for each maturity across all periods and found that model 5 tends to produce larger errors than the other two models; on the other hand, model 4 produces smaller errors than the linear model for all maturities smaller than 72 months and similar errors (or slightly larger) for longer maturities. In any case, we show in the next section that, in many cases, model 5 tends to have better out-of-sample forecasting performance than the other two models. And second, on a given date, the average difference between actual and fitted yields is not zero, as can be seen in the March 1989 yield curve. In contrast, fitted yield curves in Diebold and Li (2006) have, on average, zero errors across maturities. The reason for this difference is that Diebold and Li, by running independent OLS regressions, minimize the in-sample

difference between actual and fitted yield curves period by period. In our model, by imposing ex-ante an explicit dynamic structure on the latent factors, maximizing the likelihood function does not necessarily imply minimizing the difference between actual and fitted yield on a period-by-period basis. Of course, the same is true in Diebold, Rudebusch, and Aruoba (2006), although they do not report plots with the in-sample fit of their model.

As a reference for the out-of-sample forecasts presented in next section, Tables 4 and 5 show the estimation of the same models 1 through 5 but over a short sample that goes from January 1972 to December 1993. Overall the estimated parameter values in all models is very similar to those using the full sample. The Akaike and Bayesian information criteria now select model 5 as the preferred specification.

5.2 Out-of-Sample forecasting

In this subsection we evaluate the accuracy of out-of-sample forecasts from the empirical models discussed earlier. The comparisons are based on a series of recursive forecasts beginning in 1994:1 and extending through 2000:12 (84 sample points). We compare h-month-ahead out-of sample forecasting results from each of the five competing models for maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120 months, and forecast horizons of $h = 1, 3, 6$ and 12 months.

Several measures are used to compare forecast accuracy. These measures include: i) standard mean squared errors (MSE), including the proportion of the times that each model achieves the smallest MSE over the 84 sample points, ii) tests of equal predictive accuracy of Clark and West (2007) (CW) and Harvey, Leybourne, and Newbold (1997) modified tests of Diebold and Mariano (1995) (DM), and iii) the confusion rate (CR).

Tables 6 to 13 reports the results of the forecast exercise. Let us start by considering Tables 6 to 9 (left part); it is clear that the smallest MSE is achieved by model 3 for all maturities and forecast horizons (with the only exception for the long maturities of the 12 months forecast horizon). Note also that, from a first inspection, the linear model (model 1) seems to be generally outperformed by the nonlinear alternatives. More specifically, the gain of using nonlinear models over the alternative linear model achieves the highest point for maturity 3 and forecast horizons 3 and 6 month, with a gain of model 3 over model 1 of, respectively, 78% and 68%. Turning to the linear model, it is clear that model 1 achieves more frequently the smallest MSE against models 2 and 4. In particular, the

marginal gain of the linear model over the nonlinear alternatives on the 3 month maturity varies from around 1% (using model 2 and a forecast horizon of 6 months) to 32% (using model 4 and a forecast horizon of 12 months).

On the right part of Tables 6 to 9 we report the proportion of time that each model achieves the smallest MSE over the 84 sample points (1994:1-2000:12), calculated for each individual maturities and forecast horizons. Contrary to our previous findings, overall we find that model 5 outperforms the alternative specifications (in particular for the medium horizons), achieving the smallest MSE for 11 maturities (over the 17 presented) for the 1 month horizon, and 13 maturities for the 12 months. On the other hand, model 3 seems to do particularly well in the smallest and longest horizons, achieving the smallest MSE for 9 maturities (over the 17) at the 3 months forecasts and 11 maturities at the 6 months. It is very informative to compare these results with those presented in Tables 5 to 8 (left part), where we look at the average errors. The results show, for example, that while for the 12 month ahead the smallest MSE is achieved by model 5 only for 2 maturities (36 and 48), when we evaluate the performance in terms of the number of periods with the smallest errors, we find that model 5 outperforms the competing models for 13 out of the 17 maturities. Most importantly, we also find that the linear model 1 never clearly achieves the smallest MSE, with the only exception of the 120 maturities for the 1 month ahead. Finally, on the basis of this criterion model 4 never outperform the competing alternative models while model 2 only achieves the smallest MSE for 5 points over the total 68 calculated.

Simply comparing the values of the MSE does not tell us how significant the difference is. Therefore, we use tests of equal forecast accuracy which can be applied to examine whether the MSE of two alternative models are significantly different from each other. The comparison is made between the linear model 1 and the nonlinear alternatives. Note that models 2 and 4 nest the alternative linear model 1 while the former is not nested by models 3 and 5. This suggest the use of two classes of tests which are specifically designed to test for equal forecast accuracy between nested and non-nested models. For non-nested models (models 3 and 5), we use a test of equal forecast accuracy due to Diebold and Mariano (1995) which is applied to examine whether the MSE of two alternative models are significantly different from each other. The comparison is based on the modified statistic proposed by Harvey, Leybourne, and Newbold (1997) to overcome the problem of over-sized of the original test in small and moderate samples (particularly acute for longer forecast

horizons).¹⁴ For nested models (models 2 and 4), a direct comparison of forecast errors from the parsimonious model (linear model 1) with those from the more general models (within which they are nested) are based on Clark and West (2007), who suggest the use of MSE-based statistics which are adjusted for the noise that is introduced into the forecasting process by the more general model when its additional parameters are not helpful for prediction.¹⁵

From Tables 6 to 9, a comparison of the nonlinear specification with the parsimonious linear model that it nests reveals that, on the basis of the MSE-adjusted differences, model 2 has a general advantage over the linear model for the 3 months ahead forecasts. For the 1 month ahead (6 months ahead), however, the gain over the linear alternative is only evident for the shortest (medium) maturities. On the other hand it is clear that the Clark–West adjustment seems not to be large enough to give model 4 an advantage over the restricted model, with the only exception for the first maturity of the 1 month ahead forecast. Turning to the modified Diebold Mariano tests, results show that the null hypothesis of equal forecast accuracy is clearly rejected for all the maturities and forecast horizons (with the only exceptions of the long maturities for the 12 months forecasts) in favor of model 3. Per contra, results for model 5 appear more mixed, showing that the null hypotheses of equal forecast accuracy are more frequently rejected for the 6 months and 12 months ahead.

An alternative way of evaluating out-of-sample forecasts is to assess the ability of different models to identify the direction of change in the variable of interest, regardless of the accuracy with which the magnitude of the change is predicted. As Haefke and Helmenstein (1996) and Swanson and White (1997) point out, this evaluation method is particularly useful in situations where directional predictions are the focus of the analysis, as is the case, for instance, when we are trying to forecast the next business-cycle turning point or the future price movements of asset prices. We therefore complement our evaluation of the relative performance of the different models using traditional accuracy measures by assessing their ability to correctly identify turning points (i.e. whether the rates are rising or falling regardless of the accuracy with which the magnitude of the change is

¹⁴The 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 statistic of Diebold and Mariano (1995).

¹⁵Let τ_1 and τ_2 be respectively the MSE of the parsimonious model (M1) and the corresponding larger model (M2) that nests M1, the Clark–West adjustment term is computed as $d = (1/N) \sum_{j=1}^N (\hat{y}_{1,T-N+j} - \hat{y}_{2,T-N+j})^2$, where $\hat{y}_{1,T-N+j}$ and $\hat{y}_{2,T-N+j}$ are the step-ahead forecasts from M1 and M2, respectively. A t -statistic for testing the null hypothesis of equal predictive accuracy is then based on $\tau_1 - \tau_2 + d$ (the alternative hypothesis is that M2 has a smaller MSPE than M1).

predicted) using the so-called confusion rate.¹⁶ From Tables 10 to 13, confusion rates results show that models 3 and 5 do quite well in identifying turning points for the 1 month ahead and for the 3 months (model 3) and 6 month ahead (model 5). On the other hand, while linear model 1 does particularly well for the 3 month maturity at all forecast horizons, model 2 appear to be the most successful for the 12 months ahead.

In summary, the results presented in this section suggest that our proposed models seems to not only successfully characterize the data under scrutiny but also, and perhaps more importantly, to have a good forecasting performance (in particular for models 3 and 5). The forecasting results are particularly noteworthy because one of the major weaknesses of many existing nonlinear models is their relatively poor out-of-sample performance.

6 Conclusions

In this paper we have proposed an extension of the three-factor term structure models of Nelson and Siegel (1987) and Diebold and Li (2006). We have shown that our model is parsimonious and relatively easy to estimate. Furthermore, using US zero-coupon data, we have shown that: i) the model has the substantial flexibility required to match the changing shape of the yield curve; ii) the data present clear evidence of regime shifts that can be adequately captured by our proposed model, iii) the estimated transition probabilities implies that the regimes are highly persistent; iv) the typical stylized facts of the yield curve are associated with the boom periods, v) the model has good forecasting performance. The forecasting results are particularly noteworthy as they contradict a common empirical findings on the relatively poor out-of-sample performance of nonlinear models.

¹⁶Consider the following 2×2 contingency table:

	actual up	actual down
predicted up	a_{11}	a_{12}
predicted down	a_{21}	a_{22}

The columns correspond to actual moves, up or down, while the rows correspond to predicted moves. Hence, a_{11} and a_{22} correspond to correct directional predictions, while a_{12} and a_{21} correspond to incorrect predictions. The performance of the model can be assessed in terms of its so-called confusion rate, $CR = (a_{12} + a_{21}) / (a_{11} + a_{12} + a_{21} + a_{22})$, i.e. the ratio of the sum of the off-diagonal elements to the sum of all elements.

References

- [1] Ang, A. and M. Piazzesi. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745-787.
- [2] Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291-311.
- [3] Christensen, J.H.E., Diebold, F. X., and Rudebusch, G.D. (2011) The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics* 164, 4-20.
- [4] Coroneo, L., Nyholm, K., and Vidova-Koleva, R. (2011). How arbitrage-free is the Nelson and Siegel model? *Journal of Empirical Finance* 18, 393-407.
- [5] Costa, O. L. V., Fragoso, M. D., and Marques, R. P. (2005). *Discrete-time Markov jump linear systems*. London, Springer.
- [6] Dai, Q. and Singleton, K. (2000). Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-78.
- [7] Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247-54.
- [8] Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33-43.
- [9] Diebold, F.X., and Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* 130, 337-64
- [10] Diebold, F.X., Rudebusch, G.D. and Aruoba, S.B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics* 131, 309-38
- [11] Diebold, F.X. and Mariano R.S. (1995). Comparing predictive accuracy, *Journal of Business, Economics and Statistics* 13, 253-263.
- [12] Driffill J., Kenc T., Sola M, and Spagnolo F. (2009). The effects of different parameterizations of Markov-Switching in a CIR model of bond pricing, *Studies in Nonlinear Dynamics & Econometrics*, Berkeley Electronic Press, vol. 13(1), pages 1.
- [13] Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance* 6, 379-406.
- [14] Elliot, R.J, Abboun, L., and Moore, J.B. (1995). *Hidden markov models: estimation and control*. New York, Springer.
- [15] Fama, E.F. and Bliss, R.R. (1987). The Information in long-maturity forward rates. *American Economic Review* 77, 680-92.
- [16] Goffe, W.L., Ferrier G.D. and Rogers J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics* 60, 65-99.
- [17] Harrison, P. J. and C. F. Stevens (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Series B* 38, 207–247.

- [18] Harvey, D. I., Leybourne, S. J. and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13, 281-291.
- [19] Koopman, S. J., Mallee, M.I.P., and 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.
- [20] Kim, C. J., and Nelson C. R. (1999). *State-Space models with regime switching*. Cambridge, Massachusetts: MIT Press.
- [21] Kim, C. J. (1994). Dynamic linear models with Markov-Switching, *Journal of Econometrics* 60, 1-22.
- [22] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science. New Series* 220 (4598): 671–680.
- [23] Knez, P., Litterman, R. and Scheinkman, J. (1994). Exploration into factors explaining money market returns,. *Journal of Finance* 49, 1861-82.
- [24] Martellini and Priaulet (2001). *Fixed-income securities: dynamic methods for interest rate risk pricing and hedging*, Willey.
- [25] Murphy, K.P. (1998). *Switching Kalman filters*. Dept. of Computer Science, University of California, Berkeley, Tech. Rep.
- [26] Nelson, C.R. and Siegel, A.F. (1987). Parsimonious modeling of yield curves, *Journal of Business* 60, 473-89.
- [27] Piazzesi, M. (2010). Affine term structure models. *Handbook of Financial Econometrics*, Elsevier.
- [28] Psaradakis, Z. and Spagnolo, N. (2003). On the determination of number of regimes in Markov-switching autoregressive models, *Journal of Time Series Analysis* 24 (2) : 237- 252
- [29] Psaradakis, Z. and Spagnolo, N. (2006). Joint determination of the state dimension and autoregressive order for models with Markov regime switching, *Journal of Time Series Analysis* 27 (5) : 753- 766
- [30] Svensson L. E.O. (1994). Estimating and interpreting forward interest rates: Sweden 1992–1994. IMF Working Paper 94/114.
- [31] West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer-Verlag, New York.

Appendix A

This appendix shows that, if the yield curve is represented by the Nelson and Siegel model (1), a coupon paying bond (or portfolio of bonds) can be immunized of the capital loss and reinvesting risks using only three additional bonds/portfolios. In contrast, if the yield curve is represented by the MSDNS model (2), hedging the original portfolio needs four additional bonds/portfolios.

Suppose first that the yield curve is represented by the Nelson and Siegel model (1), where β_{1t} , β_{2t} , and β_{3t} are stochastic processes. Consider now a bond (or portfolio of bonds) P_t that pays coupons M_i at time $t + \tau_i$, where $\tau_i = \{\tau_1, \tau_2, \dots, \tau_N\}$. The price of this bond is

$$P_t = \sum_{i=1}^N M_i B_t(\tau_i),$$

where $B_t(\tau_i) = e^{-R_t(\tau_i)\tau_i}$ is the price of a zero coupon bond at time t maturing τ_i periods ahead.

The portfolio P_t can be hedged for capital and reinvesting risk by constructing a new portfolio, P_t^* , containing the original portfolio and three additional coupon paying bonds, denoted by $z = 1, 2, 3$, with price P_t^z and coupons M_i^z paid at time $t + \tau_i$. For each dollar invested in the original portfolio, the extended portfolio invests $-\phi_t^z$ dollars in bond $z = 1, 2, 3$. Thus, the price of the extended portfolio is

$$P_t^* = P_t - \phi_t^1 P_t^1 - \phi_t^2 P_t^2 - \phi_t^3 P_t^3. \quad (21)$$

Because the yield curve is assumed to be generated by model (1), the change in the prices of all bonds are a consequence of changes in the three latent factors. Thus, for small changes in the three factors, the change in the price of any bond \tilde{P}_t with associated coupons \tilde{M}_i is

$$d\tilde{P}_t = \frac{\partial \tilde{P}_t}{\partial \beta_{1t}} d\beta_{1t} + \frac{\partial \tilde{P}_t}{\partial \beta_{2t}} d\beta_{2t} + \frac{\partial \tilde{P}_t}{\partial \beta_{3t}} d\beta_{3t},$$

where

$$\begin{aligned} \frac{\partial \tilde{P}_t}{\partial \beta_{1t}} &= -\tau_i \sum_{i=1}^{\tau} \tilde{M}_i B_t(\tau_i), \\ \frac{\partial \tilde{P}_t}{\partial \beta_{2t}} &= -\tau_i \sum_{i=1}^{\tau} \tilde{M}_i \frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} B_t(\tau_i), \\ \frac{\partial \tilde{P}_t}{\partial \beta_{3t}} &= -\tau_i \sum_{i=1}^{\tau} \tilde{M}_i \left(\frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} - e^{-\lambda \tau_i} \right) B_t(\tau_i). \end{aligned}$$

Given weights ϕ_t^z , the change in the value of the hedging portfolio P_t^* is thus given by

$$dP_t^* = dP_t - \phi_t^1 dP_t^1 - \phi_t^2 dP_t^2 - \phi_t^3 dP_t^3.$$

To hedge the extended portfolio amounts to choose a vector, $(\phi_t^1, \phi_t^2, \phi_t^3)'$ that makes the change in the value of the extended portfolio, dP_t^* , equal to zero for any possible values of $d\beta_{1t}$, $d\beta_{2t}$, and

$d\beta_{3t}$. This hedging vector is

$$\begin{bmatrix} \phi_t^1 \\ \phi_t^2 \\ \phi_t^3 \end{bmatrix} = \begin{bmatrix} \partial P_t^1 / \partial \beta_{1t} & \partial P_t^2 / \partial \beta_{1t} & \partial P_t^3 / \partial \beta_{1t} \\ \partial P_t^1 / \partial \beta_{2t} & \partial P_t^2 / \partial \beta_{2t} & \partial P_t^3 / \partial \beta_{2t} \\ \partial P_t^1 / \partial \beta_{3t} & \partial P_t^2 / \partial \beta_{3t} & \partial P_t^3 / \partial \beta_{3t} \end{bmatrix}^{-1} \begin{bmatrix} \partial P_t / \partial \beta_{1t} \\ \partial P_t / \partial \beta_{2t} \\ \partial P_t / \partial \beta_{3t} \end{bmatrix}.$$

Hedging with the Markov Switching Dynamic Nelson Siegel model

The dynamic Nelson and Siegel model with Markov switching has four sources of risk: the risk associated with changes in the latent factors $\beta_{1t}(x_t)$, $\beta_{2t}(x_t)$, and $\beta_{3t}(x_t)$ conditional on the Markov regime x_t , and the risk associated with regime switches. In this section we show that the additional risk associated with the Markov switching structure changes the associated hedging portfolio: we now need four bonds instead of three.

For mathematical convenience, in this section we model the Markov regimes x_t as a continuous time Markov chain. To do this we need to depart somewhat from the notation in the body of the paper. The gain is that the continuous time structure delivers simpler expressions.¹⁷ We consider a two state Markov chain, with the state space of the chain x_t taken to be the two unit vectors $X = \{e_0, e_1\}$, where $e_0 = (1, 0)'$ and $e_1 = (0, 1)'$. The transition rate matrix of the Markov chain x_t is denoted by $H = (h_{ij})$, for $i, j = 1, 2$, where H is a Q -matrix such that, for $i \neq j$, $h_{ij} > 0$ and $h_{ii} + h_{ji} = 0$, so that the elements in the diagonal are negative and given by $h_{ii} = -h_{ji} < 0$. We find it convenient to write H as

$$H = \begin{bmatrix} -h_0 & h_1 \\ h_0 & -h_1 \end{bmatrix}.$$

The semi-martingale representation of the Markov chain is given by

$$x_t = x_0 + \int_0^t H x_v dv + N_t,$$

where N_t is an \mathbb{R}^2 -valued martingale process with respect to \mathcal{F}_t , the right continuous σ -algebra of information available at time t . The evolution of x_t can be written in differential form as

$$dx_t = H x_t dt + dN_t,$$

where $dN_{0t} + dN_{1t} = 0$ for all t (dN_{0t} and dN_{1t} are the two elements of the vector dN_t).

We now consider the possibility that the yield curve depends on the state x_t . Using the new notation, the price of the coupon paying bond can be written as $P_t(x_t) = \langle \vec{P}_t, x_t \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in \mathbb{R}^2 , $\vec{P}_t = (P_{0t}, P_{1t})'$, and $P_{jt} = P_t(e_j)$ is the price of the bond in regime $j = 0, 1$. The bond price P_t can change because the level, slope, and curvature change, or because there is a discrete regime change.

Thus, considering only first order changes in β_{1t} , β_{2t} , and β_{3t} , the change in the price of any bond $\tilde{P}_t(x_t)$ with associated coupons \tilde{M}_t can be written as

$$d\tilde{P}_t(x_t) = \frac{\partial \tilde{P}_t(x_t)}{\partial \beta_{1t}} d\beta_{1t} + \frac{\partial \tilde{P}_t(x_t)}{\partial \beta_{2t}} d\beta_{2t} + \frac{\partial \tilde{P}_t(x_t)}{\partial \beta_{3t}} d\beta_{3t} + \langle \vec{\tilde{P}}_t, dx_t \rangle,$$

¹⁷See Elliot, Abboun, and Moore (1995) for an exposition of Markov chains in continuous time and additional results mentioned in this section.

where the discrete impact term resulting from a regime shift is

$$\langle \vec{\tilde{P}}_t, dx_t \rangle = \langle \vec{\tilde{P}}_t, Hx_t dt \rangle + \langle \vec{\tilde{P}}_t, dN_t \rangle.$$

The components of the discrete impact term can be in turn expressed as follows: if $x_t = e_0$, $\langle \vec{\tilde{P}}_t, Hx_t dt \rangle = h_0 (\tilde{P}_{1t} - \tilde{P}_{0t}) dt$ and $\langle \vec{\tilde{P}}_t, dN_t \rangle = (\tilde{P}_{0t} - \tilde{P}_{1t}) dN_{0t}$; and if $x_t = e_1$, $\langle \vec{\tilde{P}}_t, Hx_t dt \rangle = h_1 (\tilde{P}_{0t} - \tilde{P}_{1t}) dt$ and $\langle \vec{\tilde{P}}_t, dN_t \rangle = (\tilde{P}_{1t} - \tilde{P}_{0t}) dN_{1t}$. In addition, the coupon paying bond in state $x_t = e_j$, for $j = 0, 1$, is given by

$$\tilde{P}_{jt} = \sum_{i=1}^N M_i B_{jt}(\tau_i),$$

where the price of the zero coupon bond is now $B_{jt}(\tau_i) = e^{-R_{jt}(\tau_i)\tau_i}$, $R_j(\tau_i)$ satisfies (2), and the state-dependent derivatives in state $j = 0, 1$, are

$$\begin{aligned} \frac{\partial \tilde{P}_{jt}}{\partial \beta_{1t}} &= -\tau_i \sum_{i=1}^{\tau} \tilde{M}_i B_j(\tau_i), \\ \frac{\partial \tilde{P}_{jt}}{\partial \beta_{2t}} &= -\tau_i \sum_{i=1}^{\tau} \tilde{M}_i \frac{1 - e^{-\lambda_j \tau_i}}{\lambda_j \tau_i} B_j(\tau_i), \\ \frac{\partial \tilde{P}_{jt}}{\partial \beta_{3t}} &= -\tau_i \sum_{i=1}^{\tau} \tilde{M}_i \left(\frac{1 - e^{-\lambda_j \tau_i}}{\lambda_j \tau_i} - e^{-\lambda_0 \tau_i} \right) B_j(\tau_i). \end{aligned}$$

Constructing the hedging portfolio To hedge the four sources of risk, it is necessary to construct an extended portfolio with four additional bonds, denoted by $z = 1, 2, 3, 4$. Conditional on state $j = 0, 1$, the extended portfolio invests $-\phi_{jt}^z$ dollars in bond $z = 1, 2, 3, 4$ for each dollar invested in the original portfolio. Given state $j = 0, 1$, we consider the portfolio

$$P_{jt}^* = P_{jt} - \phi_{jt}^1 P_{jt}^1 - \phi_{jt}^2 P_{jt}^2 - \phi_{jt}^3 P_{jt}^3 - \phi_{jt}^4 P_{jt}^4.$$

Using similar arguments as those presented when using the single regime model, we can choose a vector $(\phi_{jt}^1, \phi_{jt}^2, \phi_{jt}^3, \phi_{jt}^4)'$ that makes the change in the value of the extended portfolio, dP_{jt}^* , equal to zero given the four sources of risk. In state $j = 0$, this is accomplished with

$$\begin{bmatrix} \phi_{0t}^1 \\ \phi_{0t}^2 \\ \phi_{0t}^3 \\ \phi_{0t}^4 \end{bmatrix} = \begin{bmatrix} \partial P_{0t}^1 / \partial \beta_{1t} & \partial P_{0t}^2 / \partial \beta_{1t} & \partial P_{0t}^3 / \partial \beta_{1t} & \partial P_{0t}^4 / \partial \beta_{1t} \\ \partial P_{0t}^1 / \partial \beta_{2t} & \partial P_{0t}^2 / \partial \beta_{2t} & \partial P_{0t}^3 / \partial \beta_{2t} & \partial P_{0t}^4 / \partial \beta_{2t} \\ \partial P_{0t}^1 / \partial \beta_{3t} & \partial P_{0t}^2 / \partial \beta_{3t} & \partial P_{0t}^3 / \partial \beta_{3t} & \partial P_{0t}^4 / \partial \beta_{3t} \\ P_{1t}^1 - P_{0t}^1 & P_{1t}^2 - P_{0t}^2 & P_{1t}^3 - P_{0t}^3 & P_{1t}^4 - P_{0t}^4 \end{bmatrix}^{-1} \begin{bmatrix} \partial P_{0t} / \partial \beta_{1t} \\ \partial P_{0t} / \partial \beta_{2t} \\ \partial P_{0t} / \partial \beta_{3t} \\ P_{1t} - P_{0t} \end{bmatrix},$$

and, in state $j = 1$, with the portfolio

$$\begin{bmatrix} \phi_{1t}^1 \\ \phi_{1t}^2 \\ \phi_{1t}^3 \\ \phi_{1t}^4 \end{bmatrix} = \begin{bmatrix} \partial P_{1t}^1 / \partial \beta_{1t} & \partial P_{1t}^2 / \partial \beta_{1t} & \partial P_{1t}^3 / \partial \beta_{1t} & \partial P_{1t}^4 / \partial \beta_{1t} \\ \partial P_{1t}^1 / \partial \beta_{2t} & \partial P_{1t}^2 / \partial \beta_{2t} & \partial P_{1t}^3 / \partial \beta_{2t} & \partial P_{1t}^4 / \partial \beta_{2t} \\ \partial P_{1t}^1 / \partial \beta_{3t} & \partial P_{1t}^2 / \partial \beta_{3t} & \partial P_{1t}^3 / \partial \beta_{3t} & \partial P_{1t}^4 / \partial \beta_{3t} \\ P_{0t}^1 - P_{1t}^1 & P_{0t}^2 - P_{1t}^2 & P_{0t}^3 - P_{1t}^3 & P_{0t}^4 - P_{1t}^4 \end{bmatrix}^{-1} \begin{bmatrix} \partial P_{1t} / \partial \beta_{1t} \\ \partial P_{1t} / \partial \beta_{2t} \\ \partial P_{1t} / \partial \beta_{3t} \\ P_{0t} - P_{1t} \end{bmatrix}.$$

It is important to note that ignoring the regime specific nature of the model gives an incorrect coverage of the risk associated with changes in interest rates.

Appendix B

Proof of Proposition 1. Using Bayes's law

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

Thus

$$\begin{aligned} \Pr(f_t|Y^t, x_t = j) &= (2\pi)^{-m/2} \det(Q_j)^{-1/2} \exp \left\{ -\frac{1}{2} (y_t - \Lambda_j f_t)' Q_j^{-1} (y_t - \Lambda_j f_t) \right\} \times \\ &\quad (2\pi)^{-n/2} \det(V_{t|t-1}^j)^{-1/2} \exp \left\{ -\frac{1}{2} (f_t - \hat{f}_{t|t-1}^j)' (V_{t|t-1}^j)^{-1} (f_t - \hat{f}_{t|t-1}^j) \right\} / \\ &\quad (2\pi)^{m/2} \det(\Omega_{t|t-1}^j)^{1/2} \exp \left\{ -\frac{1}{2} (y_t - \Lambda_j \hat{f}_{t|t-1}^j)' (\Omega_{t|t-1}^j)^{-1} (y_t - \Lambda_j \hat{f}_{t|t-1}^j) \right\}. \end{aligned}$$

Consider the terms inside the exponentials. Straightforward but tedious algebra shows that

$$\begin{aligned} (f_t - \tilde{f})' \tilde{V}^{-1} (f_t - \tilde{f}) &= (y_t - \Lambda_j f_t)' Q_j^{-1} (y_t - \Lambda_j f_t) + (f_t - \hat{f}_{t|t-1}^j)' (V_{t|t-1}^j)^{-1} (f_t - \hat{f}_{t|t-1}^j) \\ &\quad - (y_t - \Lambda_j \hat{f}_{t|t-1}^j)' (\Omega_{t|t-1}^j)^{-1} (y_t - \Lambda_j \hat{f}_{t|t-1}^j) \end{aligned}$$

where

$$\begin{aligned} \tilde{f} &= \hat{f}_{t|t-1}^j + V_{t|t-1}^j \Lambda_j' (\Omega_{t|t-1}^j)^{-1} (y_t - \Lambda_j \hat{f}_{t|t-1}^j) \\ \tilde{V} &= V_{t|t-1}^j - V_{t|t-1}^j \Lambda_j' (\Omega_{t|t-1}^j)^{-1} \Lambda_j V_{t|t-1}^j. \end{aligned}$$

Moreover,

$$\det(\tilde{V}) = \frac{\det(Q_j) \det(V_{t|t-1}^j)}{\det(\Omega_{t|t-1}^j)}.$$

To see this, consider

$$\begin{aligned} \Lambda_j \tilde{V} &= \Lambda_j V_{t|t-1}^j - \Lambda_j V_{t|t-1}^j \Lambda_j' (\Omega_{t|t-1}^j)^{-1} \Lambda_j V_{t|t-1}^j \\ &= \Lambda_j V_{t|t-1}^j - (\Omega_{t|t-1}^j - Q_j) (\Omega_{t|t-1}^j)^{-1} \Lambda_j V_{t|t-1}^j \\ &= Q_j (\Omega_{t|t-1}^j)^{-1} \Lambda_j V_{t|t-1}^j. \end{aligned}$$

Therefore

$$\det(\Lambda_j \tilde{V}) = \det(Q_j (\Omega_{t|t-1}^j)^{-1} \Lambda_j V_{t|t-1}^j).$$

Using standard properties of the determinant gives the desired result. Therefore, $\Pr(f_t|Y^t, x_t = j)$ is a Gaussian distribution with the proposed mean and covariance matrix.

Table 1: 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 5				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 2: Yield Estimation Results with Full-Sample

	Model (1)	Model (2)		Model (3)	
λ	0.0777 (0.0021)	0.1307 (0.0034)	0.0520 (0.0017)	0.1185 (0.0028)	0.0485 (0.0018)
μ_1	0.0675 (0.0691)	0.0535 (0.0735)		0.0799 (0.0706)	
μ_2	0.1887 (0.1386)	0.2442 (0.1477)		-0.0759 (0.0432)	
μ_3	-0.2220 (0.2015)	-0.6969 (0.1858)		-0.0972 (0.0480)	
$A(1, 1)$	0.9957 (0.0177)	1.0009 (0.0086)		0.9904 (0.0081)	
$A(1, 2)$	0.0285 (0.0211)	0.0343 (0.0086)			
$A(1, 3)$	-0.0222 (0.0232)	-0.0244 (0.0124)			
$A(2, 1)$	-0.0306 (0.0260)	-0.0435 (0.0174)			
$A(2, 2)$	0.9389 (0.0309)	0.9303 (0.0174)		0.9506 (0.0167)	
$A(2, 3)$	0.0393 (0.0084)	0.0497 (0.0250)			
$A(3, 1)$	0.0242 (0.0113)	0.0856 (0.0210)			
$A(3, 2)$	0.0229 (0.0305)	0.0111 (0.0211)			
$A(3, 3)$	0.8438 (0.0186)	0.7760 (0.0348)		0.7533 (0.0338)	
$H(1, 1)$	0.0947 (0.0084)	0.0992 (0.0090)		0.1024 (0.0087)	
$H(1, 2)$	-0.0140 (0.0113)	-0.0287 (0.0122)			
$H(1, 3)$	0.0438 (0.0186)	0.0533 (0.0176)			
$H(2, 1)$	-0.0140 (0.0113)	-0.0287 (0.0122)			
$H(2, 2)$	0.3822 (0.0305)	0.4041 (0.0332)		0.3965 (0.0323)	
$H(2, 3)$	0.0094 (0.0344)	-0.0074 (0.0380)			
$H(3, 1)$	0.0438 (0.0186)	0.0533 (0.0176)			
$H(3, 2)$	0.0094 (0.0344)	-0.0074 (0.0380)			
$H(3, 3)$	0.8007 (0.0813)	0.8563 (0.0922)		1.0854 (0.0964)	
p_{00}		0.8895 (0.0237)		0.9753 (0.0068)	
p_{11}		0.9189 (0.0202)		0.9167 (0.0256)	
Log likelihood	8594.3195	8771.2319		8732.4813	
AIC	-17144.639	-17492.4638		-17438.9626	
BIC	-17059.89055	-17396.15874		-17388.88397	

Table 3: Yield Estimation Results with Full-Sample (Continuation)

	Model (4)		Model (5)	
λ	0.1311 (0.0035)	0.0527 (0.0016)	0.1274 (0.0035)	0.0530 (0.0019)
μ_1	0.0557 (0.0733)	0.0604 (0.0813)	0.0462 (0.0839)	0.1225 (0.1416)
μ_2	0.2818 (0.1463)	0.0789 (0.1637)	0.1102 (0.0404)	-0.6930 (0.0870)
μ_3	-0.5333 (0.1941)	-1.0119 (0.2081)	0.0165 (0.0745)	-0.0895 (0.0817)
$A(1, 1)$	0.9998 (0.0087)		0.9950 (0.0097)	0.9840 (0.0161)
$A(1, 2)$	0.0339 (0.0092)			
$A(1, 3)$	-0.0195 (0.0127)			
$A(2, 1)$	-0.0380 (0.0175)			
$A(2, 2)$	0.9130 (0.0188)		0.9958 (0.0067)	0.7776 (0.0276)
$A(2, 3)$	0.0381 (0.0251)			
$A(3, 1)$	0.0898 (0.0228)			
$A(3, 2)$	-0.0201 (0.0235)			
$A(3, 3)$	0.7338 (0.0370)		0.8276 (0.0513)	0.7500 (0.0519)
$H(1, 1)$	0.0968 (0.0088)		0.1096 (0.0093)	
$H(1, 2)$	-0.0292 (0.0119)			
$H(1, 3)$	0.0509 (0.0176)			
$H(2, 1)$	-0.0292 (0.0119)			
$H(2, 2)$	0.4027 (0.0329)		0.3390 (0.0260)	
$H(2, 3)$	-0.0020 (0.0372)			
$H(3, 1)$	0.0509 (0.0176)			
$H(3, 2)$	-0.0020 (0.0372)			
$H(3, 3)$	0.8726 (0.0922)		0.9669 (0.0933)	
p_{00}	0.9057 (0.0225)		0.9311 (0.0171)	
p_{11}	0.9156 (0.0206)		0.9137 (0.0223)	
Log likelihood	8779.3223		8764.6252	
AIC	-17502.6446		-17491.2504	
BIC	-17394.78293		-17418.05855	

Table 4: Yield Estimation Results with Short-Sample

	Model (1)	Model (2)		Model (3)	
λ	0.0862 (0.0028)	0.1409 (0.0048)	0.0560 (0.0026)	0.1193 (0.0035)	0.0487 (0.0025)
μ_1	0.1303 (0.1041)	0.1137 (0.1080)		0.1821 (0.1020)	
μ_2	0.3424 (0.2227)	0.4694 (0.2320)		-0.0955 (0.0556)	
μ_3	-0.5310 (0.3293)	-0.5115 (0.3510)		-0.1129 (0.0663)	
$A(1, 1)$	0.9904 (0.0221)	0.9952 (0.0119)		0.9812 (0.0109)	
$A(1, 2)$	0.0337 (0.0277)	0.0344 (0.0102)			
$A(1, 3)$	-0.0225 (0.0350)	-0.0280 (0.0154)			
$A(2, 1)$	-0.0464 (0.0330)	-0.0677 (0.0256)			
$A(2, 2)$	0.9330 (0.0413)	0.9269 (0.0222)		0.9492 (0.0205)	
$A(2, 3)$	0.0507 (0.0113)	0.0856 (0.0333)			
$A(3, 1)$	0.0547 (0.0160)	0.0814 (0.0373)			
$A(3, 2)$	0.0474 (0.0462)	0.0624 (0.0347)			
$A(3, 3)$	0.8052 (0.0267)	0.6894 (0.0502)		0.7350 (0.0432)	
$H(1, 1)$	0.1083 (0.0113)	0.1142 (0.0122)		0.1177 (0.0119)	
$H(1, 2)$	0.0035 (0.0160)	-0.0140 (0.0178)			
$H(1, 3)$	0.0562 (0.0267)	0.0828 (0.0295)			
$H(2, 1)$	0.0035 (0.0160)	-0.0140 (0.0178)			
$H(2, 2)$	0.4952 (0.0462)	0.5269 (0.0511)		0.5169 (0.0494)	
$H(2, 3)$	0.0220 (0.0532)	-0.0016 (0.0646)			
$H(3, 1)$	0.0562 (0.0267)	0.0828 (0.0295)			
$H(3, 2)$	0.0220 (0.0532)	-0.0016 (0.0646)			
$H(3, 3)$	1.0738 (0.1276)	1.2613 (0.1472)		1.3969 (0.1504)	
p_{11}		0.8864 (0.0309)		0.9699 (0.0099)	
p_{22}		0.9348 (0.0236)		0.9232 (0.0290)	
Log likelihood	5766.4026	5869.0481		5842.4822	
AIC	-11488.806	-11688.096		-11658.964	
BIC	-11411.071	-11599.761		-11613.030	

Table 5: Yield Estimation Results with Short-Sample (Continuation)

	Model (4)		Model (5)	
λ	0.1409 (0.0049)	0.0575 (0.0028)	0.1364 (0.0049)	0.0579 (0.0024)
μ_1	0.1452 (0.1173)	0.1359 (0.1052)	0.1658 (0.1646)	0.2148 (0.1526)
μ_2	0.8964 (0.2305)	0.4076 (0.2198)	0.1618 (0.0668)	-0.6901 (0.1481)
μ_3	0.0852 (0.3594)	-0.3581 (0.3117)	0.2731 (0.1268)	-0.0625 (0.097)
$A(1, 1)$	0.9919 (0.0119)		0.9877 (0.0171)	0.9734 (0.0167)
$A(1, 2)$	0.0334 (0.0120)			
$A(1, 3)$	-0.0202 (0.0149)			
$A(2, 1)$	-0.0891 (0.0237)			
$A(2, 2)$	0.8770 (0.0239)		0.9925 (0.0311)	0.7772 (0.0438)
$A(2, 3)$	0.0648 (0.0306)			
$A(3, 1)$	0.0370 (0.0369)			
$A(3, 2)$	0.0252 (0.0393)			
$A(3, 3)$	0.6829 (0.0511)		0.7452 (0.0735)	0.7164 (0.0650)
$H(1, 1)$	0.1148 (0.0123)		0.1245 (0.0126)	
$H(1, 2)$	-0.0132 (0.0167)			
$H(1, 3)$	0.0647 (0.0287)			
$H(2, 1)$	-0.0132 (0.0167)			
$H(2, 2)$	0.4902 (0.0474)		0.4473 (0.0438)	
$H(2, 3)$	-0.0357 (0.0565)			
$H(3, 1)$	0.0647 (0.0287)			
$H(3, 2)$	-0.0357 (0.0565)			
$H(3, 3)$	1.2655 (0.1472)		1.2474 (0.1448)	
p_{11}	0.8760 (0.0326)		0.8620 (0.0323)	
p_{22}	0.9335 (0.0211)		0.9240 (0.0233)	
Log likelihood	5877.6828		5870.6989	
AIC	-11699.366		-11703.398	
BIC	-11600.431		-11636.264	

Table 6: Forecasting 1 month ahead

Maturity	MSE: 1 month ahead					MSE percentage of times: 1 month ahead				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
3	0.0396	0.0351*	0.0282*	0.0380**	0.0408	16 %	15 %	40 %	13 %	14 %
6	0.0397	0.0395	0.0341*	0.0427	0.0431	14 %	15 %	36 %	11 %	21 %
9	0.0487	0.0480**	0.0410*	0.0520	0.0506	8 %	17 %	34 %	14 %	25 %
12	0.0534	0.0524**	0.0506*	0.0564	0.0631	11 %	23 %	25 %	15 %	23 %
15	0.0617	0.0619	0.0617	0.0649	0.0742	10 %	30 %	16 %	14 %	27 %
18	0.0658	0.0667	0.0644**	0.0708	0.0762	15 %	28 %	13 %	13 %	29 %
21	0.0720	0.0736	0.0692*	0.0778	0.0803	19 %	26 %	14 %	9 %	30 %
24	0.0772	0.0797	0.0731*	0.0828	0.0820	17 %	22 %	17 %	11 %	29 %
30	0.0746	0.0778	0.0719*	0.0813	0.0792	15 %	26 %	17 %	9 %	30 %
36	0.0734	0.0774	0.0719**	0.0803	0.0764	16 %	25 %	19 %	5 %	33 %
48	0.0755	0.0801	0.0745	0.0823	0.0773	14 %	25 %	16 %	2 %	41 %
60	0.0798	0.0852	0.0784	0.0852	0.0783**	17 %	19 %	20 %	5 %	36 %
72	0.0720	0.0770	0.0716**	0.0779	0.0723	20 %	16 %	16 %	5 %	40 %
84	0.0752	0.0799	0.0739	0.0806	0.0751	25 %	14 %	20 %	2 %	38 %
96	0.0702	0.0741	0.0686**	0.0754	0.0705	30 %	10 %	19 %	4 %	34 %
108	0.0714	0.0749	0.0694**	0.0761	0.0712	34 %	10 %	15 %	4 %	34 %
120	0.0720	0.0770	0.0715	0.0777	0.0727	33 %	17 %	14 %	3 %	30%

Table 7: Forecasting 3 months ahead

Maturity	MSE: 3 months ahead					MSE percentage of times: 3 months ahead				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
3	0.1428	0.1388**	0.0932*	0.1751	0.1547	11 %	13 %	38 %	15 %	21 %
6	0.1726	0.1689**	0.1345*	0.2122	0.1861	16 %	9 %	35 %	13 %	25 %
9	0.2053	0.1986*	0.1607*	0.2446	0.2131	13 %	16 %	30 %	14 %	25 %
12	0.2185	0.2099*	0.1909*	0.2581	0.2498	16 %	21 %	26 %	8 %	27 %
15	0.2430	0.2330*	0.2230*	0.2816	0.2830	7 %	38 %	20 %	4 %	29 %
18	0.2616	0.2518*	0.2352*	0.3010	0.2899	11 %	35 %	21 %	5 %	25 %
21	0.2808	0.2714*	0.2488*	0.3188	0.2986	10 %	34 %	21 %	5 %	27 %
24	0.2948	0.2885**	0.2585*	0.3322	0.3019	16 %	27 %	21 %	5 %	28 %
30	0.2873	0.2842**	0.2569*	0.3253	0.2919	16 %	27 %	22 %	4 %	28 %
36	0.2848	0.2857	0.2592*	0.3227	0.2857	16 %	26 %	23 %	2 %	30 %
48	0.2810	0.2863	0.2606*	0.3168	0.2790	17 %	25 %	27 %	1 %	28 %
60	0.2877	0.2995	0.2697*	0.3203	0.2800**	19 %	20 %	28 %	3 %	28 %
72	0.2623	0.2734	0.2496*	0.2930	0.2593	27 %	16 %	32 %	2 %	21 %
84	0.2655	0.2740	0.2499*	0.2913	0.2613	26 %	15 %	35 %	2 %	20 %
96	0.2446	0.2512	0.2283*	0.2667	0.2393**	28 %	14 %	29 %	4 %	22 %
108	0.2459	0.2516	0.2276*	0.2646	0.2385*	26 %	14 %	30 %	4 %	23 %
120	0.2442	0.2543	0.2288*	0.2638	0.2386**	27 %	11 %	28 %	5 %	26 %

Table 8: Forecasting 6 months ahead

Maturity	MSE: 6 months ahead					MSE percentage of times: 6 months ahead				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
3	0.3803	0.3878	0.2138*	0.4929	0.3253*	8 %	14 %	39 %	16 %	21 %
6	0.4159	0.4273	0.2827*	0.5304	0.3727*	10 %	15 %	40 %	14 %	19 %
9	0.4518	0.4603	0.3184*	0.5566	0.4027*	14 %	14 %	30 %	16 %	23 %
12	0.4675	0.4668	0.3740*	0.5709	0.4668	10 %	20 %	32 %	14 %	22 %
15	0.4886	0.4816**	0.4139*	0.5858	0.5080	8 %	26 %	27 %	13 %	25 %
18	0.5207	0.5135**	0.4381*	0.6134	0.5212	5 %	26 %	26 %	17 %	23 %
21	0.5521	0.5455**	0.4637*	0.6381	0.5383*	7 %	25 %	26 %	13 %	28 %
24	0.5790	0.5780	0.4860*	0.6626	0.5498*	9 %	25 %	23 %	10 %	30 %
30	0.5695	0.5735	0.4912*	0.6486	0.5412**	11 %	22 %	27 %	7 %	30 %
36	0.5710	0.5812	0.5052*	0.6471	0.5434*	13 %	22 %	28 %	5 %	29 %
48	0.5715	0.5880	0.5224*	0.6408	0.5481*	13 %	25 %	30 %	4 %	26 %
60	0.6022	0.6296	0.5633*	0.6694	0.5806*	20 %	20 %	26 %	5 %	27 %
72	0.5561	0.5820	0.5302*	0.6179	0.5458**	21 %	21 %	26 %	3 %	27 %
84	0.5579	0.5795	0.5333*	0.6117	0.5514**	21 %	21 %	27 %	3 %	26 %
96	0.5242	0.5430	0.4981*	0.5712	0.5166**	21 %	21 %	28 %	3 %	25 %
108	0.5255	0.5442	0.4938*	0.5670	0.5137	23 %	17 %	29 %	4 %	23 %
120	0.5379	0.5643	0.5070*	0.5820	0.5269**	25 %	17 %	25 %	3 %	28 %

Table 9: Forecasting 12 months ahead

Maturity	MSE: 12 months ahead					MSE percentage of times: 12 months ahead				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
3	1.0213	1.1396	0.6069*	1.3496	0.7638*	7 %	21 %	20 %	9 %	41 %
6	0.9995	1.1602	0.7087*	1.3169	0.8321*	10 %	15 %	16 %	13 %	44 %
9	0.9976	1.1765	0.7603*	1.2937	0.8592 *	7 %	17 %	20 %	9 %	45 %
12	0.9502	1.1186	0.8129*	1.2243	0.9058**	7 %	16 %	22 %	10 %	42 %
15	0.9344	1.0871	0.8601*	1.1924	0.9486	10 %	19 %	21 %	7 %	41 %
18	0.9647	1.1192	0.9018*	1.2031	0.9688	14 %	17 %	23 %	5 %	38 %
21	0.9950	1.1493	0.9389*	1.2151	0.9866	14 %	17 %	22 %	7 %	38 %
24	1.0338	1.1957	0.9802*	1.2416	1.0078**	13 %	17 %	21 %	8 %	39 %
30	1.0244	1.1794	0.9993**	1.2071	1.0033**	11 %	17 %	22 %	8 %	39 %
36	1.0305	1.1809	1.0254	1.1961	1.0130**	13 %	15 %	22 %	10 %	38 %
48	1.0320	1.1662	1.0524	1.1696	1.0257**	14 %	15 %	23 %	7 %	39 %
60	1.0985	1.2297	1.1362	1.2273	1.1005	19 %	14 %	23 %	8 %	34 %
72	1.0296	1.1417	1.0842	1.1464	1.0492	20 %	17 %	27 %	7 %	27 %
84	1.0123	1.1085	1.0678	1.1183	1.0400	25 %	15 %	29 %	8 %	21 %
96	0.9703	1.0546	1.0137	1.0682	0.9918	27 %	15 %	30 %	7 %	19 %
108	0.9811	1.0620	1.0087	1.0751	0.9930	28 %	13 %	30 %	4 %	22 %
120	1.0405	1.1270	1.0613	1.1433	1.0516	27 %	11 %	30 %	5 %	23 %

Table 10: Confusion Rates: 1 month ahead

Maturity	Percentage of Confusion rate: 1 Month ahead				
	Model 1	Model 2	Model 3	Model 4	Model 5
3	41 %	45 %	45 %	44 %	42 %
6	40 %	42 %	45 %	46 %	46 %
9	41 %	41 %	44 %	42 %	41 %
12	46 %	47 %	47 %	44 %	44 %
15	42 %	41 %	41 %	40 %	38 %
18	36 %	39 %	38 %	36 %	36 %
21	35 %	36 %	35 %	35 %	35 %
24	35 %	36 %	35 %	36 %	35 %
30	36 %	38 %	34 %	36 %	36 %
36	36 %	36 %	35 %	36 %	38 %
48	38 %	38 %	36 %	36 %	39 %
60	38 %	39 %	36 %	36 %	39 %
72	36 %	38 %	34 %	36 %	36 %
84	38 %	36 %	35 %	35 %	35 %
96	38 %	36 %	35 %	35 %	35 %
108	36 %	35 %	34 %	34 %	34 %
120	39 %	39 %	38 %	39 %	38 %

Table 11: Confusion Rates: 3 months ahead

Maturity	Percentage of Confusion rate: 3 Months ahead				
	Model 1	Model 2	Model 3	Model 4	Model 5
3	46 %	50 %	46 %	46 %	44 %
6	44 %	44 %	47 %	47 %	46 %
9	38 %	41 %	42 %	42 %	41 %
12	46 %	46 %	50 %	47 %	44 %
15	45 %	40 %	42 %	45 %	46 %
18	47 %	46 %	45 %	50 %	48 %
21	47 %	46 %	47 %	46 %	50 %
24	45 %	46 %	45 %	46 %	45 %
30	45 %	42 %	46 %	47 %	45 %
36	45 %	42 %	45 %	46 %	44 %
48	44 %	45 %	44 %	45 %	41 %
60	46 %	46 %	44 %	45 %	45 %
72	47 %	46 %	41 %	44 %	44 %
84	46 %	45 %	42 %	45 %	44 %
96	47 %	46 %	44 %	46 %	45 %
108	51 %	52 %	46 %	50 %	48 %
120	45 %	48 %	42 %	48 %	44 %

Table 12: Confusion Rates: 6 months ahead

	Percentage of Confusion rate: 6 Months ahead				
Maturity	Model 1	Model 2	Model 3	Model 4	Model 5
3	50 %	50 %	48 %	51 %	42 %
6	52 %	57 %	48 %	46 %	40 %
9	42 %	51 %	44 %	42 %	42 %
12	47 %	50 %	47 %	47 %	45 %
15	47 %	52 %	44 %	46 %	44 %
18	47 %	47 %	45 %	42 %	44 %
21	48 %	47 %	46 %	44 %	45 %
24	44 %	46 %	44 %	41 %	45%
30	44 %	41 %	45 %	41 %	41 %
36	42 %	39 %	44 %	40 %	40 %
48	46 %	44 %	45 %	45 %	42 %
60	48 %	48 %	46 %	47 %	45 %
72	47 %	47 %	46 %	46 %	45 %
84	47 %	47 %	47 %	46 %	47 %
96	47 %	48 %	48 %	47 %	48 %
108	42 %	42 %	44 %	44 %	45 %
120	39 %	40 %	42 %	42 %	44%

Table 13: Confusion Rates: 12 months ahead

	Percentage of Confusion rate: 12 Months ahead				
Maturity	Model 1	Model 2	Model 3	Model 4	Model 5
3	47 %	57 %	57 %	58 %	53 %
6	54 %	52 %	54 %	54 %	54 %
9	65 %	61 %	57 %	60 %	53 %
12	57 %	57 %	54 %	57 %	48 %
15	57 %	59 %	53 %	53 %	51 %
18	50 %	55 %	54 %	53 %	51 %
21	53 %	55 %	57 %	57 %	54 %
24	53 %	55 %	58 %	57 %	55 %
30	51 %	52 %	55 %	53 %	52 %
36	51 %	51 %	54 %	54 %	51 %
48	54 %	48 %	57 %	58 %	53 %
60	58 %	53 %	57 %	60 %	54 %
72	55 %	51 %	58 %	57 %	54 %
84	61 %	57 %	61 %	60 %	58 %
96	55 %	54 %	58 %	57 %	54 %
108	57 %	53 %	57 %	55 %	53 %
120	55 %	53 %	58 %	54 %	54 %

Figure 1: Stylized facts in booms and recessions

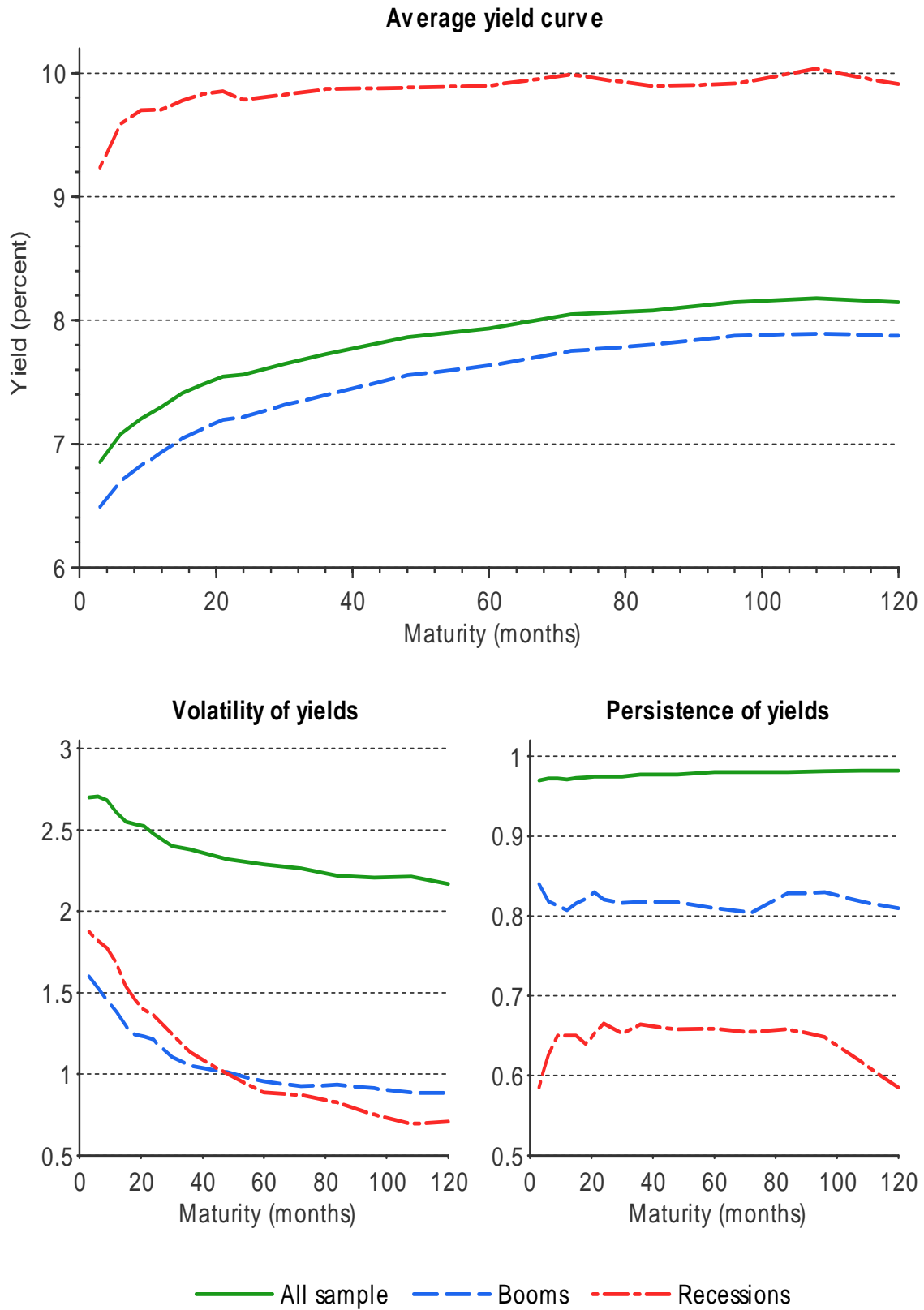


Figure 2: Factor loadings

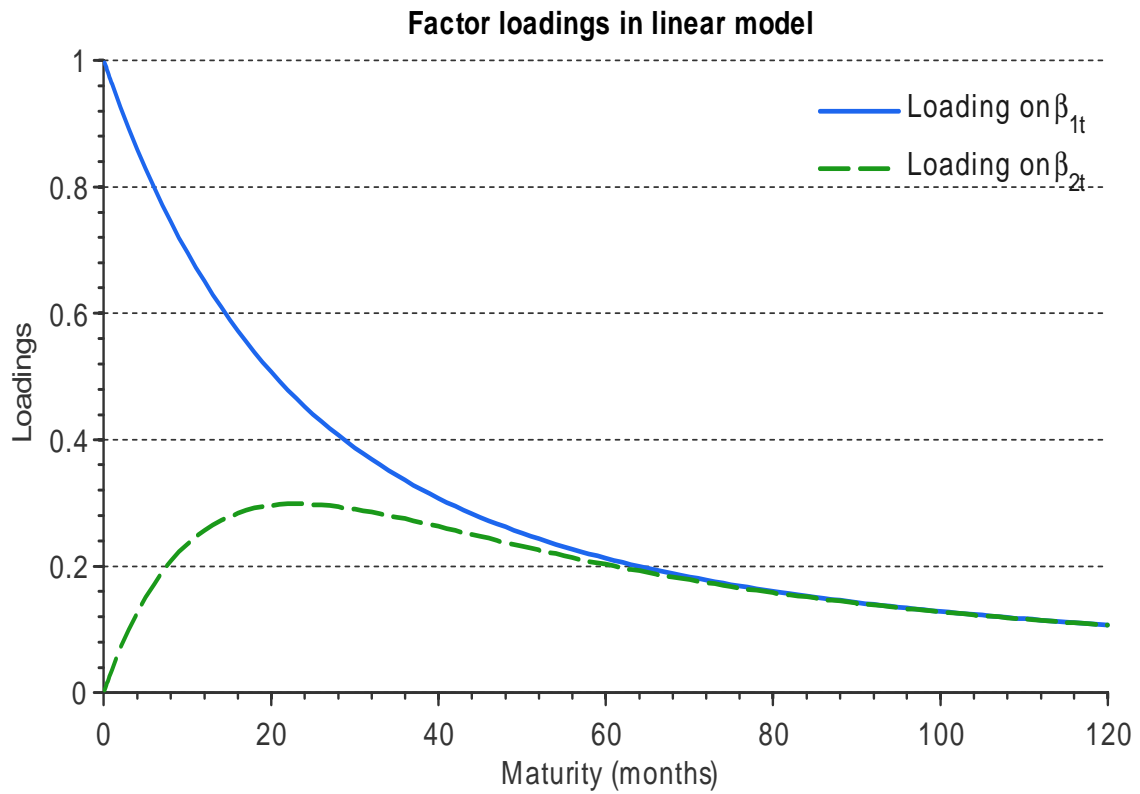
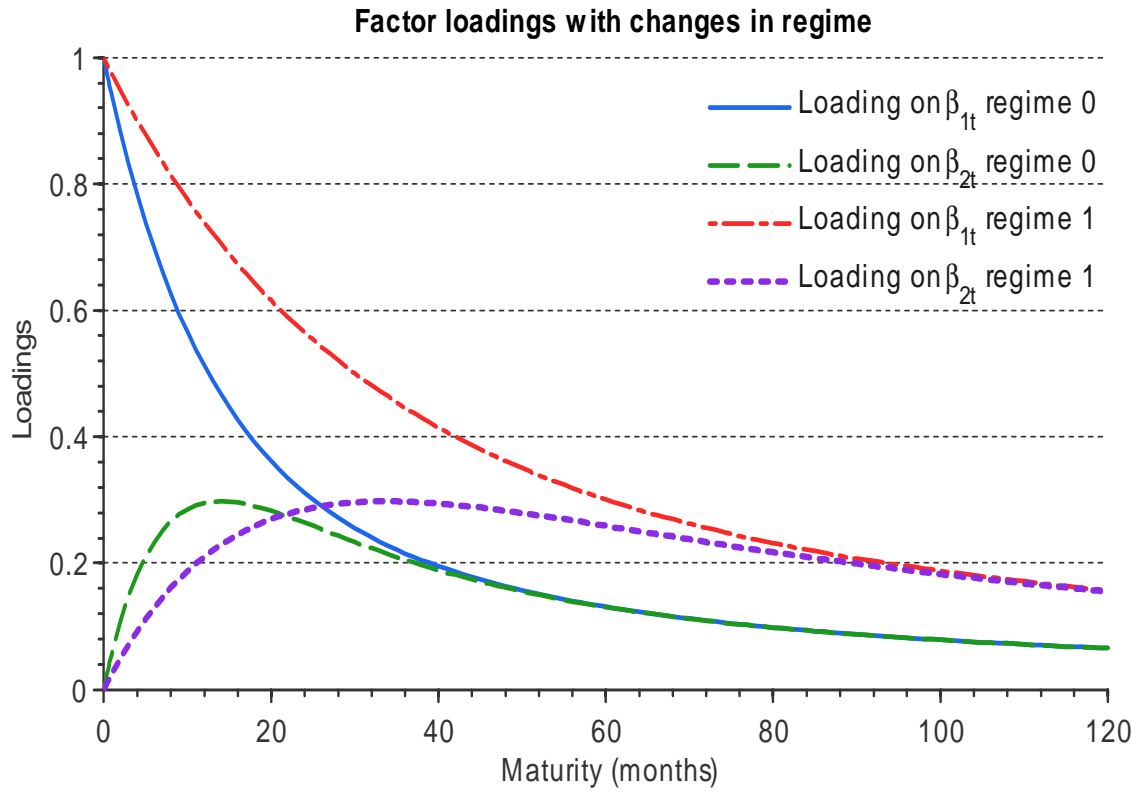


Figure 3: Estimated latent factors and level, slope, and curvature

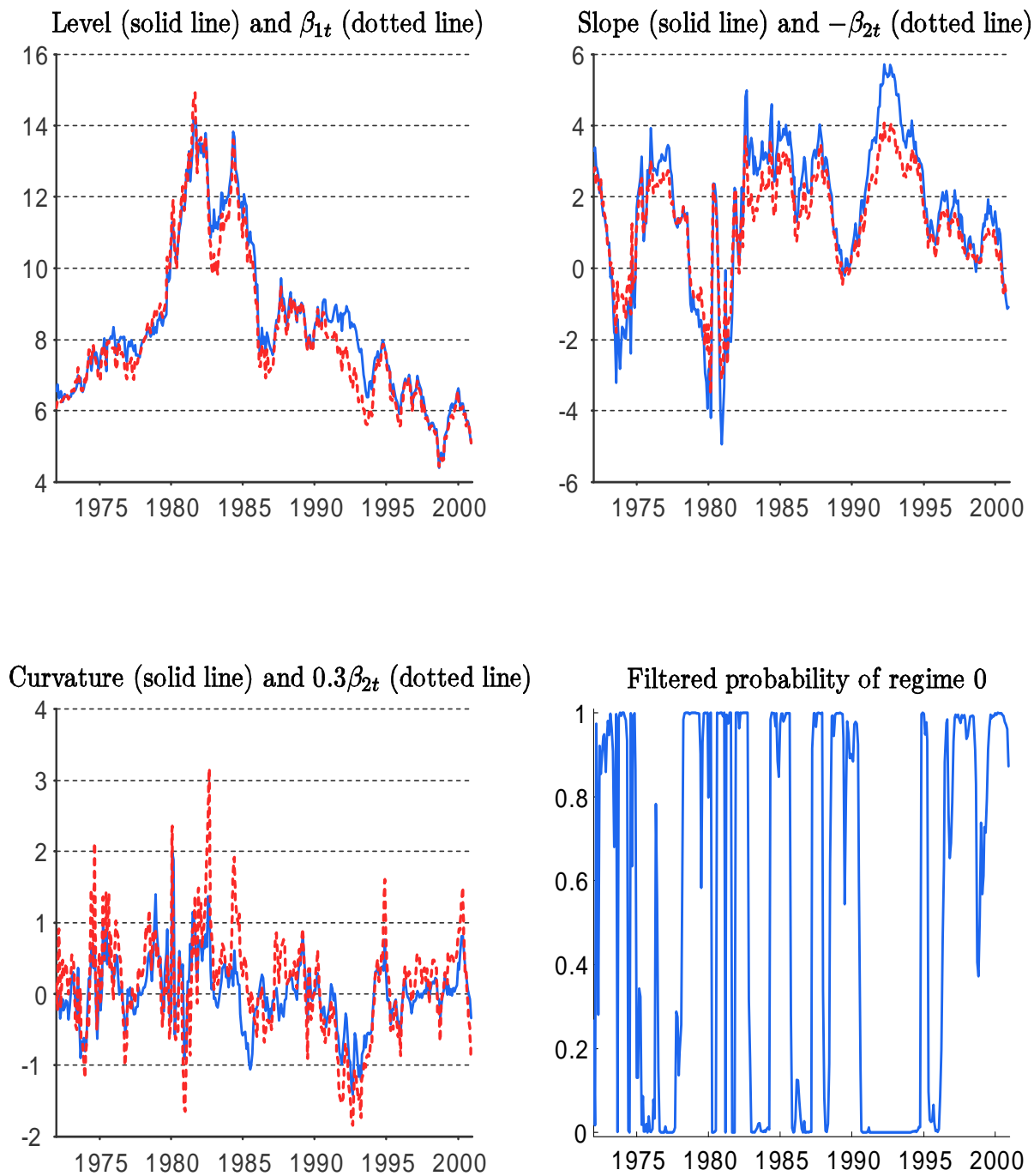


Figure 4: Regime specific level, volatility, and persistence of yields

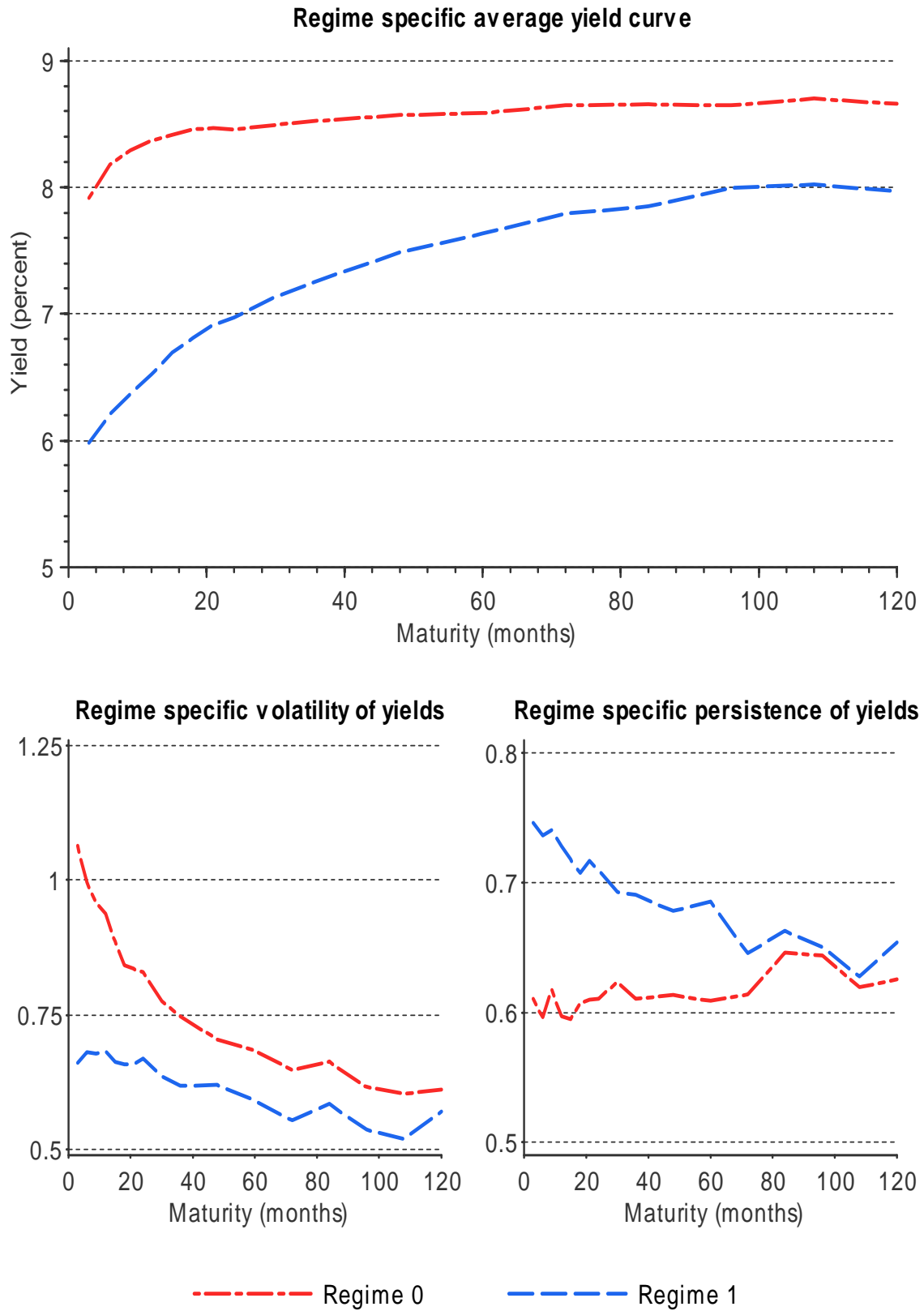


Figure 5: Actual and fitted yield curve for selected dates

