DISCUSSION PAPER SERIES

No. 4165

ON MODEL SELECTION AND MARKOV SWITCHING: A EMPIRICAL EXAMINATION OF TERM STRUCTURE MODELS WITH REGIME SHIFTS

John Driffill, Turalay Kenc, Martin Sola
and Fabio Spagnolo

FINANCIAL ECONOMICS and
INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP4165.asp**

ISSN 0265-8003

ON MODEL SELECTION AND MARKOV SWITCHING: A EMPIRICAL EXAMINATION OF TERM STRUCTURE MODELS WITH REGIME SHIFTS

John Driffill, Birkbeck College, London and CEPR
Turalay Kenc, Imperial College, London
Martin Sola, Birkbeck College, London
Fabio Spagnolo, Brunel University

Discussion Paper No. 4165
January 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS and INTERNATIONAL MACROECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: John Driffill, Turalay Kenc, Martin Sola and Fabio Spagnolo

CEPR Discussion Paper No. 4165

January 2004

ABSTRACT

On Model Selection and Markov Switching: A Empirical Examination of Term Structure Models with Regime Shifts*

We examine several continuous-time term-structure models, in which the short rate is subject to discrete shifts. Our empirical analysis suggests that inquiring which parameters of the short-term interest rate equation are allowed to switch is crucial, as failing to do so may result in switching pricing models that produce no improvement (in terms of pricing) with respect to models which do not allow for regime switching, even when there are clear breaks in the data.

JEL Classification: E43 and G12
Keywords: bond yields, regime switching, stochastic discount factor/pricing kernel and term structure of interest rates

John Driffill
School of Economics, Mathmatics and Statistics
Birkbeck College
Malet Street
LONDON
WC1E 7HX
Tel: (44 20) 7631 6417
Fax: (44 20) 7631 6416
Email: jdriffill@econ.bbk.ac.uk

Turalay Kenc
The Business School
Imperial College London
53 Prince's Gate, Exhibition Road
London
SW7 2AZ
Tel: (44 20) 7594 9212
Email: t.kenc@ic.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=102049

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=136504

Martin Sola
School of Economics, Mathmatics
and Statistics
Birkbeck College
Malet Street
LONDON
WC1E 7HX
Tel: (44 20) 7631 6411
Fax: (44 20) 7631 6416
Email: msola@econ.bbk.ac.uk

Fabio Spagnolo
Department of Economics & Finance
Brunel University
Uxbridge
UB8 3PH
Tel: (44 1895) 816383
Fax: (44 1895) 203303
Email: fabio.spagnolo@brunel.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=119222

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=159989

*We thank David Barr, Keith Cuthbertson, Demian Pouzo, Marzia Raybaudi, Daniel Thornton, Giorgio Valente for helpful comments. We are also grateful for comments from seminar participants at the Computing in Economic and Finance Annual Meeting, the Royal Economic Society Meeting, the European Financial Management Association Meeting, the ISMA center at the University of Reading, Imperial College, the University of Kent at Canterbury, the Federal Reserve Bank of St. Louis, and the University of Essex. John Driffill, Martin Sola and Fabio Spagnolo are grateful to the ESRC for support under grant L138251003. The usual disclaimer applies.

Submitted 07 November 2003

1 Introduction

A popular way of characterizing processes which are subject to structural breaks is to assume that the breaks follow a Markov chain as in Hamilton (1988, 1989). Even though there is an extensive literature that uses this technique, only few papers systematically study the specification of the switching regression. Questions such as how to select (i) the numbers of states [see Hansen (1992, 1996), Garcia (1998), and Psaradakis and Spagnolo (2003)], (ii) the number of lags which should be included in the switching regression [see Kapetanios (2001)], or (iii) which parameters are supposed to be allowed to switch [see Hall and Sola (1993)] have attracted comparatively less attention in the literature.

The importance of each of these questions depends very much on the application at hand. For example, Kim and Piger (2002) found that ignoring some of the dynamics of the output growth does not affect, and sometimes even improves, the ability of the filter to correctly separate booms and recessions. On the other hand, the correct specification of the switching process seems to be of great importance when the process is used for forecasting purposes as in all the rational expectations applications. Different specifications of the switching driving process will typically imply different forecasts and different pricing equations. This will also potentially affect any conclusion about the validity of the theory under scrutiny.[1]

One of the areas where this technique has been widely used is in characterizing the evolution of the short/long-term interest rates as a process subject to Markov changes. There is increasing evidence in the literature that both short and long-term interest rates can be characterized as a stochastic process subject to regime switches [see, for example, Hamilton (1988), Sola and Driffill (1994), Garcia and Perron (1996), Gray (1996), and Dahlquist and Gray (2000)]. In particular, Gray (1996) showed that a time-varying parameter version of the Cox, Ingersoll and Ross (1985; CIR) model best characterizes the US short-term data, and this model has widely been used in the literature. Some of the papers that use the switching CIR-specification of the short-term rate [e.g., Naik and Lee (1997)] only allow to switch the volatility of the short-term interest rate. On the other hand, Bansal and Zhou (2002) allow all the parameters of the short rate to switch [see also Dahlquist and Gray (2000) and Ang and Bekaert (2002)].

A key problem arising in such applications is to assess how to determine which parametrization is the best characterization of the observed data. None of the mentioned papers questioned how different specifications of the CIR model (for the instantaneous interest rate) perform in

[1] The issues studied in this paper are less relevant in cases where the specification of the switching model is dictated by economic reasons (identifying assumptions).

terms of fit and, most importantly, in terms of forecasting (which is crucial for pricing bonds). This paper attempts to fill this gap by evaluating how different parameterizations of the switching process for the short-term interest rates affect the bond prices.[2]

The idea is to obtain the best model for bond pricing (the model with generated prices closest to those observed in the data), and use the pricing performance as a criterion to assess which parameterization of the short-term interest rate is preferred. Our approach is based on recursively estimating the different parameterizations of the switching CIR process for the short-term interest rate described above, and use the results to price bonds for different maturities. In this way we generate a series of prices which are then compared with the actual prices in terms of fit, and in terms of the shape of the term structure at different points in time.[3]

These results are then compared with those obtained using standard likelihood ratio tests, complexity-penalized likelihood criteria[4], and out-of-sample multi-step ahead forecasting (for the short-term interest rate). Interestingly, the results obtained for the whole sample, using standard likelihood ratio tests and goodness of fit criteria, do not coincide with those obtained using the pricing strategies described above. The specifications that produce the best prices tend to be those with better forecasting performance. This highlights (in particular for Markov switching models) that the models which provide the best fit do not necessarily provide the best forecast and therefore the best price.

[2]We consider different versions of the CIR short rate process which include: (1) a benchmark case with no regime-switching; models with regime-switching in: (2) volatility; (3) volatility and the speed of adjustment; (4) volatility and the long-run value of the short rate; and (5) volatility, the speed of adjustment and the long-run value of the short rate.

[3]To carry out an extensive analysis of the implications of studying the effects of the choice of the parameters that are allowed to switch and, more importantly, the relevance of the issue, we use a simple Markov switching CIR model. We prefer this to using other switching affine models with more factors (arguably superior in many metrics), because an extensive analysis of the issue at hand, using those models, is more cumbersome. Needless is to say that the point raised in this paper is equally important for other affine switching models.

[4]Such methods have enjoyed much popularity in statistics as a means of choosing among competing models and, under appropriate regularity conditions, are known to be capable of selecting with probability 1 the model with lowest Kullback-Leibler divergence from the data-generating mechanism [Nishii (1988); Sin and White (1996)]. Furthermore, as Granger, King, and White (1995) pointed out, these methods are arguably more appropriate for model selection than procedures based on formal hypothesis testing, partly because, unlike testing, they do not unfairly favor the model chosen to be the null hypothesis. Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999) use complexity-penalized likelihood criteria to select among linear models for prediction of stock returns. The use of formal statistical selection criteria as a means of selecting the number of components in independent and Markov-dependent finite mixture models has been studied by Leroux (1992), Leroux and Puterman (1992) and Ryden (1995).

The plan of the paper is as follows. In Section 2 we discuss how to price bonds when the instantaneous rate switches between two Brownian motions. Sections 3 and 4 describe the data and the estimation results. Section 5 examines the forecast performance of the five competing models for the short-term interest rate. Section 6 considers using bond pricing as model selection criteria for the instantaneous interest rate. Finally, Section 7 summarizes and concludes.

2 Pricing bonds when the instantaneous rate switches between two Brownian motions

We consider a financial market model in which all activity takes place in the time interval $[0, T]$ and operates on a stochastic basis $(\Omega, F, (F_t)_{(t\in[0,\mathcal{T}])}, P)$, where Ω denotes the set of all possible states of the economy, F is the set of distinguishable events in the economy and is a given σ-field of subsets of Ω, $(F_t)_{(t\in[0,\mathcal{T}]}$ (with $F_T = F$) represents a filtration describing the information arrival, and P denotes the probability beliefs of the agents.

It is assumed that the short-term interest rate is driven by a Brownian motion with state dependent drift and diffusion parameters, where the discrete states X take values 0 and 1. The switch between the regimes is governed by a Markov chain with an intensity rate matrix $H = (h_{ij})$:

$$H = \begin{bmatrix} h_{00} & h_{01} \\ h_{10} & h_{11} \end{bmatrix} = \begin{bmatrix} -h_{01} & h_{01} \\ h_{10} & -h_{10} \end{bmatrix}. \tag{1}$$

The probability that a transition occurs from a given source state depends not only on the source state, but also on the length of the observation interval. Then, the probability that a transition occurs from state i (say $i = 0$) to state j (say $j = 1$) in the interval $[t, t + \Delta t)$ is equal to $h_{01}\Delta t + o(\Delta t)$. Similarly, $1 - h_{01}\Delta t + o(\Delta t)$ is the probability that the process remains in state i. Moreover, if the Markov chain is stationary, the long-run probabilities are given by $\pi_\infty(0) = \frac{h_{10}}{h_{01}+h_{10}}$ and $\pi_\infty(1) = \frac{h_{01}}{h_{01}+h_{10}}$, while the conditional probability (given the source state 1) that the process would be in regime 1 at time $t+s$ is given by $\frac{h_{01}}{h_{01}+h_{10}} + \frac{h_{01}\exp[-(h_{10}+h_{01})(s)]}{h_{01}+h_{10}}$. For analytical tractability, it is assumed that the discrete states X_t are independent of the Wiener process $W(t)$. It is also assumed that agents in the financial markets know the actual state of the system X_t.[5]

The short rate process r is thus formally represented by a stochastic differential equation for

[5]However, the econometrician does not observe the actual state, and has to make inferences of it based on the observable history of the system.

which we specify a "regime-switching mean-reverting square root process" (the RSCIR process)[6],

$$dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t). \tag{2}$$

where $dW(t)$ is the increment of a standard Wiener process. The drift term, $\kappa(X_t)[\alpha(X_t) - r(t)]dt$, in (2) captures (by setting $\kappa(X_t) > 0$) the regime-dependent mean-reversion; the parameter $\alpha(X_t)$ (> 0) is the implied regime-dependent long-run mean interest rate; $\kappa(X_t)$ determines the regime-dependent adjustment speed of r toward the long-term mean; and $\sigma(X_t)^2 r$ is the regime-dependent variance of the unexpected interest rate changes. The term $\sigma(X_t)$ is the regime-dependent volatility factor and serves as a scale factor.[7]

In order to price bonds based on the short rate specified in (2), we use a stochastic discount factor process [see Cochrane (2001)] of the form

$$\frac{d\Lambda(t)}{\Lambda(t)} = -r(t)dt - \sigma_\Lambda\sqrt{r(t)}dW(t) \tag{3}$$

where $\Lambda(t)$ is the stochastic discount factor.[8] Note that, for simplicity, it is assumed that uncertainty in the instantaneous interest rate and the stochastic discount rate is governed by the same Brownian motion $Z(t)$.

We assume that there is a market for every bond at every choice of maturity T and the market is arbitrage free. Furthermore, we assume that, for every T, the price of a maturity T-bond has the form

$$P(t,T) = F(t, r(t), X, T), \tag{4}$$

where $F(t, r(t), X, T)$ is a function of four variables and more precisely takes a semi-affine form

$$\ln F(t, r(t), X, T) = A(r, X, T) - B(r, X, T)r \tag{5}$$

where A and B are deterministic functions. This leads to a semi-affine form term structure (SATS).[9]

[6]Evans (2003) and Bansal and Zhou (2002) use a discrete version of the RSCIR process. For the relation between discrete-time and continuous-time single regime models see Sun (1992).

[7]This implies that the volatility of the interest rate is parameterized as a function of interest rate levels and produces conditional heteroskedasticity (which is the cause of the leptokurtosis in the unconditional distribution of changes in the short rate).

[8]This is equivalent to the pricing kernel or state-price deflator [see for example Fisher and Gilles (2000)], which is determined by marginal utility

$$\Lambda(t) = e^{-\delta t}u'(c(t))$$

where c is the consumption and δ is the time preference rate.

[9]Duarte (2000) also uses a similar terminology "semi-affine square-root" model in which he develops a different (flexible) parameterization for the price of risk.

Our aim is to price bonds using (i) the above semi-affine term structure and (ii) the no-arbitrage condition. To do this we make two further assumptions: one about the hedging behavior of investors and the other about the market price of risk. The first assumption is that investors do not hedge the regime switching risk; however they hedge the risk of the continuous changes in short rates. The second is that the market price of risk is the same in different regimes as in Naik and Lee (1997) and Landen (2000). To satisfy the no-arbitrage condition we will use the fundamental pricing equation

$$\mathcal{E}(t)\left(\frac{dF}{F}\right) - \left(\frac{1}{P}\frac{\partial F}{\partial T}\right)dt = rdt - \mathcal{E}(t)\left(\frac{dF}{F}\frac{d\Lambda}{\Lambda}\right) \tag{6}$$

where $\mathcal{E}(t)$ is the expectation operator conditional on information at time t. The left hand side is the expected (holding period) rate of return on the T-Bond. This should be equal to the sum of the risk-free rate and the covariance of the return of the bond with the proportional change in the discount factor [the last term in (6)]. Using Itô's lemma we get the following price dynamics for the T-bond

$$dF = aFdt + bFdW \tag{7}$$

where

$$a = \frac{\kappa(X_t)[\alpha(X_t) - r]F_r + \frac{1}{2}\sigma(X_t)^2 rF_{rr} + \sum_{i=0}^{1}\sum_{j=0}^{1} h_{ij}\Delta F}{F} \tag{8}$$

$$b = \frac{\sigma(X_t)\sqrt{r}F_r}{F} \tag{9}$$

Substituting for a and b in (6) (and cancelling dt), we obtain the differential equation,

$$\kappa(X_t)[\alpha(X_t) - r]F_r + \tfrac{1}{2}\sigma(X_t)^2 rF_{rr} - F_T + \sum_{i=0}^{1}\sum_{j=0}^{1} h_{ij}\Delta F - rF = F_r\sigma\sigma_\Lambda r. \tag{10}$$

The right hand-side of (10) is the risk premium. Setting $\sigma_\Lambda = \lambda$, the market price of risk, and rearranging (10) we obtain the following "term structure equation"

$$(\kappa(X_t)[\alpha(X_t) - r] - \sigma(X_t)\lambda r)F_r + \tfrac{1}{2}\sigma(X_t)^2 rF_{rr} - F_T + \sum_{i=0}^{1}\sum_{j=0}^{1} h_{ij}\Delta F - rF = 0, \tag{11}$$

with the boundary condition.[10]

$$F(T, X, r, T) = 1. \tag{12}$$

[10]Details about the numerical solution of the differential equation are presented in the Appendix A.

3 Data description

We use Duffee (2002) data set (monthly estimates of annualized continuously-compounded zero-coupon US government bond yields) which is based on the Bliss (1997) extension of McCulloch and Kwon (1993). The data set ranges from January 1952 to December 1998.[11] There are 564 monthly observations with 6 maturities: 3, 6 month, and 1, 2, 5, and 10 year.

To find out how the shapes of the yield curves evolve over the sample period we plot the surfaces of the yield curves in Figure 1. Likewise, Table 1 reports the summary statistics for the yields. The yield curve is typically upward sloping, but inverted around 1979 to 1981. The yield volatilities, measured by standard deviations, decrease with the maturities. Figure 1 reveals the widely reported empirical fact that there has been a significant increase in interest rate volatility in the last two decades.

[Figure 1 approximately here]

[Table 1 approximately here]

4 Estimation

To estimate the parameters of the instantaneous interest rate for the regime-switching Cox, Ingersoll and Ross model (RSCIR), we use as a proxy, the 3 month T-Bill yield as in Andersen and Lund (1997) and Duffee (2002).[12] We use quarterly data to avoid potential serial correlation which would be induced by overlapping expectations when higher (than the maturity of the short interest rate) sampling frequency is used. The five models that are specified in Table 2 are estimated for the period 1964:1–1998:4 (note that Table 2 does not include regime switching

[11] There are two widely used data sets of US bond yields: (i) McCulloch and Kwon (1993) who use a cubic spline to construct the yield curve and (ii) Fama and Bliss (1987) who use bootstrap methods to construct the yield curve. The data used in this paper is available on the web page associated with Duffie and Epstein (1992). The address is http://www.haas.berkeley.edu/~duffee/affine.htm.

[12] Since our model is in continuous time, its state variable is the instantaneous interest rate which is unobservable. Chapman, John B. Long, and Pearson (1997) explores this proxy problem and show that it is not economically significant for single-factor affine models. But it can be economically significant when applied to a two-factor affine model and a nonlinear single-factor model. In the literature, the one-month rate is also used a popular proxy for the instantaneous rate [see for example Chan, Karolyi, Longstaff, and Sanders (1992) and Nowman (1997)]. But it is well documented that estimating the model with the one-month rate is relatively difficult. Another argument against using 1-month yields is that they are more likely to be influenced by liquidity needs. For the same reason Bansal and Zhou (2002) use the 6 month T-Bill yield.

models with constant volatility scale factor, σ, because those specifications are strongly rejected by the data).

The estimation of the Markov-switching models can be carried out by using the recursive algorithm discussed in Hamilton (1988, 1989). This gives as a by-product the sample likelihood function which can be maximized numerically with respect to $(\kappa_0, \kappa_1, \alpha_0, \alpha_1, \sigma_0, \sigma_1, p, q)$, subject to the constraint that $p = P(X_t = 1|X_{t-1} = 1)$ and $q = P(X_t = 0|X_{t-1} = 0)$ lie in the open unit interval.

To estimate the RSCIR process (2) we discretize it using the exact discretization method as follows:

$$r_{t+\Delta} = e^{-\kappa(X_t)\Delta t} r_t + (1 - e^{-\kappa(X_t)\Delta t})\alpha(X_t) + \sigma(X_t)\sqrt{r_t}\sqrt{\frac{1 - e^{-2\kappa(X_t)\Delta t}}{2\kappa(X_t)}}\,\epsilon_{t+\Delta}, \tag{13}$$

where $\epsilon_{t+\Delta}$ is assumed to be i.i.d. standard normal in the model. Notice that as Δt becomes small equation (13) reduces to the Euler discretization equation

$$r_{t+\Delta} = \varphi(X_t) r_t - (1 - \varphi(X_t))\alpha(X_t) + \sigma(X_t)\sqrt{r_t \Delta t}\,\epsilon_{t+\Delta}, \tag{14}$$

where $\varphi \approx 1 - \kappa\Delta t$ captures the autocorrelation of the interest rate.

In Table 3, we report Gaussian standard pseudo-maximum likelihood (S–PML) estimates of the parameters of the models presented in Table 2, along with the corresponding asymptotic standard errors.[13] The estimated parameters for the Markov switching models, show significant evidence of shifts between regimes. For example, estimates of the long-run rate range from 0.0149 to 0.0337 (model 4), while the estimated volatilities in regime 1 are almost three times larger than those in regime 0. Finally, the estimated transition probabilities suggest that the Markov chains that drive changes in regime are highly persistent, thus, if the system is in either of the two regimes, it is likely to remain in that regime.

The allocation of time periods to the two states for the four switching models under consideration is shown in Figure 2. The period between 1965 and 1980 is assigned with high probability to state 1, with a brief departure from it around 1971. The period from 1980 to 1982 is assigned to the high interest rates/low-variance state (state 0). The remaining observations fall into state 1.

Table 3 shows that the hypothesis that model 4 and model 2 are valid simplifications of model 5 are rejected [the likelihood ratio test (LR) statistics are 5.994, distributed $\chi^2(1)$, and 6.539, distributed $\chi^2(2)$, respectively]. On the other hand, the null hypothesis that model 3 is

[13] The likelihood function was maximized by using the Broyden–Fletcher–Goldfarb–Shanno quasi-Newton algorithm with numerically computed derivatives.

a valid reduction is not rejected [the likelihood ratio test statistic is 2.299, distributed $\chi^2(1)$]. The Akaike, Schwarz, and Hannan-Quinn specification criteria, give conflicting results. While model 5 is favored by the AIC, model 2 is favored by the SIC and model 3 by the HQ criteria.

It is clear that neither the likelihood ratio test nor the selection criteria give us a clear cut indication of which should be the preferred model. In order to establish whether these results are sensitive to the sample specifications we recursively estimate the five models described in Table 2 (starting from 1964:1-1981:4 and sequentially enlarging the sample up to1998:4) and calculate, for each iteration, LR tests and the different complexity-penalized likelihood measures.[14] The latter is a worthwhile exercise since procedures based on recursive estimation are used in the remaining of the paper for forecasting and bond prices evaluation purposes and are, therefore, directly comparable with these results.

In Table 4 we report results of recursive goodness of fit criteria and indicate periods during which each model is selected. On the basis of the AIC criterion, model 3 is preferred during the periods 1982:1 to 1990:3 and 1992:4 to 1995:3; model 5 is selected in the remaining part of the sample. Using the SIC and HQ criteria, with the exception of the period 1990:2-1992:4 (where model 2 is chosen by the SIC), model 3 is always chosen. Recursive LR tests suggest that the hypotheses that models 1, 2 and 4 are valid simplifications of model 5 can be rejected at 5% significance level during all the sub-samples [model 2 is not rejected only for three quarters (1982:4-1983:2)]. On the other hand, the null hypothesis that model 3 is a valid reduction of model 5 is never rejected.

These results are further corroborated by Table 5, which shows that on the basis of all the complexity-penalized likelihood cumulative measures, which capture both the time series and the cross section dimension, model 3 outperform the competing models.

[Tables 2-3 approximately here]

[Figures 2-3 approximately here]

[Tables 4-5 approximately here]

5 Forecast evaluation

In this section we evaluate the multi step ahead out-of-sample forecasts performance of the 5 models considered in the previous section.

[14] Pesaran and Timmermann (1995) use a similar approach to asses the economic significance of the predictability of U.S. stock returns. See also Bossaerts and Hillion (1999).

Forecast comparisons are based on series of recursive forecasts computed in the following way. For a given time series $\{w_t\}_{t=1}^{T}$, the five alternative models are fitted to the sub-series $\{w_t\}_{t=1}^{T-\bar{h}-n}$, where $\bar{h}$ is the longest forecast horizon under consideration and n is the desired number of forecasts. Using $t = T - \bar{h} - n$ as the forecast origin, a sequence of h-step-ahead forecasts are generated from the fitted models for $h \in \{1, \ldots, \bar{h}\}$. The forecast origin is then rolled forward one period to $t = T - \bar{h} - n + 1$, the parameters of the forecast models are re-estimated and another sequence of one-step-ahead to $\bar{h}$-step-ahead forecasts is generated. The procedure is repeated until n forecasts are obtained for each $h \in \{1, \ldots, \bar{h}\}$, and these are used to compute measures of forecast performance for each forecast horizon. We calculate traditional accuracy measures (defined on the forecast errors $e_{t+h} = x_{t+h} - \widehat{x}_{t+h}$, $h \geqslant 1$, where $\widehat{x}_{t+h}$ denotes the h-step-ahead forecast at the forecast origin t) such as the mean squared error, MSE(h), the relative mean square error, RMSE(h), the mean absolute error, MAE(h), the relative mean absolute error, RMAE(h), and the Theil's inequality coefficient, U(h).[15] In addition we use a test of equal forecast accuracy due to Diebold and Mariano (1995), DM(h), to assess whether the MSE(h) from the benchmark linear model (model 1) is statistically different from the MSE(h) of the nonlinear alternatives.[16]

The first noteworthy result that emerges from the comparison is that simpler specifications of the instantaneous interest rate, such as the SWCIR model with only a regime-dependent volatility parameter, perform better than the alternative linear and nonlinear models. Models with more general Markov switching parametrization result in substantial losses of multi-period forecast accuracy.

Table 6 summarizes the results for the competing models based on several performance indicators using a ten periods ahead horizon. The results show that the forecasting performance of models 3 and 5 is very poor and that these models are always outperformed by the linear alternative. On the other hand, model 2 outperforms the linear specification for all design points, with an average gain, when the MSE and the RMSE criteria are adopted, of over 12%. The results for model 4 are mixed since they only show a gain over the linear alternative for

[15] The scaling of the coefficient is such that $0 \leqslant \mathrm{U}(h) \leqslant 1$. If $\mathrm{U}(h) = 0$, there is a perfect prediction; if, on the other hand, $\mathrm{U}(h) = 1$, the forecast performance of the model is as bad as it can be.

[16] If $\{d_i(h)\}_{i=1}^{n}$ are the loss differentials associated with the h-step-ahead forecasts from M_1 and M_2, the test is based on the statistic $\mathrm{DM}(h) = \frac{\sum_{i=1}^{n} d_i(h)}{\sqrt{n\widehat{\tau}_h^2}}$, where $\widehat{\tau}_h^2$ is a consistent estimator of $\tau_h^2 = \lim_{n\to\infty}(1/n)\mathrm{Var}[\sum_{i=1}^{n} d_i(h)]$. Under the null hypothesis of equal forecast accuracy [which entails $\mathbb{E}[d_i] = 0$], DM(h) has a standard normal asymptotic distribution. Consistent estimates of τ_h^2 are obtained by using the prewhitened kernel estimator of Andrews and Monahan (1992) in conjunction with the Parzen kernel and their data-based bandwidth selector.

a forecast horizon up to $h = 2$, and do not appear to show any substantial gain over longer forecast horizons.

These findings are further supported by the results of the test of equal forecast accuracy, as reported in Table 7. While for models 3 to 5 the DM test fails to reject the null hypothesis of equal forecast accuracy for any forecast horizon, the null hypothesis of equal accuracy cannot be rejected for any forecast horizon for model 2.

[Tables 6-7 approximately here]

6 Using bond pricing as model selection criteria for the instantaneous interest rate

The experiments carried out above give conflicting results, while the in-sample comparisons seem to select model 3 (with some evidence which favors model 5), the forecasting comparisons favor model 2 (with some evidence which favors model 4).[17] Given that pricing is intrinsically a forecasting exercise we expect model 2 (and 4) to perform better when we compare the specifications on a pricing basis. In this section we use the information contained in the term structure to decide which is the model that produces the best bond prices. To do this, we recursively estimate the five models described in Table 2 and obtain [using equation (11)] bond prices for 6 month, 1, 2 and 5 year, and compare these prices (returns) with the actual data. Therefore, we use the observations from 1964:1-1981:4 to start the pricing exercise and sequentially enlarge the sample up to 1998:4. In other words, at time τ a yield curve,$-\frac{1}{T-\tau}\ln(F(\tau, r(\tau), X_\tau, T))$, can be constructed using the pricing equation (11) and recursively estimate the instantaneous interest rate (for the alternative parameterizations) using information up to time $\tau = t_1, ...T-1, T$. This produces a series of $T-t$ long-term interest rates for each maturity and estimated model. We then compare the actual and generated yields.

Notice that in order to obtain these prices (the yield curve), we need to calculate the market price of risk λ. The latter, is calculated as in Backus, Foresi, and Telmer (2000) by equating the observed yield on the 10 year bond to the yield generated by the model. We generate yield curves for our 5 models, using for each period of time from 1981:4 to 1998:4, the estimates of the alternative parameterizations for the short-term interest rate (and the calculated market prices of risk). Figures 4-6 plot the actual and model generated yield curves for 1982:1, 1989:3 and 1998:4, respectively. A simple visual inspection of the plots shows that only model 4 is able

[17] See Hamilton, Waggoner, and Zha (2002) for a possible explanation of this finding.

to generate, for the three chosen dates, a yield curve which resembles the actual curve. Even though model 4 (and to some extent model 2) manages to reproduce the shape of the actual yield curve, a visual inspection of the shapes of the yield curve for every data point is not a feasible (or formal) selection criterion. Therefore, we attempt to use all the information contained in our generated prices to asses which of the models has best predictive power. Figure 7 plots the actual and model generated yields for the four maturities. Table 8 reports some goodness of fit statistics which are used to compare the empirical performance of each of the 5 models. We present performance indicators used in the previous section, such as the MSE, RMSE, MAE and RMAE of the difference between the generated yields and the actual data for each maturity. An aggregate measure (column 5), which captures both the time series and the cross section dimension, is also presented. A general feature of all the pricing models is that they perform better in predicting the lower maturities than in predicting higher maturities (the reduction of the error for the 5-year bond has to be attributed to the way the price of the risk is computed). More importantly, we find that models 2 and 4 significantly outperform the other models in terms of producing prices closer to the actual data. Interestingly, a straight comparison between model 1 and model 5 shows that using the very general Markov switching parameterization may not produce better prices than those obtained using the standard CIR model, even when there are apparent structural breaks in the sample. This implies that any improvement of model 5 over model 1 in terms of fit is undone by the poor forecasting performance of model 5. Figure 7 shows a deterioration in the fit of the simple CIR model after the first half of the 1980s when the structural break takes place.

We also report, in Table 9, the descriptive statistics of the yield differences. The results show that model 1 overestimates (underprices) yields (bonds), while other models underestimate (overprice) yields (bonds). In terms of the first two moments (mean and variance), models 2 and 4 perform better than the other three models. When the third (skewness) and fourth (kurtosis) moments are considered, model 4 gains more support.

Finally, we present Kolmogorov-Smirnov statistics for testing the hypothesis of equal cumulative probability distributions between the generated yields and the actual data for each model and maturity. From Table 10, it is clear that models 2 and 4 significantly outperform their alternative models and the null hypothesis of equal distributions cannot be rejected at any significance level. On the other hand, when models 3 and 5 are considered the null hypothesis is rejected at the 5% significance level for the 1 and 2 year maturities. For the linear model only the 1 year maturity is rejected.

Consider an individual who intends to price a bond using a Markov switching model in 1998:4 (the end of the sample). A clear advice can be given to this individual using the results drawn from our analysis: If she attempted to use standard statistical criteria, she would find that models 2 and 4 are not valid simplifications of model 5, and that model 3 is preferred to model 5. On the other hand, by carrying out a more extensive analysis, that is, using forecasting criteria and obtaining the bond prices by means of estimating recursively the Markov switching short rate, she would find that model 4 (and to some extent model 2) are not only, the only models that manage to reproduce the actual shape of the yield curve, but also seem to have produced on average much better prices than the other parameterizations. Most importantly, the poor performance of model 5 compared with model 1 highlight that attempting to correctly specify the switching model is crucial (especially when the model is used to forecast). This exercise suggests how important is to carry out a careful model selection of the switching interest rate process and that failing to do so, may give prices that do not represent an improvement over those obtained with models which do not allow for regime switching, even in cases where there are clear breaks in the data.

[Figures 4–6 approximately here]

[Table 8 approximately here]

[Figure 7 approximately here]

[Table 9-10 approximately here]

7 Conclusions

In this paper, we have provided an analysis of several regime-switching characterizations of the Cox, Ingersoll and Ross (1985; CIR) term structure process. We investigate how the pricing performance of the model is affected by different assumptions about which parameters (drift and diffusion) are specified as regime-dependent. We have estimated recursively Markov switching models for the short-term interest rate and generated bond yields which are then compared with actual yields. Our results reveal that simpler specifications, such as a CIR short rate with,only regime-dependent volatility and, with both regime-dependent volatility and regime-dependent long-run rate, produce better bond prices than those obtained using models with no regime switching, models where all the parameters are allowed to switch, and models with both

regime-dependent volatility and regime-dependent speed of adjustment. Interestingly, we find that the preferred (those which produce bond prices closer to the actual data) models differ from those that one would have had selected, on the basis of the goodness of fit of the instantaneous interest rate. This can only be interpreted as a sign that the models which better fit the data in sample, are not necessarily those with better forecasting performance. This issue seems to be particularly important when using Markov switching models for pricing purposes.

References

Andersen, T. G., and J. Lund (1997): "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343–377.

Andrews, D. W. K., and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

Ang, A., and G. Bekaert (2002): "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 163–182.

Backus, D., S. Foresi, and C. Telmer (2000): "Discrete-Time Models of Bond Pricing," in *Advanced Fixed-Income Valuation Tools*, ed. by N. Jagadeesh, and B. Tuckman, chap. 4, pp. 87–127. Wiley & Sons, Inc.

Bansal, R., and H. Zhou (2002): "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance*, 57, 1997–2043.

Bliss, R. R. (1997): "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research*, 9, 197–231.

Bossaerts, P., and B. Hillion (1999): "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *The Review of Financial Studies*, 12, 405–428.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992): "An Empricial Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1227.

Chapman, D. A., J. John B. Long, and N. D. Pearson (1997): "Using Proxies for the Short Rate: When are Three Months Like an Instant?," *Review of Financial Studies*, 12, 763–806.

Cochrane, J. H. (2001): *Asset Pricing*. Princeton University Press, Princeton and Oxford.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "A Theory of the Term Structure of Interes Rates," *Econometrica*, 53, 385–407.

Dahlquist, M., and S. F. Gray (2000): "Regime-Switching and Interest Rates in the European Monetary System," *Journal of International Economics*, 50, 399–419.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Duarte, J. (2000): "The Relevance of the Price of Risk in Affine Term Structure Models," *University of Chicago*, Mimeo.

Duffee, G. R. (2002): "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405–443.

Duffie, D., and L. G. Epstein (1992): "Stochastic Differential Utility," *Econometrica*, 60, 353–394.

Evans, M. (2003): "Regime Shifts, Risk and the Term Structure," *Economic Journal*, 113, 345–389.

Fama, E., and R. Bliss (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

Fisher, M., and C. Gilles (2000): "Modeling the State-Price Deflator and the Term Structure of Interest Rates," *Federal Reserve Bank of Atlanta*, Mimeo.

Garcia, R. (1998): "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39, 763–788.

Garcia, R., and P. Perron (1996): "An Analysis of the Real Interest Rate under Regime Shifts," *The Review of Economics and Statistics*, 78, 111–125.

Granger, C. W. J., M. L. King, and H. White (1995): "Comments on Testing Economic Theories and the Use of Model Selection Criteria," *Journal of Econometrics*, 67, 172–187.

Gray, S. F. (1996): "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

Hall, S., and M. Sola (1993): "A Generalised Model of Regime Changes Applied to the US Treasury Bill Rate," *London Business School*, CEF Working Paper No. 7.

Hamilton, J. D. (1988): "Rational Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control*, 12, 385–423.

——— (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hamilton, J. D., D. F. Waggoner, and T. Zha (2002): "Normalization in Econometrics," *Federal Reserve Bank of Atlanta*, Mimeo.

Hansen, B. E. (1992): "The Likelihood Ratio Test under Non-standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61–S82.

——— (1996): "Erratum: The Likelihood Ratio Test under Non-standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 11, 195–198.

Hull, J. (2000): *Options, Futures and Other Derivative Securities*. Prentice Hall, 2 edn.

Kapetanios, G. (2001): "Model Selection in Threshold Models," *Journal of Time Series Analysis*, 22, 733–754.

Kim, C.-J., and J. M. Piger (2002): "Common Stochastic Trends, Common Cycles, and Asymmetry in Economic Fluctuations," *Journal of Monetary Economics*, 49, 1189–1211.

Landen, C. (2000): "Bond Pricing in a Hidden Markov Model of the Short Rate," *Finance and Stochastics*, 4, 371–390.

Leroux, B. G. (1992): "Consistent Estimation of a Mixing Distribution," *Annals of Statistics*, 20, 1350–1360.

Leroux, B. G., and M. L. Puterman (1992): "Maximum Penalized Likelihood Estimation for Independent and Markov-Dependent Poisson Mixtures," *Biometrics*, 48, 545–558.

McCulloch, J. H., and H.-C. Kwon (1993): "US Term Structure Data, 1947-1991," *Ohio State University*, Working Paper, 93-6.

Naik, V., and M. H. Lee (1997): "Yield Curve Dynamics with Discrete Shifts in Economic Regimes," *Faculty of Commerce, University of British Columbia*, Working Paper.

Nishii, R. (1988): "Maximum Likelihood Principle and Model Selection when the True Model is Unspecified," *Journal of Multivariate Analysis*, 27, 392–403.

Nowman, B. (1997): "Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates," *Journal of Finance*, 52, 1695–1706.

Pesaran, M. H., and A. Timmermann (1995): "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50, 1201–1228.

Psaradakis, Z., and N. Spagnolo (2003): "On the Determination of the Number of Regimes in Markov-Switching Autoregressive Models," *Journal of Time Series Analysis*, 24, 237–252.

Ryden, T. (1995): "Estimating the Order of Hidden Markov Models," *Statistics*, 26, 345–354.

Sin, C.-Y., and H. White (1996): "Information Criteria for Selecting Possibly Misspecified Parametric Models," *Journal of Econometrics*, 71, 207–225.

Sola, M., and J. Driffill (1994): "Testing the Term Structure of Interest Rates from a Stationary Switching Regime VAR," *Journal of Economic Dynamics and Control*, 18, 601–628.

Sun, T.-S. (1992): "Real and Nominal Interest Rates: A Discrete-Time Model and its Continuous-Time Limit," *Review of Financial Studies*, 5, 581–611.

A Appendix: Solution

A.1 Model 1: The single regime Cox-Ingersoll-Ross model

The stochastic processes for the two state variables (the stochastic discount factor and the short rate) are given by

$$\frac{d\Lambda(t)}{\Lambda(t)} = -r(t)dt - \sigma_\Lambda \sqrt{r(t)} dW(t) \tag{A. 1}$$

and

$$dr(t) = \kappa[\alpha - r(t)]dt + \sigma\sqrt{r(t)}dW(t). \tag{A. 2}$$

The fundamental bond pricing equation (10) becomes

$$\kappa[\alpha - r]F_r + \tfrac{1}{2}\sigma^2 r F_{rr} - F_T - rF = F_r \sigma\sigma_\Lambda r. \tag{A. 3}$$

Using the following affine functional form for bond prices

$$F(t, r(t), T) = e^{A(t,T) - B(t,T)r},$$

we obtain the partial derivatives required in (A. 3). Substituting them into (A. 3) and separating the coefficients on the constant and on the terms in r result in a set of *ordinary* differential equations for $A(t,T)$ and $B(t,T)$

$$\begin{aligned} B'(t,T) &= 1 - \tfrac{1}{2}\sigma^2 B(t,T)^2 - (\sigma\sigma_\Lambda + \kappa)B(t,T), \\ A'(t,T) &= -B(t,T)\kappa\alpha. \end{aligned} \tag{A. 4}$$

One can solve them, subject to the boundary condition imposed by $F(T, r(t), T) = 1$, which implies $A(T,T) = 0$ and $B(T,T) = 0$,

$$B(t,T) = \frac{2(1 - e^{\gamma\tau})}{(\gamma + \kappa + \sigma\sigma_\Lambda)(e^{\gamma\tau} - 1) + 2\gamma},$$

$$A(t,T) = \frac{\kappa\alpha}{\sigma^2}\left(2\ln\left(\frac{2\gamma}{\psi(e^{\gamma\tau - 1}) + 2\gamma}\right) + \psi\tau\right),$$

where

$$\gamma = \sqrt{(\kappa + \sigma\sigma_\Lambda)^2 + 2\sigma^2},$$

$$\psi = \kappa + \sigma\sigma_\Lambda + \gamma,$$

$$\tau = T - t.$$

A.2 Models 2-5: The Cox-Ingersoll-Ross model with regime switching

The stochastic processes for the two state variables (the stochastic discount factor and the short rate) are given by

$$\frac{d\Lambda(t)}{\Lambda(t)} = -r(t)dt - \sigma_\Lambda \sqrt{r(t)} dW(t) \tag{A. 5}$$

and

$$dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t) \qquad X = 0, 1. \tag{A. 6}$$

The "term structure equation" is now given by

$$(\kappa(X_t)[\alpha(X_t) - r] - \sigma(X_t)\lambda r)F_r + \tfrac{1}{2}\sigma(X_t)^2 r F_{rr} - F_T + \sum_{i=0}^{1}\sum_{j=0}^{1} h_{ij}\Delta F - rF = 0, \tag{A. 7}$$

with the boundary condition

$$F(T, X, r, T) = 1.$$

Using the following affine functional form for bond prices

$$F(t, X, r(t), T) = e^{A(t,X,T) - B(t,X,T)r},$$

we obtain the partial derivatives required in (A. 7). Substituting them into (A. 7) and separating the coefficients on the constant and on the terms in r, result in a set of *ordinary* differential equations for $A(t,T)$ and $B(t,T)$,

$$\left\{ \begin{array}{l} B_{0t}(t,T) - \kappa_0 B_0(t,T) - \frac{1}{2}\sigma_0 B_0(t,T)^2 + 1 \\ B_{1t}(t,T) - \kappa_1 B_1(t,T) - \frac{1}{2}\sigma_1 B_1(t,T)^2 + 1 \end{array} \right\} r + \tag{A. 8}$$

$$\left\{ \begin{array}{lll} A_{0t}(t,T) - & \big(\kappa_0\alpha_0 - \lambda_0\sigma_0\big)B_0(t,T) + & \sum_{j=0}^{1} h_{0j}\Delta F_0 \\ A_{1t}(t,T) - & \big(\kappa_1\alpha_1 - \lambda_1\sigma_1\big)B_1(t,T) + & \sum_{j=0}^{1} h_{1j}\Delta F_1 \end{array} \right\} = 0$$

where

$$\Delta F_0 = e^{A_j(t,T) - rB_j(t,T) - [A_0(t,T) - rB_0(t,T)]},$$

and

$$\Delta F_1 = e^{A_j(t,T) - rB_j(t,T) - [A_1(t,T) - rB_1(t,T)]}.$$

The boundary value $F(T, r, X, T) \equiv 1$ implies

$$\begin{cases} A_i(T,T) &= 0, \qquad i = 0, 1, \\ B_i(T,T) &= 0, \qquad i = 0, 1. \end{cases}$$

The system of equations (A. 8) may be solved by applying the commonly adopted log-linear approximation $e^y - 1 \approx y$ [see, for example, Bansal and Zhou (2002)].

$$\left\{ \begin{array}{l} B_{0t}(t,T) - \kappa_0 B_0(t,T) - \frac{1}{2}\sigma_0 B_0(t,T)^2 + h_{01}[B_0(t,T) - B_1(t,T)] + 1 \\ B_{1t}(t,T) - \kappa_1 B_1(t,T) - \frac{1}{2}\sigma_1 B_1(t,T)^2 + h_{10}[B_1(t,T) - B_0(t,T)] + 1 \end{array} \right\} r +$$

$$\left\{ \begin{array}{lll} A_{0t}(t,T) - & \big(\kappa_0\alpha_0 - \lambda_0\sigma_0\big)B_0(t,T) + & h_{01}[A_1(t,T) - A_0(t,T)] \\ A_{1t}(t,T) - & \big(\kappa_1\alpha_1 - \lambda_1\sigma_1\big)B_1(t,T) + & h_{10}[A_0(t,T) - A_1(t,T)] \end{array} \right\} = 0. \qquad \text{(A. 9)}$$

Unfortunately, the above equations system (A. 9) has only an approximate numerical solution. One way to improve the approximation is to use a technique similar to the *control variate technique* used as a variance reduction procedures in the option pricing literature [see Hull (2000)]. This involves calculating the bond pricing equation for the single regime CIR model (A. 3) using the approximation adopted in obtaining (A. 9). The difference between these two gives the approximation error.

Table 1. Summary Statistics for the U.S. Term Structure of Interest Rates

	Period 1964-1978		Period 1979-1981		Period 1982-1998	
Maturity	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
3	4.15023	1.98836	12.0278	2.47187	6.50504	2.26995
6	4.38941	2.04902	12.2170	2.43729	6.74247	2.36220
12	4.56006	2.03908	12.1098	2.30643	7.07753	2.43233
24	4.72129	1.96039	11.8012	2.24064	7.55160	2.40562
60	4.96496	1.87292	11.4083	2.15478	8.15998	2.28925
120	5.10668	1.84290	11.2681	1.92634	8.54856	2.12881

Table 2. Estimated Models

- Model 1: No regime switching.
  - $dr(t) = \kappa[\alpha - r(t)]dt + \sigma\sqrt{r(t)}dW(t)$
- Model 2: Regime switching in volatility.
  - $dr(t) = \kappa[\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$
- Model 3: Regime switching in volatility and adjustment speed.
  - $dr(t) = \kappa(X_t)[(\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$
- Model 4: Regime switching in volatility and long-run rate.
  - $dr(t) = \kappa[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$
- Model 5: Regime switching in all parameters.
  - $dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Table 3. Parameters Estimates of Models

	Model 1	Model 2	Model 3	Model 4	Model 5
α_0	0.0161	–	0.0296	0.0149	0.0154
	(0.0028)		(0.0016)	(0.0021)	(0.0025)
σ_0	0.0177	0.0128	0.0130	0.0128	0.0129
	(0.0011)	(0.0009)	(0.0008)	(0.0008)	(0.0008)
κ_0	0.0684	–	0.0118	–	0.0614
	(0.0344)		(0.0080)		(0.0340)
α_1	–	0.0154	–	0.0337	0.0296
		(0.0024)		(0.0239)	(0.0016)
σ_1	–	0.0399	0.0310	0.0388	0.0311
		(0.0082)	(0.0064)	(0.0080)	(0.0064)
κ_1	–	0.0635	1.0234	0.0725	0.9905
		(0.0322)	(0.3387)	(0.0360)	(0.3417)
p	–	0.9039	0.9115	0.9107	0.9136
		(0.0849)	(0.0796)	(0.0806)	(0.0777)
q	–	0.9899	0.9925	0.9906	0.9926
		(0.0108)	(0.0075)	(0.0098)	(0.0075)
Log L	363.1274	384.9890	387.1091	385.2617	388.2586
AIC	-720.2548	-757.9780	-760.2182	-756.5230	-760.5172
SIC	-711.4298	-740.3281	-739.6287	-735.9319	-736.9840
HQ	-716.6681	-750.8056	-751.8504	-748.1556	-750.9540

Notes: The table reports the recursive Markov switching maximum likelihood estimates as of 1998:4 for the 5 models, which are:

Model 1: $dr(t) = \kappa[\alpha - r(t)]dt + \sigma\sqrt{r(t)}dW(t)$

Model 2: $dr(t) = \kappa[\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Model 3: $dr(t) = \kappa(X_t)[\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Model 4: $dr(t) = \kappa[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Model 5: $dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

The figures in parentheses are asymptotic standard errors.

Table 4. Recursive AIC, SIC, HQ and LR Tests

	AIC	SIC	HQ	LR
Model 1	–	–	–	–
Model 2	–	1990:2 - 1992:4	–	1982:4 - 1983:2
Model 3	1982:1 - 1990:3	1982:1 - 1990:2	1982:1 - 1998:4	1982:1 - 1998:4
	1992:4 - 1995:3	1992:4 - 1998:4		
Model 4	–	–	–	–
Model 5	1990:3 - 1992:4	–	–	–
	1995:3 - 1998:4			

Table 5. Cumulative Recursive AIC, SIC and HQ

	AIC	SIC	HQ
Model 1	-35838.177	-35297.451	-35619.420
Model 2	-37573.112	-36491.661	-37135.599
Model 3	-37774.079	-36512.386	-37263.648
Model 4	-37476.543	-36214.849	-36966.110
Model 5	-37745.557	-36303.622	-37162.206

Table 6. Out of Sample Performance Results of Models

h	1	2	3	4	5	6	7	8	9	10
Mean Squared Error (MSE)										
Model 1	1.00	2.21	3.21	4.40	5.70	7.06	8.22	9.28	10.08	11.05
Model 2	0.92	1.91	2.79	3.84	5.01	6.28	7.39	8.46	9.22	10.01
Model 3	31.15	29.71	29.56	29.17	28.83	28.28	27.74	26.97	26.51	25.70
Model 4	0.94	2.03	3.43	5.11	6.88	8.67	10.31	11.83	12.91	13.73
Model 5	29.63	30.20	30.52	30.50	30.50	30.34	30.19	29.78	29.46	28.77
Relative Mean Squared Error (RMSE)										
Model 1	0.02	0.04	0.06	0.09	0.12	0.15	0.17	0.20	0.23	0.26
Model 2	0.01	0.03	0.05	0.08	0.10	0.13	0.15	0.18	0.21	0.23
Model 3	0.73	0.70	0.70	0.71	0.71	0.71	0.71	0.72	0.72	0.72
Model 4	0.01	0.04	0.06	0.09	0.13	0.16	0.20	0.24	0.27	0.30
Model 5	0.71	0.72	0.74	0.75	0.76	0.77	0.77	0.78	0.79	0.80
Mean Absolute Error (MAE)										
Model 1	0.71	1.10	1.41	1.69	1.92	2.16	2.33	2.52	2.67	2.82
Model 2	0.67	1.04	1.31	1.54	1.78	1.99	2.18	2.40	2.55	2.67
Model 3	5.22	5.10	5.07	5.02	4.98	4.95	4.91	4.88	4.82	4.76
Model 4	0.74	1.12	1.42	1.78	2.10	2.37	2.63	2.87	3.04	3.17
Model 5	5.12	5.19	5.20	5.18	5.19	5.19	5.18	5.17	5.13	5.09
Relative Mean Absolute Error (RMAE)										
Model 1	0.11	0.17	0.21	0.25	0.29	0.34	0.36	0.40	0.43	0.46
Model 2	0.10	0.15	0.19	0.23	0.27	0.31	0.34	0.38	0.41	0.43
Model 3	0.83	0.81	0.81	0.81	0.81	0.82	0.82	0.82	0.82	0.83
Model 4	0.11	0.16	0.21	0.26	0.31	0.35	0.40	0.44	0.47	0.50
Model 5	0.81	0.83	0.84	0.84	0.85	0.86	0.86	0.87	0.87	0.88
Theil's Inequality										
Model 1	0.07	0.12	0.15	0.18	0.21	0.25	0.28	0.30	0.33	0.36
Model 2	0.07	0.10	0.13	0.16	0.19	0.22	0.25	0.28	0.30	0.33
Model 3	0.62	0.62	0.63	0.63	0.64	0.64	0.65	0.65	0.66	0.66
Model 4	0.07	0.11	0.15	0.19	0.24	0.28	0.32	0.35	0.38	0.41
Model 5	0.60	0.63	0.65	0.67	0.68	0.70	0.71	0.73	0.74	0.75

Table 7. Diebold-Mariano Tests

h	1	2	3	4	5	6	7	8	9	10
Model 2	0.03	0.00	0.00	0.02	0.01	0.01	0.02	0.02	0.04	0.03
Model 3	0.78	0.83	0.88	0.87	0.93	0.97	0.99	0.99	0.99	0.99
Model 4	0.35	0.23	0.70	0.79	0.78	0.81	0.95	0.99	1.00	1.00
Model 5	0.83	0.84	0.90	0.89	0.94	0.97	0.99	0.99	0.99	0.99

Notes: The figures are p-values

Table 8. In Sample Performance Results of Models

	Maturity				
	6 Month	1 Year	2 Year	5 Year	Total
Mean Squared Error (MSE)					
Model 1	0.5306	1.0997	0.8339	0.1728	2.6370
Model 2	0.0614	0.2042	0.2986	0.1181	0.6823
Model 3	0.3592	1.2558	1.7106	0.4848	3.8104
Model 4	0.0704	0.2711	0.4625	0.2185	1.0225
Model 5	0.3635	1.2511	1.6432	0.4449	3.7027
Relative Mean Square Error (RMSE)					
Model 1	0.0237	0.0382	0.0238	0.0039	0.0896
Model 2	0.0011	0.0035	0.0047	0.0016	0.0109
Model 3	0.0167	0.0591	0.0667	0.0124	0.1549
Model 4	0.0014	0.0055	0.0086	0.0032	0.0187
Model 5	0.0166	0.0561	0.0613	0.0112	0.1452
Mean Absolute Error (MAE)					
Model 1	0.5960	0.8395	0.7163	0.3301	2.4819
Model 2	0.1498	0.2951	0.3858	0.2529	1.0836
Model 3	0.4736	0.9414	1.1535	0.6303	3.1988
Model 4	0.1828	0.4129	0.5902	0.4098	1.5957
Model 5	0.4836	0.9439	1.1258	0.6013	3.1546
Relative Mean Absolute Error (RMAE)					
Model 1	0.1083	0.1375	0.1082	0.0462	0.4002
Model 2	0.0221	0.0420	0.0515	0.0314	0.1470
Model 3	0.0915	0.1846	0.2106	0.0960	0.5827
Model 4	0.0281	0.0616	0.0837	0.0523	0.2257
Model 5	0.0926	0.1818	0.2025	0.0909	0.5678

Model 1: $dr(t) = \kappa[\alpha - r(t)]dt + \sigma\sqrt{r(t)}dW(t)$

Model 2: $dr(t) = \kappa[\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Model 3: $dr(t) = \kappa(X_t)[\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Model 4: $dr(t) = \kappa[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Model 5: $dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$

Table 9. Descriptive Statistics of Yield Diffusion

Maturity	Mean	Variance	Skewness	Kurtosis
Model 1: $dr(t) = \kappa[\alpha - r(t)]dt + \sigma\sqrt{r(t)}dW(t)$				
6 Month	0.5601	0.5306	1.52787	2.7184
1 Year	0.7827	1.0997	1.5763	2.9075
2 Year	0.6570	0.8339	1.6436	3.2214
5 Year	0.2871	0.1728	1.5960	3.2651
Model 2: $dr(t) = \kappa[\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$				
6 Month	-0.1135	0.0614	-0.0462	0.0466
1 Year	-0.2495	0.2042	-0.2381	0.3687
2 Year	-0.3492	0.2986	-0.3605	0.5860
5 Year	-0.1984	0.1181	-0.0776	0.0668
Model 3: $dr(t) = \kappa(X_t)[\alpha - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$				
6 Month	-0.4118	0.3592	-0.3205	0.4094
1 Year	-0.8881	1.2558	-1.9432	3.4012
2 Year	-1.1255	1.7106	-2.8431	5.0770
5 Year	-0.6273	0.4848	-0.4073	0.3610
Model 4: $dr(t) = \kappa[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$				
6 Month	-0.1659	0.0704	-0.0434	0.0354
1 Year	-0.3812	0.2711	-0.2459	0.2891
2 Year	-0.5654	0.4625	-0.4464	0.5269
5 Year	-0.3965	0.2185	-0.1404	0.1061
Model 5: $dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)\sqrt{r(t)}dW(t)$				
6 Month	-0.4264	0.3635	-0.3260	0.3926
1 Year	-0.8930	1.2511	-1.9041	3.2248
2 Year	-1.1031	1.6432	-2.6718	4.6273
5 Year	-0.5979	0.4449	-0.3614	0.3102

Table 10. Kolmogorov-Smirnov Tests

	Maturity			
	6 Months	1 Year	2 Year	5 Year
Model 1	0.1912	0.2647	0.2206	0.1176
	(0.1665)	(0.0171)	(0.0731)	(0.7344)
Model 2	0.0735	0.1324	0.1176	0.0735
	(0.9929)	(0.5907)	(0.7344)	(0.9929)
Model 3	0.2206	0.3088	0.3382	0.1765
	(0.0731)	(0.0031)	(0.0008)	(0.2402)
Model 4	0.1176	0.1765	0.1912	0.1176
	(0.7344)	(0.2402)	(0.1665)	(0.7344)
Model 5	0.2206	0.3088	0.3382	0.1765
	(0.0731)	(0.0031)	(0.0008)	(0.2402)

Notes: The figures in parentheses are p-values

Figure 1: The Historical U.S Term Structure of Interest Rates

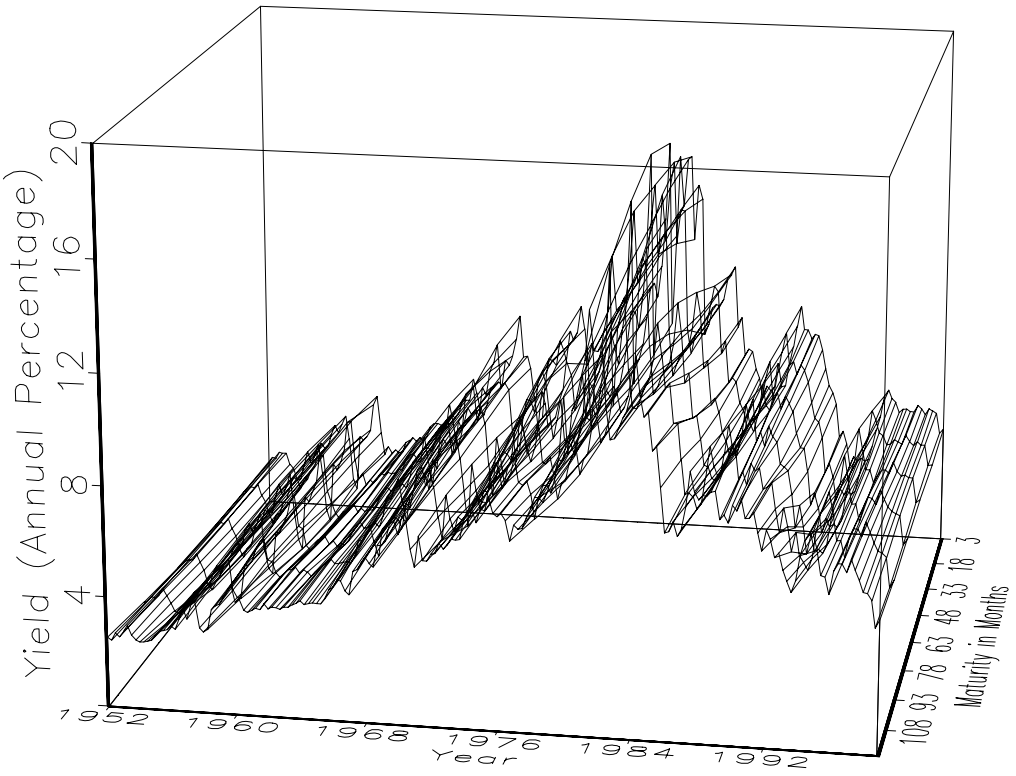

Figure 2: Probability of Regime 1 - Univariate Filter. (These are filtered probabilities conditional on information available at time t).

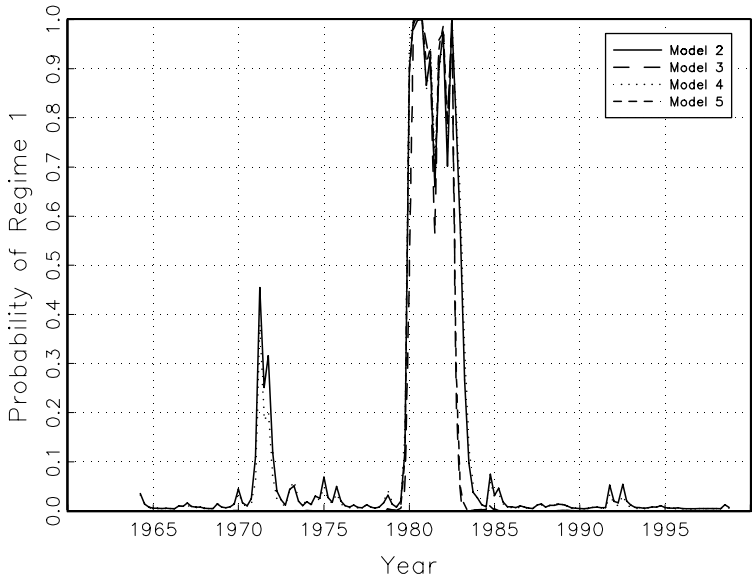

Figure 3: Actual Yields on 3-Month Bond

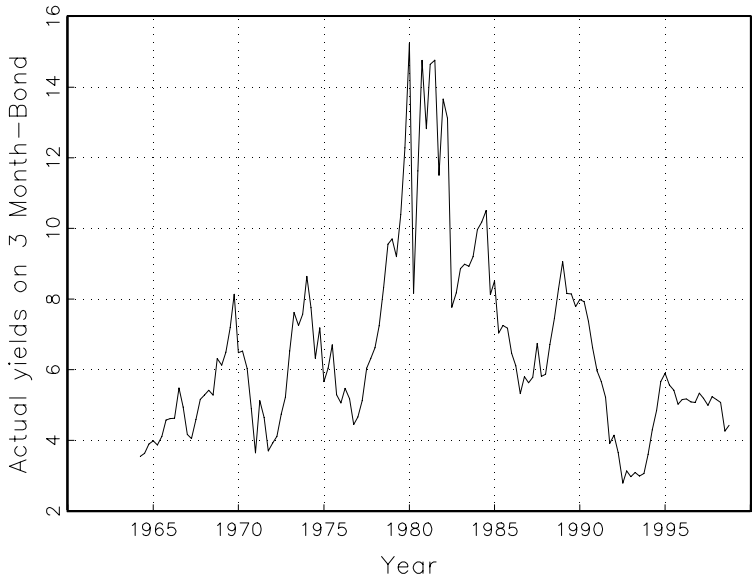

Figure 4: Model Generated Yield Curves. Date: 1982:1

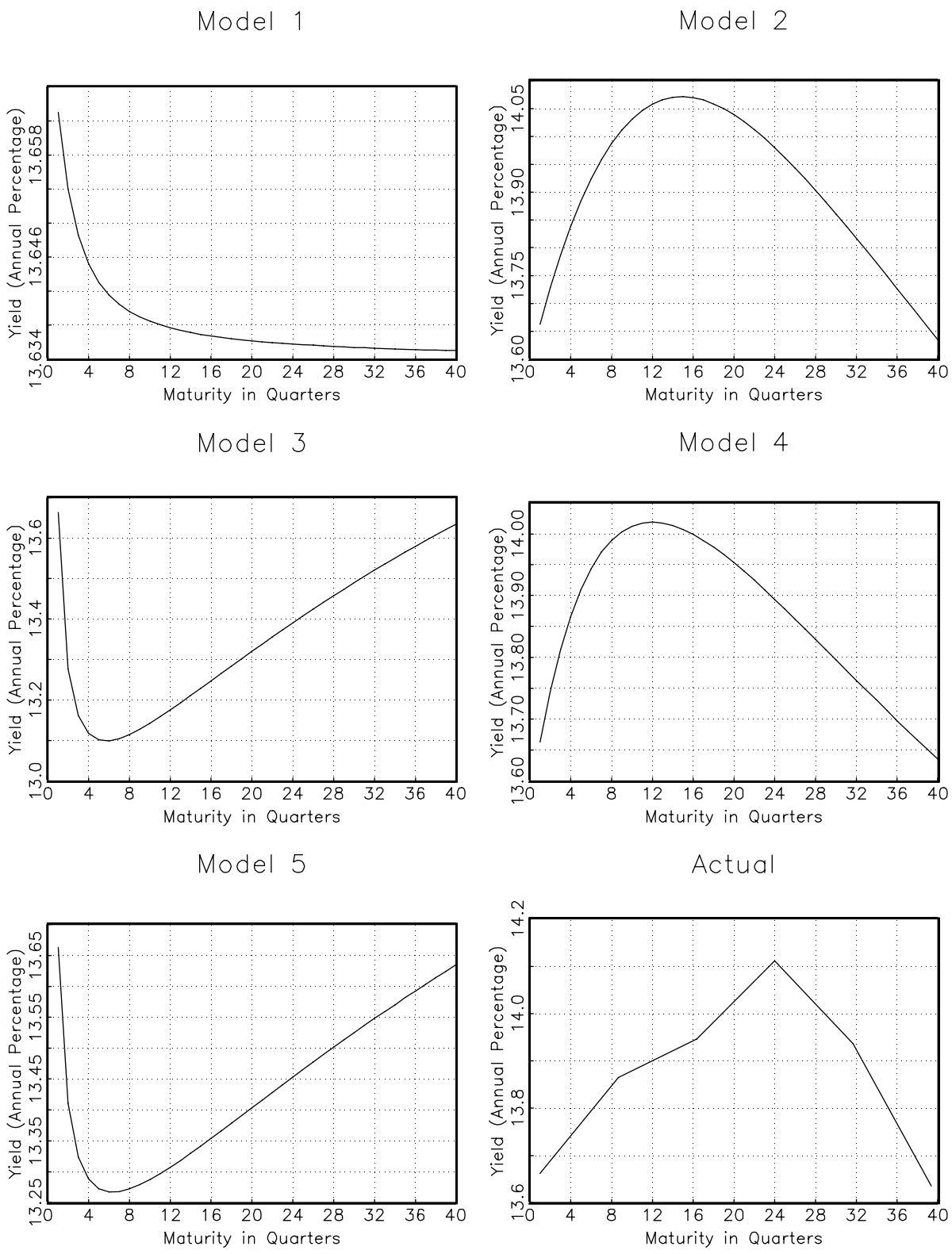

Figure 5: Model Generated Yield Curves. Date: 1989:3

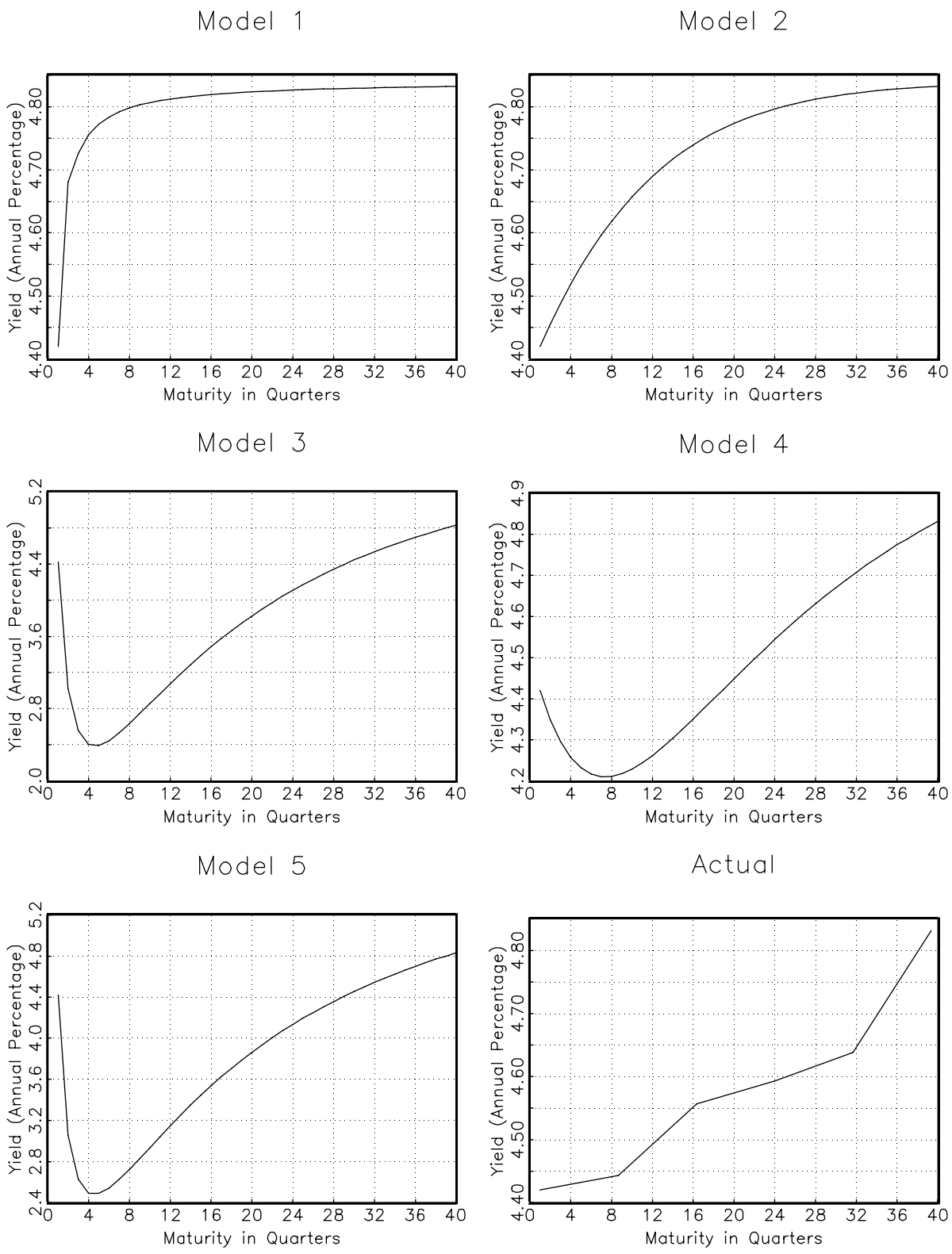

Figure 6: Model Generated Yield Curves. Date: 1998:4

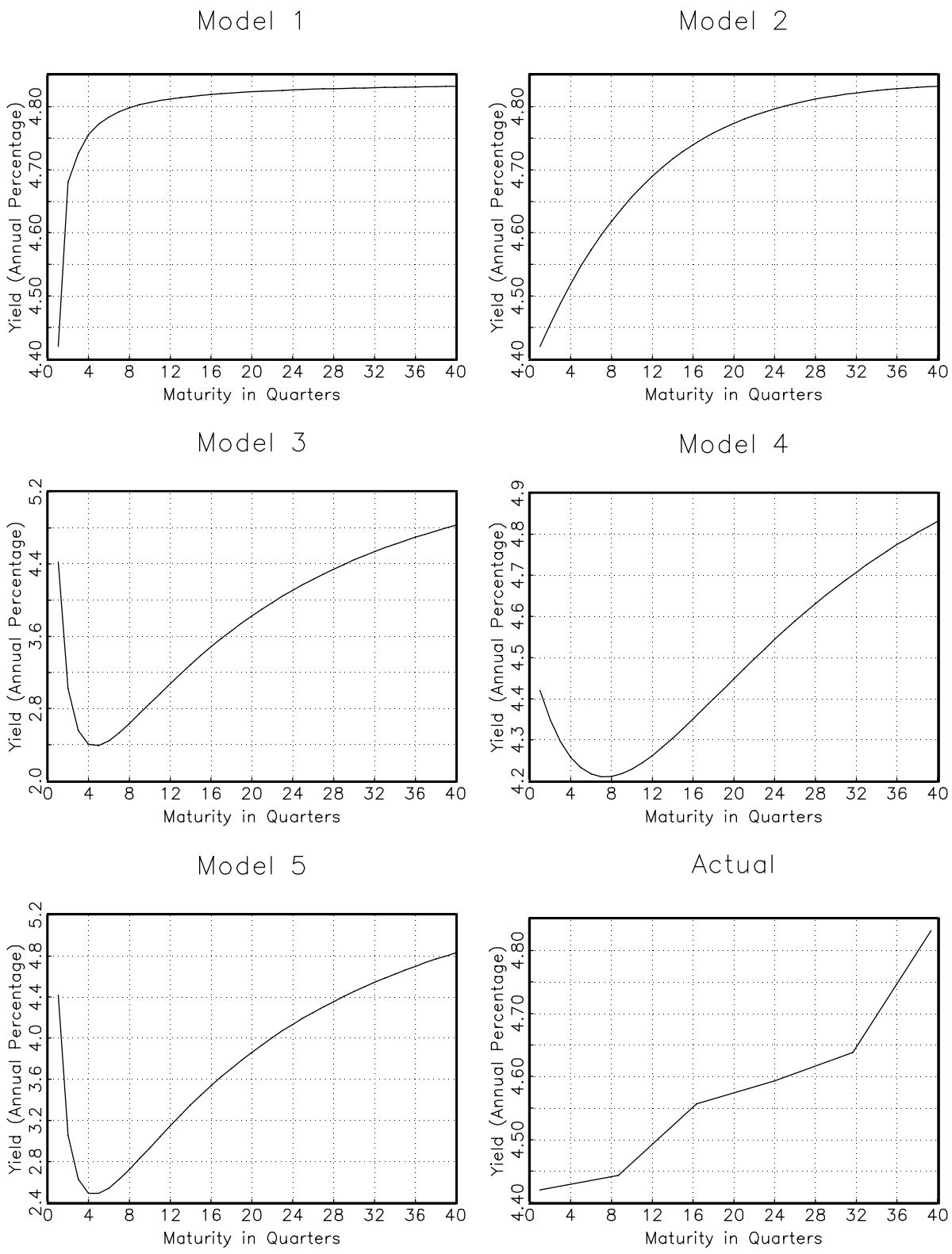
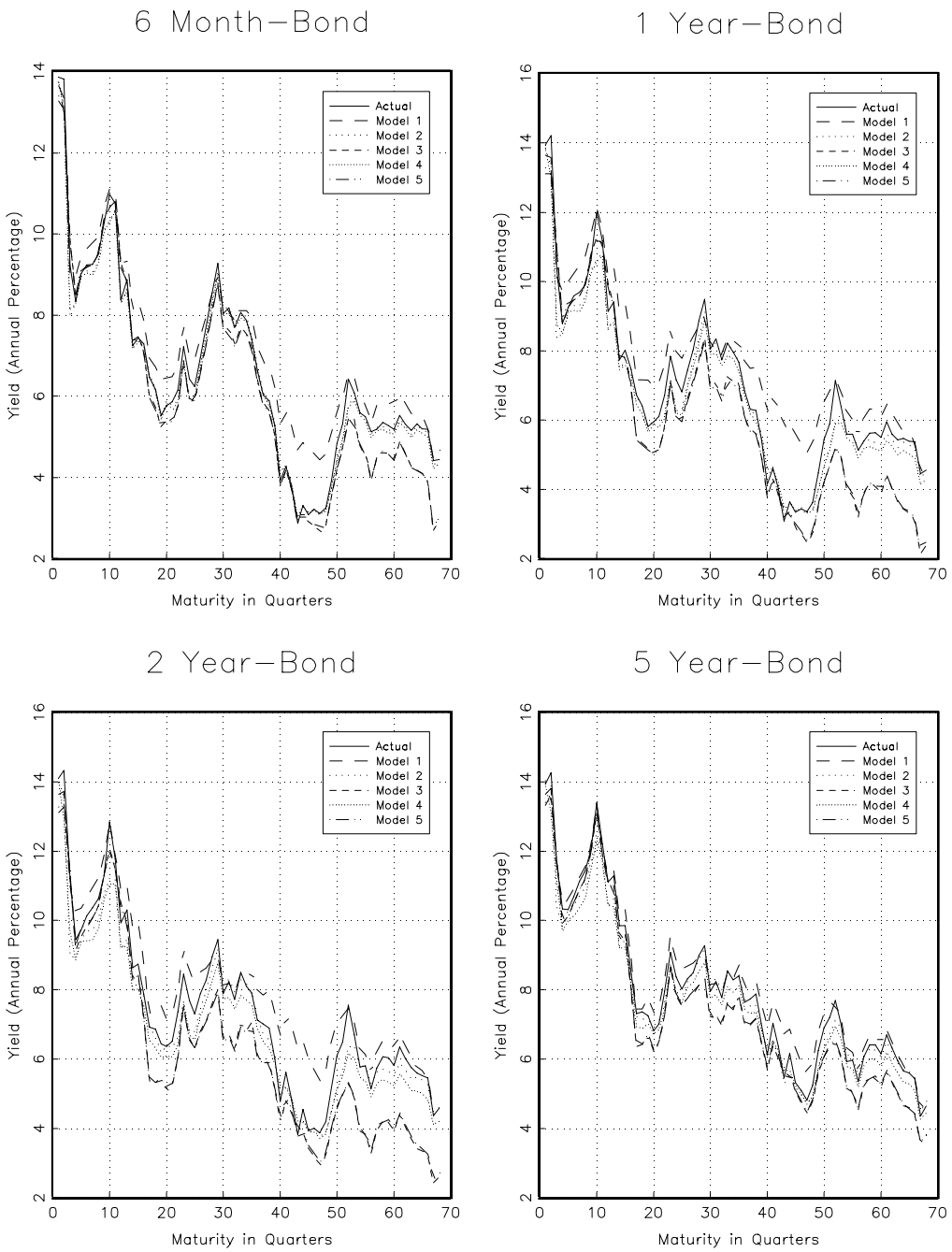
Figure 7: Actual and Model Generated Yields