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

John Driffill\textsuperscript{a}, Turalay Kenc\textsuperscript{b}, Martin Sola\textsuperscript{a,c} and Fabio Spagnolo\textsuperscript{d}

\textsuperscript{a}Birkbeck College, University of London
\textsuperscript{b}Imperial College, University of London
\textsuperscript{c}Universidad Torcuato Di Tella
\textsuperscript{d}Brunel University

This version: August 2003

Abstract

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 it 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.

Key Words: Term structure of interest rates; bond yields; stochastic discount factor/pricing kernel; and regime switching.

JEL Classification: E43; G12.

\footnote{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. Address correspondence to: Martin Sola, School of Economics, Mathematics and Statistics, Birkbeck College, University of London, Malet Street, Bloomsbury, London WC1E 7HX. E-mail address: msola@econ.bbk.ac.uk}
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 Driffl (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

\(^1\)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).
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 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 criteria4, and out-of-sample multi-step ahead forecasting (for

---

2We 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.

3To 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.

4Such 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 ? 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
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.

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

In this section, we present the regime switching term structure model. Before incorporating regime shifts into the term structure model we first consider the benchmark Cox-Ingersoll-Ross (CIR) model in which a single factor $x$, typically associated with the short rate $r$, governs the term structure and follows a mean-reverting square root process. Following Sun (1992) eq. 6 the discrete-time version of the CIR process for the single factor is written as

$$x_{t+1} - x_t = \kappa[\alpha - x_t] + \sigma \sqrt{x_t} u_{t+1}.$$  \hspace{1cm} (1)

with $\{u_{t+1}\}$ distributed normally and independently with mean zero and variance one. $\alpha$ is the long term mean the factor reverts to, $\kappa$ determines the regime-dependent adjustment speed of $x$ toward the long-term mean, and $\sigma^2 x$ is the variance of the unexpected factor changes. The term $\sigma$ is the local volatility factor and serves as a scale factor.\(^5\)

Markov-dependent finite mixture models has been studied by Leroux (1992), Leroux and Puterman (1992) and ?.\(^5\)

This implies that the volatility of the interest rate (factor) 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).
The pricing kernel (stochastic discount factor), $M$, for a discrete time version of CIR is

$$M_{t+1} = \exp \left[ -r_t^f - \frac{\lambda^2}{2} x_t - \lambda \sqrt{\pi_t} u_{t+1} \right].$$

We refer to $\lambda$ as the *market price of diffusion risk*, since it determines the covariance between shocks to $M$ and $x$, and thus the risk characteristics of bonds and related assets. Note that $E[M_{t+1}] = \exp(-r_t^f)$, where $r_t^f$ is the continuous one-period risk-free rate.

We are now going to incorporate regime switching effects into the above term structure model. This is done by assuming that the short rate (factor) in eq. (1) follows a switching process as the parameters $\kappa_{st}$, $\alpha_{st}$ and $\sigma_{st}$ take different values in different regimes $s_t$. Following the previous works such as Hamilton (1989) and Bansal and Zhou (2002) we also model the regime switching process $s_t$ as a two state Markov process, i.e., $s_t$ is either 0 (regime 0) or 1 (regime 1). The switch between the regimes is governed by a Markov chain with a transition probability matrix $\Pi = (\pi_{ij})$:

$$\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}. \quad (3)$$

where $\sum_{j=0,1} \pi_{ij} = 1$ and $0 < \pi_{ij} < 1$. The probability that a transition occurs from state $s_t = i$ (say $i = 0$) to state $s_{t+1} = j$ (say $j = 1$) in the interval $[t, t+1)$ is equal to $\pi_{01}$. Similarly, $1 - \pi_{01}$ is the probability that the process remains in state $i$. For analytical tractability, it is assumed that the discrete states $s_{t+1}$ are independent of the random process $u_{t+1}$. It is also assumed that agents in the financial markets know the actual state of the system $s_{t+1}$.

With regime shifts the process in (1) becomes a “regime-switching mean-reverting square root process” (the RSCIR process) as follows:

$$x_{t+1} - x_t = \kappa_{s_{t+1}} [\alpha_{s_{t+1}} - x_t] + \sigma_{s_{t+1}} \sqrt{x_t} u_{s_{t+1}}. \quad (4)$$

Following Bansal and Zhou (2002) we model the market price of random risk as regime depen-

---

6However, the econometrician does not observe the actual state, and has to make inferences of it based on the observable history of the system.

7Evans (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).
dent: $\lambda_{s_{t+1}}$. The pricing kernel therefore needs to be adjusted for regime shifts as follows

$$M_{t+1} = \exp \left[ -r^f_t - \frac{\lambda^2_{s_{t+1}}}{2} x_t - \lambda_{s_{t+1}} \sqrt{x_t} u_{t+1} \right].$$

(5)

To incorporate the regime switching risk we introduce a useful representation for a Markov chain which is obtained by letting $\Psi_t$ denote a random $(2 \times 1)$ vector whose $j$th element is equal to unity if $s_t = j$ and whose $s_t = j$ element equals to zero otherwise. Following Hamilton (1989) it is possible to express a Markov chain in the form

$$\Psi_{t+1} = \Pi \Psi_t + V_{t+1}$$

where $V_{t+1}$ is the innovation vector.

The pricing kernel (stochastic discount factor) for a discrete time version of CIR is

$$M_{t+1} = \exp \left[ -r^f_t - \frac{\lambda^2_{s_{t+1}}}{2} x_t - \lambda_{s_{t+1}} \sqrt{x_t} u_{t+1} - \frac{\Gamma j^2}{2} - \sum_{i \neq j} \Gamma^j (\Psi_{t+1} - H \Psi_t) \right].$$

(6)

We refer to $\Gamma$ as the market price of regime switching risk.

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 log price of a maturity $T$-bond has the form

$$P_i(t, T) = \exp \left[ -A_i(T) - B_i(T) x_t \right]$$

(7)

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

Our aim is to price bonds using (i) the above semi-affine term structure and (ii) the no-arbitrage condition. To satisfy the no-arbitrage condition we will use the fundamental pricing equation

$$P_i(t, T) = E_t[M_{t+1} P_i(t + 1, T - 1)]$$

(8)

where $E_t$ is the expectation operator conditional on information at time $t$.

---

*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.*
We assume that the distribution of the stochastic discount factor $M_{t+1}$ is conditionally lognormal. We specify models in which bond prices are jointly lognormal with $M_{t+1}$. We can then take logs of (8) to obtain

$$\log P_i(t, T) = E_t[\log M_{t+1} + \log P_i(t + 1, T - 1)] + \frac{1}{2} Var[\log M_{t+1} + \log P_i(t + 1, T - 1)].$$ (9)

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. 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.

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

\(^9\)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.
and Lund (1997) and Duffee (2002). 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 models with constant volatility scale factor, \( \sigma \), 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.

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. 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

\[10\] 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.

\[11\] The likelihood function was maximized by using the Broyden–Fletcher–Goldfarb–Shanno quasi-Newton algorithm with numerically computed derivatives.
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 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 to 1998:4) and calculate, for each iteration, LR tests and the different complexity-penalized likelihood measures. 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

---

12 Pesaran and Timmermann (1995) use a similar approach to assess the economic significance of the predictability of U.S. stock returns. See also ?. 
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.

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 \( \{w_t\}_{t=1}^{T-h-n} \), where \( h \) is the longest forecast horizon under consideration and \( n \) is the desired number of forecasts. Using \( t = T - h - n \) 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 \( e_{t+h} = x_{t+h} - \hat{x}_{t+h}, h \geq 1 \), where \( \hat{x}_{t+h} \) denotes the \( h \)-step-ahead forecast at the forecast origin \( t \) ) such as the mean squared error, \( \text{MSE}(h) \), the relative mean square error, \( \text{RMSE}(h) \), the mean absolute error, \( \text{MAE}(h) \), the relative mean absolute error, \( \text{RMAE}(h) \), and the Theil’s inequality coefficient, \( \text{U}(h) \).\(^\text{13}\) In addition we use a test of equal forecast accuracy due to Diebold and Mariano (1995), \( \text{DM}(h) \), to assess whether the

\(^{13}\) The scaling of the coefficient is such that \( 0 \leq \text{U}(h) \leq 1 \). If \( \text{U}(h) = 0 \), there is a perfect prediction; if, on the other hand, \( \text{U}(h) = 1 \), the forecast performance of the model is as bad as it can be.
MSE\((h)\) from the benchmark linear model (model 1) is statistically different from the MSE\((h)\) of the nonlinear alternatives.\(^\text{14}\)

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 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.

\[^{14}\text{If } \{d_i(h)\}_{i=1}^n \text{ are the loss differentials associated with the } h\text{-step-ahead forecasts from } M_1 \text{ and } M_2, \text{ the test is based on the statistic } DM(h) = \frac{\sum_{i=1}^n d_i(h)}{\sqrt{\hat{\tau}_h^2}} \text{, where } \hat{\tau}_h^2 = \lim_{n \to \infty} \frac{1}{n} \text{Var}[\sum_{i=1}^n d_i(h)]. \text{ Under the null hypothesis of equal forecast accuracy [which entails } E[d_i] = 0], \text{ } DM(h) \text{ has a standard normal asymptotic distribution. Consistent estimates of } \tau_h^2 \text{ are obtained by using the prewhitened kernel estimator of Andrews and Monahan (1992) in conjunction with the Parzen kernel and their data-based bandwidth selector.}\]
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).\textsuperscript{15} 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 (??)] 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 $\tau$ a yield curve, $\frac{1}{T-\tau} \ln(F(\tau, r(\tau), X_{\tau}, T))$, can be constructed using the pricing equation (??) 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 $\lambda$. 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 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

\textsuperscript{15}See ? for a possible explanation of this finding.
our generated prices to assess 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.

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


Appendix A. 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

\[ M_{t+1} = \exp \left[ -r_t^f - \frac{\lambda^2}{2} x_t - \lambda \sqrt{x_t} u_{t+1} \right], \quad (A. 1) \]

and

\[ x_{t+1} - x_t = \kappa [\alpha - x_t] + \sigma \sqrt{x_t} u_{t+1}. \quad (A. 2) \]

To price bonds these two expressions are used in the fundamental pricing equation (9) in the text:

\[ \log P(t, T) = E_t[\log M_{t+1} + \log P(t+1, T-1)] + \frac{1}{2} \text{Var}[\log M_{t+1} + \log P(t+1, T-1)]. \quad (A. 3) \]

Using the following affine functional form for bond prices

\[ P(t, T) = \exp [-A(T) - B(T) x_t], \quad (A. 4) \]

with the boundary condition

\[ P(T, 0) = 1, \]

we obtain the expressions \( P(t, T) \) and hence \( P(t+1, T-1) \) required in (A. 3). Substituting them into (A. 3) and using the fact that \( M_{t+1} = \exp[-r_t^f] = \exp[-x_t] \) yield

\[ A(T) + B(T) x_t = [A(T - 1) + B(T - 1) \kappa \alpha] + [1 + B(T - 1)(1 - \kappa)] x_t \]

\[ - \frac{1}{2} B(T - 1)^2 \sigma^2 x_t - B(T - 1) \lambda \sigma x_t. \quad (A. 5) \]

The right side is obtained as follows:

\[ \log M_{t+1} + \log P(t+1, T-1) = -x_t - \frac{\lambda^2}{2} x_t - \lambda \sqrt{x_t} u_{t+1} - A(T - 1) - B(T - 1) x_{t+1} \]

\[ = -[A(T - 1) + B(T - 1) \kappa \alpha] - [1 + \frac{\lambda^2}{2} + B(T - 1)(1 - \kappa)] x_t \]

\[ -[\lambda + B(T - 1) \sigma] \sqrt{x_t} u_{t+1} \]

which has the conditional moments

\[ E_t[\log M_{t+1} + \log P(t+1, T-1)] = -[A(T - 1) + B(T - 1) \kappa \alpha] - [1 + \frac{\lambda^2}{2} + B(T - 1)(1 - \kappa)] x_t, \]
and

\[ \text{Var}[\log M_{t+1} + \log P(t+1, T-1)] = [\lambda + B(T-1)\sigma]^2 x_t. \]

Separating the coefficients on the constant and on the terms in \( x \) in equation (A.5) gives us a set of difference equations for \( A(T) \) and \( B(T) \)

\[
B(T) = 1 + (1 - \kappa - \lambda \sigma)B(T-1) - \frac{1}{2}B(T-1)^2\sigma^2, \\
A(T) = A(T-1) + B(T-1)\kappa \alpha. 
\]

(A.6)

The boundary condition \( P(T, 0) = 1 \) implies that

\[ A(0) = B(0) = 0. \]

Given values for \( \alpha, \kappa, \sigma, \lambda \) and subject to the above boundary condition we can easily evaluate \( A(T) \) and \( B(T) \) in (A.6). The exponential form of (A.4) means that log prices and log yields are linear functions of the interest rate (factor)

\[ y(t, T) = -\log P(t, T) = \frac{A(T)}{T} + \frac{B(T)}{T} x_t \]

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

The stochastic processes with regime dependent parameters for the stochastic discount factor and the short rate are given by

\[ M_{t+1} = \exp \left[ -\lambda \frac{x_{t+1}}{2} \right] \]

and

\[ x_{t+1} - x_t = \kappa [\alpha x_{t+1} - x_t] + \sigma \sqrt{x_t} u_{t+1}. \]

(B.1)

(B.2)

With regime shifts the fundamental bond pricing equation (9) becomes

\[
P_i(t, T) = \sum_{\sigma_{t+1}=0,1} \pi_{i\sigma_{t+1}} E_t [M_{\sigma_{t+1},t+1} P_{\sigma_{t+1}}(t+1, T-1)] \\
= \pi_{i0} E_t [M_{0,t+1} P_0(t+1, T-1)] + \pi_{i1} E_t [M_{1,t+1} P_1(t+1, T-1)].
\]

(B.3)
Conditional on the current regime \(i\) we can write

\[
P_0(t,T) = E_i [M_{0,t+1}P_0(t+1,T-1)] + \pi_{0i} E_i [M_{1,t+1}P_1(t+1,T-1) - M_{0,t+1}P_0(t+1,T-1)] \\
= E_i [M_{0,t+1}P_0(t+1,T-1)] + \pi_{01} [P_1(t,T) - P_0(t,T)] \quad \text{if} \quad i = 0, \quad (B.4)
\]

\[
P_1(t,T) = E_i [M_{1,t+1}P_1(t+1,T-1)] + \pi_{10} E_i [M_{0,t+1}P_0(t+1,T-1) - M_{1,t+1}P_1(t+1,T-1)] \\
= E_i [M_{1,t+1}P_1(t+1,T-1)] + \pi_{10} [P_0(t,T) - P_1(t,T)] \quad \text{if} \quad i = 1. \quad (B.5)
\]

In order to obtain the second terms in (B.4) and (B.5) we use a simple trick. For example, for the second term in (B.5) we write

\[
M_{0,t+1}P_0(t+1,T-1) - M_{1,t+1}P_1(t+1,T-1) = \\
[M_{0,t+1}P_0(t+1,T-1) + \pi_{10}M_{0,t+1}P_0(t+1,T-1) - \pi_{10}M_{0,t+1}P_0(t+1,T-1) \\
- M_{1,t+1}P_1(t+1,T-1) - \pi_{10}M_{1,t+1}P_1(t+1,T-1) + \pi_{10}M_{1,t+1}P_0(t+1,T-1)]
\]

Rearranging and using the fundamental asset pricing expression (8) yield

\[
M_{0,t+1}P_0(t+1,T-1) - M_{1,t+1}P_1(t+1,T-1) = \\
[M_{0,t+1}P_0(t+1,T-1) + \pi_{10}M_{0,t+1}P_0(t+1,T-1) - (1-\pi_{11})M_{0,t+1}P_0(t+1,T-1) \\
- M_{1,t+1}P_1(t+1,T-1) - \pi_{10}M_{1,t+1}P_1(t+1,T-1) + (1-\pi_{11})M_{1,t+1}P_0(t+1,T-1)]
\]

\[
M_{0,t+1}P_0(t+1,T-1) - M_{1,t+1}P_1(t+1,T-1) = \\
[\pi_{10}M_{0,t+1}P_0(t+1,T-1) + \pi_{11}M_{0,t+1}P_0(t+1,T-1) \\
- \pi_{10}M_{1,t+1}P_1(t+1,T-1) - \pi_{11}M_{1,t+1}P_0(t+1,T-1)]
\]

\[
E_t[M_{0,t+1}P_0(t+1,T-1) - M_{1,t+1}P_1(t+1,T-1)] = P_0(t,T) - P_1(t,T)
\]

Now, the affine function for bond prices takes the following regime specific form

\[
P_t(t,T) = \exp[-A_t(T) - B_t(T)x_t],
\]

we obtain the expressions \(P_1(t,T)\) and hence \(P_{n+1}(t+1,T-1)\) required in (B.4) and (B.5).
The terms of (B.4) and (B.5) are obtained as follows:

\[
M_{0,t+1}P_0(t+1, T-1) + \pi_{01}[P_t(t, T) - P_0(t, T)]
\]

\[
= \exp \left[ -x_t - \frac{\lambda_0^2}{2}x_t - \lambda_0 \sqrt{x_t}u_{t+1} - A_0(T-1) - B_0(T-1)x_{t+1} \right]
\]

(B.6)

which have the conditional moments

\[
E_t[M_{0,t+1}P_0(t+1, T-1) + \pi_{01}E_t[P_t(t, T) - P_0(t, T)]]
\]

\[
= \exp \left[ -x_t - \frac{\lambda_0^2}{2}x_t - A_0(T-1) - B_0(T-1)E_t[x_{t+1}] \right]
\]

\[
+ \pi_{01} \left\{ \exp \left[ A_1(T) + B_1(T)x_t \right] - \exp \left[ -A_0(T) - B_0(T)x_t \right] \right\}
\]

(B.7)

\[
Var[M_{0,t+1}P_0(t+1, T-1)] = \exp \left[ \lambda_0 + B_0(T-1)\sigma_0 \right]^2 x_t
\]

Similarly, conditional on i=1 we have

\[
M_{1,t+1}P_1(t+1, T-1) + \pi_{10}[P_0(t, T) - P_1(t, T)]
\]

\[
= \exp \left[ -x_t - \frac{\lambda_1^2}{2}x_t - \lambda_1 \sqrt{x_t}u_{t+1} - A_1(T-1) - B_1(T-1)x_{t+1} \right]
\]

(B.8)

with the conditional moments

\[
E_t[M_{1,t+1}P_1(t+1, T-1) + \pi_{10}E_t[P_0(t, T) - P_1(t, T)]]
\]

\[
= \exp \left[ -x_t - \frac{\lambda_1^2}{2}x_t - A_1(T-1) - B_1(T-1)E_t[x_{t+1}] \right]
\]

\[
+ \pi_{10} \left\{ \exp \left[ A_0(T) + B_0(T)x_t \right] - \exp \left[ -A_1(T) - B_1(T)x_t \right] \right\}
\]

(B.9)

\[
Var[M_{1,t+1}P_1(t+1, T-1)] = \exp \left[ \lambda_1 + B_1(T-1)\sigma_1 \right]^2 x_t
\]

Substituting the above conditional moments into (B.4) and (B.5) we write

\[
\exp \left[ -A_0(T) - B_0(T)x_t \right] =
\]

\[
\exp \left\{ -x_t - \frac{\lambda_0^2}{2}x_t - A_0(T-1) - B_0(T-1)E_t[x_{t+1}] + \frac{1}{2}[\lambda_0 + B_0(T-1)\sigma_0]^2 x_t \right\}
\]

\[
+ \pi_{01} \left\{ \exp \left[ A_1(T) + B_1(T)x_t \right] - \exp \left[ -A_0(T) - B_0(T)x_t \right] \right\}
\]

(B.10)
\[
\exp \left[ -A_1(T) - B_1(T)x_t \right] = \\
\exp \left\{ -x_t - \frac{\kappa^2}{2} x_t - A_1(T - 1) - B_1(T - 1)E_t[x_{t+1}] + \frac{1}{2} [\lambda_1 + B_1(T - 1)\sigma_1]^2 x_t \right\} \\
+ \pi_{10} \left\{ \exp \left[ A_0(T) + B_0(T)x_t \right] - \exp \left[ -A_1(T) - B_1(T)x_t \right] \right\} 
\]

Using the log-linear approximation \( \exp^x \approx 1 + x \) as in Bansal and Zhou (2002) and separating the terms we obtain

\[
\begin{align*}
A_0(T) &= A_0(T - 1) + \kappa_0\alpha_0 B_0(T - 1) + \pi_{01} [A_1(T) - A_0(T)] \\
A_1(T) &= A_1(T - 1) + \kappa_1 \alpha_1 B_1(T - 1) + \pi_{10} [A_0(T) - A_1(T)]
\end{align*}
\]

which is in matrix form:

\[
\begin{bmatrix}
A_0(T) \\
A_1(T)
\end{bmatrix} = 
\begin{bmatrix}
1 + \pi_{01} & -\pi_{01} \\
-\pi_{10} & 1 + \pi_{10}
\end{bmatrix}^{-1} 
\begin{bmatrix}
A_0(T - 1) + \kappa_0\alpha_0 B_0(T - 1) \\
A_1(T - 1) + \kappa_1 \alpha_1 B_1(T - 1)
\end{bmatrix},
\]

\[
- B_0(T) = -1 - (1 - \kappa_0) B_0(T - 1) - \pi_{01} [B_1(T) - B_0(T)] \\
+ \frac{1}{2} B_0(T - 1)^2 \sigma_0^2 + \lambda_0 B_0(T - 1) \sigma_0 \quad (B.14)
\]

\[
- B_1(T) = -1 - (1 - \kappa_1) B_1(T - 1) - \pi_{10} [B_0(T) - B_1(T)] \\
+ \frac{1}{2} B_1(T - 1)^2 \sigma_1^2 + \lambda_1 B_1(T - 1) \sigma_1 \quad (B.15)
\]

which is in matrix form:

\[
\begin{bmatrix}
B_0(T) \\
B_1(T)
\end{bmatrix} = 
\begin{bmatrix}
1 + \pi_{01} & -\pi_{01} \\
-\pi_{10} & 1 + \pi_{10}
\end{bmatrix}^{-1} 
\begin{bmatrix}
1 + (1 - \kappa_0 - \lambda_0) B_0(T - 1) - \frac{1}{2} \sigma_0^2 B_0(T - 1)^2 \\
1 + (1 - \kappa_1 - \lambda_1) B_1(T - 1) - \frac{1}{2} \sigma_1^2 B_1(T - 1)^2
\end{bmatrix} 
\]

Unfortunately, the above equations system 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 (B.13) and (B.16). The difference between these two gives the approximation error.
Appendix C. Models 2-5: The Cox-Ingersoll-Ross model with regime switching and regime switching risk

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

\[ M_{t+1} = \exp \left[ -r_t^f - \frac{\lambda_{s_t+1}^2}{2} x_t - \lambda_{s_t+1} \sqrt{x_t} u_{t+1} - \frac{\Gamma_{s_t+1}^2}{2} - \sum_{i \neq j} \Gamma^j (\Psi_{t+1}^j - H \Psi_t^j) \right] \], \quad (C.1) 

and

\[ x_{t+1} - x_t = \kappa_{s_t+1} [x_t] + \sigma_{s_t+1} \sqrt{x_t} u_{t+1}. \] \quad (C.2)

The fundamental bond pricing equation (9) for this regime switching case becomes

\[ P_t(t, T) = E_t \left\{ \left( \Psi_{t+1}^i \right)^{0} \left[ \begin{array}{c} M_{t+1,0} P_0(t + 1, T - 1) \\ M_{t+1,1} P_1(t + 1, T - 1) \end{array} \right] \right\}. \] \quad (C.3)

Conditional on the current regime \( i \) we can write

\[ P_0(t, T) = E_t \left[ M_{0,t+1} P_0(t + 1, T - 1) + \Phi_0 \left[ M_{1,t+1} P_1(t + 1, T - 1) - M_{0,t+1} P_0(t + 1, T - 1) \right] \right] \]

\[ = E_t \left[ M_{0,t+1} P_0(t + 1, T - 1) + \pi_{01} [P_1(t, T) - P_0(t, T)] \right. \]

\[ + \Gamma^0 [P_1(t, T) - P_0(t, T)] \quad \text{if } i = 0, \] \quad (C.4)

\[ P_1(t, T) = E_t \left[ M_{1,t+1} P_1(t + 1, T - 1) + \Phi_1 \left[ M_{0,t+1} P_0(t + 1, T - 1) - M_{1,t+1} P_1(t + 1, T - 1) \right] \right] \]

\[ = E_t \left[ M_{1,t+1} P_1(t + 1, T - 1) + \pi_{10} [P_0(t, T) - P_1(t, T)] \right. \]

\[ + \Gamma^1 [P_0(t, T) - P_1(t, T)] \quad \text{if } i = 1. \] \quad (C.5)

Using the following regime adjusted affine functional form for bond prices

\[ P_t(t, T) = \exp \left[ -A_t(T) - B_t(T) x_t \right], \]

we obtain the expressions \( P_t(t, T) \) and hence \( P_{s_t+1}(t + 1, T - 1) \) required in (C.4) and (C.5).

The terms of (C.4) and (C.5) are obtained as follows:

\[ A_t(T) + B_t(T) x_t = [A_{s_t+1}(T - 1) + B_{s_t+1}(T - 1) \kappa \alpha] + [1 + B(T - 1)(1 - \kappa)] x_t \]

\[ - \frac{1}{2} B(T - 1)^2 \sigma^2 x_t - B(T - 1) \lambda \sigma x_t. \] \quad (C.6)
The first terms on the right side are obtained as follows:

\[
M_{0,t+1}P_0(t + 1, T - 1) = \exp \left[ -x_t - \frac{\lambda_0^2}{2} x_t - \lambda_0 \sqrt{x_t} u_{t+1} - \frac{\Gamma_0^2}{2} - \sum_{i \neq j} \Gamma^j (\Psi^j_{t+1} - H \Psi^j_t) \right. \\
\left. - A_0(T - 1) - B_0(T - 1)x_{t+1} \right] \tag{C.7}
\]

which have the conditional moments

\[
E_t[M_{0,t+1}P_0(t + 1, T - 1)] = \exp \left[ -x_t - \frac{\lambda_0^2}{2} x_t - \frac{\Gamma_0^2}{2} - A_0(T - 1) - B_0(T - 1)E_t[x_{t+1}] \right] \tag{C.8}
\]

\[
\text{Var}[M_{0,t+1}P_0(t + 1, T - 1)] = \exp \left[ \lambda_0 + B_0(T - 1)\sigma_0 \right]^2 x_t + \Gamma_0^2
\]

Similarly, conditional on \( i=1 \) we have

\[
M_{1,t+1}P_1(t + 1, T - 1) = \exp \left[ -x_t - \frac{\lambda_1^2}{2} x_t - \lambda_1 \sqrt{x_t} u_{t+1} - \frac{\Gamma_1^2}{2} - \sum_{i \neq j} \Gamma^j (\Psi^j_{t+1} - H \Psi^j_t) \right. \\
\left. - A_1(T - 1) - B_1(T - 1)x_{t+1} \right] \tag{C.9}
\]

with the conditional moments

\[
E_t[M_{1,t+1}P_1(t + 1, T - 1)] = \exp \left[ -x_t - \frac{\lambda_1^2}{2} x_t - \frac{\Gamma_1^2}{2} - A_1(T - 1) - B_1(T - 1)E_t[x_{t+1}] \right]. \tag{C.10}
\]

\[
\text{Var}[M_{1,t+1}P_1(t + 1, T - 1)] = \exp \left[ \lambda_1 + B_1(T - 1)\sigma_1 \right]^2 x_t + \Gamma_1^2
\]

Substituting the above conditional moments into (C.4) and (C.5) we write

\[
\exp \left[ -A_0(T) - B_0(T)x_t \right] = \\
\exp \left\{ -x_t - \frac{\lambda_0^2}{2} x_t - \frac{\Gamma_0^2}{2} - A_0(T - 1) - B_0(T - 1)E_t[x_{t+1}] + \frac{1}{2} [\lambda_0 + B_0(T - 1)\sigma_0]^2 x_t + \frac{\Gamma_0^2}{2} \right\} \\
+ (\pi_{01} + \Gamma_0) \left\{ \exp \left[A_1(T) + B_1(T)x_t \right] - \exp \left[ -A_0(T) - B_0(T)x_t \right] \right\} \tag{C.11}
\]

\[
\exp \left[ -A_1(T) - B_1(T)x_t \right] = \\
\exp \left\{ -x_t - \frac{\lambda_1^2}{2} x_t - \frac{\Gamma_1^2}{2} - A_1(T - 1) - B_1(T - 1)E_t[x_{t+1}] + \frac{1}{2} [\lambda_1 + B_1(T - 1)\sigma_1]^2 x_t + \frac{\Gamma_1^2}{2} \right\} \\
+ (\pi_{10} + \Gamma_1) \left\{ \exp \left[A_0(T) + B_0(T)x_t \right] - \exp \left[ -A_1(T) - B_1(T)x_t \right] \right\} \tag{C.12}
\]

24
Using the log-linear approximation \( \exp^x \approx 1 + x \) as in Bansal and Zhou (2002) we obtain

\[
A_0(T) = A_0(T - 1) + \kappa_0 \alpha_0 B_0(T - 1) + (\pi_{01} + \Gamma_0)[A_1(T) - A_0(T)]
\]

\[
A_1(T) = A_1(T - 1) + \kappa_1 \alpha_1 B_1(T - 1) + (\pi_{10} + \Gamma_1)[A_0(T) - A_1(T)]
\]

which is in matrix form:

\[
\begin{bmatrix}
A_0(T) \\
A_1(T)
\end{bmatrix} = 
\begin{bmatrix}
1 + \pi_{01} + \Gamma_0 & -\pi_{01} - \Gamma_0 \\
-\pi_{10} - \Gamma_1 & 1 + \pi_{10} + \Gamma_1
\end{bmatrix}^{-1}
\begin{bmatrix}
A_0(T - 1) + \kappa_0 \alpha_0 B_0(T - 1) \\
A_1(T - 1) + \kappa_1 \alpha_1 B_1(T - 1)
\end{bmatrix},
\]

\[
(B_0(T)) = -1 - (1 - \kappa_0)B_0(T - 1) - (\pi_{01} + \Gamma_0)[B_1(T) - B_0(T)] + \frac{1}{2}B_0(T - 1)^2 \sigma_0^2 + \lambda_0 B_0(T - 1) \sigma_0
\]

\[
(B_1(T)) = -1 - (1 - \kappa_1)B_1(T - 1) - (\pi_{10} + \Gamma_1)[B_0(T) - B_1(T)] + \frac{1}{2}B_1(T - 1)^2 \sigma_1^2 + \lambda_1 B_1(T - 1) \sigma_1
\]

which is in matrix form:

\[
\begin{bmatrix}
B_0(T) \\
B_1(T)
\end{bmatrix} = 
\begin{bmatrix}
1 + \pi_{01} + \Gamma_0 & -\pi_{01} - \Gamma_0 \\
-\pi_{10} - \Gamma_1 & 1 + \pi_{10} + \Gamma_1
\end{bmatrix}^{-1}
\begin{bmatrix}
1 + (1 - \kappa_0 - \lambda_0)B_0(T - 1) - \frac{1}{2} \sigma_0^2 B_0(T - 1)^2 \\
1 + (1 - \kappa_1 - \lambda_1)B_1(T - 1) - \frac{1}{2} \sigma_1^2 B_1(T - 1)^2
\end{bmatrix}
\]

(C.13)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard</td>
<td>Mean</td>
</tr>
<tr>
<td>3</td>
<td>4.15023</td>
<td>1.98836</td>
<td>12.0278</td>
</tr>
<tr>
<td>6</td>
<td>4.38941</td>
<td>2.04902</td>
<td>12.2170</td>
</tr>
<tr>
<td>12</td>
<td>4.56006</td>
<td>2.03908</td>
<td>12.1098</td>
</tr>
<tr>
<td>24</td>
<td>4.72129</td>
<td>1.96039</td>
<td>11.8012</td>
</tr>
<tr>
<td>60</td>
<td>4.96496</td>
<td>1.87292</td>
<td>11.4083</td>
</tr>
<tr>
<td>120</td>
<td>5.10668</td>
<td>1.84290</td>
<td>11.2681</td>
</tr>
</tbody>
</table>
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.
  \[ dr(t) = \kappa(X_t)(\alpha - r(t))dt + \sigma(X_t)\sqrt{r(t)}dW(t) \]

- 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>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.0161</td>
<td>-</td>
<td>0.0296</td>
<td>0.0149</td>
<td>0.0154</td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0016)</td>
<td>(0.0021)</td>
<td>(0.0025)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.0177</td>
<td>0.0128</td>
<td>0.0130</td>
<td>0.0128</td>
<td>0.0129</td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0009)</td>
<td>(0.0008)</td>
<td>(0.0008)</td>
<td>(0.0008)</td>
</tr>
<tr>
<td>$\kappa_0$</td>
<td>0.0684</td>
<td>-</td>
<td>0.0118</td>
<td>-</td>
<td>0.0614</td>
</tr>
<tr>
<td></td>
<td>(0.0344)</td>
<td></td>
<td>(0.0080)</td>
<td></td>
<td>(0.0340)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-</td>
<td>0.0154</td>
<td>-</td>
<td>0.0337</td>
<td>0.0296</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0239)</td>
<td></td>
<td>(0.0016)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-</td>
<td>0.0399</td>
<td>0.0310</td>
<td>0.0388</td>
<td>0.0311</td>
</tr>
<tr>
<td></td>
<td>(0.0082)</td>
<td>(0.0064)</td>
<td>(0.0080)</td>
<td>(0.0064)</td>
<td></td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>-</td>
<td>0.0635</td>
<td>1.0234</td>
<td>0.0725</td>
<td>0.9905</td>
</tr>
<tr>
<td></td>
<td>(0.0322)</td>
<td>(0.3387)</td>
<td>(0.0360)</td>
<td>(0.3417)</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>-</td>
<td>0.9039</td>
<td>0.9115</td>
<td>0.9107</td>
<td>0.9136</td>
</tr>
<tr>
<td></td>
<td>(0.0849)</td>
<td>(0.0796)</td>
<td>(0.0806)</td>
<td>(0.0777)</td>
<td></td>
</tr>
<tr>
<td>$q$</td>
<td>-</td>
<td>0.9899</td>
<td>0.9925</td>
<td>0.9906</td>
<td>0.9926</td>
</tr>
<tr>
<td></td>
<td>(0.0108)</td>
<td>(0.0075)</td>
<td>(0.0098)</td>
<td>(0.0075)</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>SIC</th>
<th>HQ</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Model 4</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Model 5</td>
<td>1990:3 - 1992:4</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>1995:3 - 1998:4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Cumulative Recursive AIC, SIC and HQ

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>SIC</th>
<th>HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>-35838.177</td>
<td>-35297.451</td>
<td>-35619.420</td>
</tr>
<tr>
<td>Model 2</td>
<td>-37573.112</td>
<td>-36491.661</td>
<td>-37135.599</td>
</tr>
<tr>
<td>Model 3</td>
<td>-37774.079</td>
<td>-36512.386</td>
<td>-37263.648</td>
</tr>
<tr>
<td>Model 4</td>
<td>-37476.543</td>
<td>-36214.849</td>
<td>-36966.110</td>
</tr>
<tr>
<td>Model 5</td>
<td>-37745.557</td>
<td>-36303.622</td>
<td>-37162.206</td>
</tr>
</tbody>
</table>
Table 6. Out of Sample Performance Results of Models

<table>
<thead>
<tr>
<th>$h$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Squared Error (MSE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>1.00</td>
<td>2.21</td>
<td>3.21</td>
<td>4.40</td>
<td>5.70</td>
<td>7.06</td>
<td>8.22</td>
<td>9.28</td>
<td>10.08</td>
<td>11.05</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.92</td>
<td>1.91</td>
<td>2.79</td>
<td>3.84</td>
<td>5.01</td>
<td>6.28</td>
<td>7.39</td>
<td>8.46</td>
<td>9.22</td>
<td>10.01</td>
</tr>
<tr>
<td>Model 3</td>
<td>31.15</td>
<td>29.71</td>
<td>29.56</td>
<td>29.17</td>
<td>28.83</td>
<td>28.28</td>
<td>27.74</td>
<td>26.97</td>
<td>26.51</td>
<td>25.70</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.94</td>
<td>2.03</td>
<td>3.43</td>
<td>5.11</td>
<td>6.88</td>
<td>8.67</td>
<td>10.31</td>
<td>11.83</td>
<td>12.91</td>
<td>13.73</td>
</tr>
<tr>
<td>Model 5</td>
<td>29.63</td>
<td>30.20</td>
<td>30.52</td>
<td>30.50</td>
<td>30.34</td>
<td>30.19</td>
<td>29.78</td>
<td>29.46</td>
<td>28.77</td>
<td></td>
</tr>
<tr>
<td>Relative Mean Squared Error (RMSE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.09</td>
<td>0.12</td>
<td>0.15</td>
<td>0.17</td>
<td>0.20</td>
<td>0.23</td>
<td>0.26</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
<td>0.08</td>
<td>0.10</td>
<td>0.13</td>
<td>0.15</td>
<td>0.18</td>
<td>0.21</td>
<td>0.23</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.73</td>
<td>0.70</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.01</td>
<td>0.04</td>
<td>0.06</td>
<td>0.09</td>
<td>0.13</td>
<td>0.16</td>
<td>0.20</td>
<td>0.24</td>
<td>0.27</td>
<td>0.30</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.71</td>
<td>0.72</td>
<td>0.74</td>
<td>0.75</td>
<td>0.76</td>
<td>0.77</td>
<td>0.78</td>
<td>0.79</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Mean Absolute Error (MAE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>0.71</td>
<td>1.10</td>
<td>1.41</td>
<td>1.69</td>
<td>1.92</td>
<td>2.16</td>
<td>2.33</td>
<td>2.52</td>
<td>2.67</td>
<td>2.82</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.67</td>
<td>1.04</td>
<td>1.31</td>
<td>1.54</td>
<td>1.78</td>
<td>1.99</td>
<td>2.18</td>
<td>2.40</td>
<td>2.55</td>
<td>2.67</td>
</tr>
<tr>
<td>Model 3</td>
<td>5.22</td>
<td>5.10</td>
<td>5.07</td>
<td>5.02</td>
<td>4.98</td>
<td>4.95</td>
<td>4.91</td>
<td>4.88</td>
<td>4.82</td>
<td>4.76</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.74</td>
<td>1.12</td>
<td>1.42</td>
<td>1.78</td>
<td>2.10</td>
<td>2.37</td>
<td>2.63</td>
<td>2.87</td>
<td>3.04</td>
<td>3.17</td>
</tr>
<tr>
<td>Model 5</td>
<td>5.12</td>
<td>5.19</td>
<td>5.20</td>
<td>5.18</td>
<td>5.19</td>
<td>5.18</td>
<td>5.17</td>
<td>5.13</td>
<td>5.09</td>
<td></td>
</tr>
<tr>
<td>Relative Mean Absolute Error (RMAE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>0.11</td>
<td>0.17</td>
<td>0.21</td>
<td>0.25</td>
<td>0.29</td>
<td>0.34</td>
<td>0.36</td>
<td>0.40</td>
<td>0.43</td>
<td>0.46</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.10</td>
<td>0.15</td>
<td>0.19</td>
<td>0.23</td>
<td>0.27</td>
<td>0.31</td>
<td>0.34</td>
<td>0.38</td>
<td>0.41</td>
<td>0.43</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.83</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>0.11</td>
<td>0.16</td>
<td>0.21</td>
<td>0.26</td>
<td>0.31</td>
<td>0.35</td>
<td>0.40</td>
<td>0.44</td>
<td>0.47</td>
<td>0.50</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.81</td>
<td>0.83</td>
<td>0.84</td>
<td>0.84</td>
<td>0.85</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>Theil’s Inequality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>0.07</td>
<td>0.12</td>
<td>0.15</td>
<td>0.18</td>
<td>0.21</td>
<td>0.25</td>
<td>0.28</td>
<td>0.30</td>
<td>0.33</td>
<td>0.36</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.07</td>
<td>0.10</td>
<td>0.13</td>
<td>0.16</td>
<td>0.19</td>
<td>0.22</td>
<td>0.25</td>
<td>0.28</td>
<td>0.30</td>
<td>0.33</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.62</td>
<td>0.62</td>
<td>0.63</td>
<td>0.64</td>
<td>0.64</td>
<td>0.65</td>
<td>0.65</td>
<td>0.66</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>0.07</td>
<td>0.11</td>
<td>0.15</td>
<td>0.19</td>
<td>0.24</td>
<td>0.28</td>
<td>0.32</td>
<td>0.35</td>
<td>0.38</td>
<td>0.41</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.60</td>
<td>0.63</td>
<td>0.65</td>
<td>0.67</td>
<td>0.68</td>
<td>0.70</td>
<td>0.71</td>
<td>0.73</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>$h$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.78</td>
<td>0.83</td>
<td>0.88</td>
<td>0.87</td>
<td>0.93</td>
<td>0.97</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.35</td>
<td>0.23</td>
<td>0.70</td>
<td>0.79</td>
<td>0.78</td>
<td>0.81</td>
<td>0.95</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.83</td>
<td>0.84</td>
<td>0.90</td>
<td>0.89</td>
<td>0.94</td>
<td>0.97</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Notes: The figures are p-values
Table 8. In Sample Performance Results of Models

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

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

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: ( dr(t) = \kappa[\alpha - r(t)]dt + \sigma_r(t)dW(t) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Month</td>
<td>0.5601</td>
<td>0.5306</td>
<td>1.52787</td>
<td>2.7184</td>
</tr>
<tr>
<td>1 Year</td>
<td>0.7827</td>
<td>1.0997</td>
<td>1.5763</td>
<td>2.9075</td>
</tr>
<tr>
<td>2 Year</td>
<td>0.6570</td>
<td>0.8339</td>
<td>1.6436</td>
<td>3.2214</td>
</tr>
<tr>
<td>5 Year</td>
<td>0.2871</td>
<td>0.1728</td>
<td>1.5960</td>
<td>3.2651</td>
</tr>
<tr>
<td>Model 2: ( dr(t) = \kappa[\alpha - r(t)]dt + \sigma(X_t)^2r(t)dW(t) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Month</td>
<td>-0.1135</td>
<td>0.0614</td>
<td>-0.0462</td>
<td>0.0466</td>
</tr>
<tr>
<td>1 Year</td>
<td>-0.2495</td>
<td>0.2042</td>
<td>-0.2381</td>
<td>0.3687</td>
</tr>
<tr>
<td>2 Year</td>
<td>-0.3492</td>
<td>0.2986</td>
<td>-0.3605</td>
<td>0.5860</td>
</tr>
<tr>
<td>5 Year</td>
<td>-0.1984</td>
<td>0.1181</td>
<td>-0.0776</td>
<td>0.0668</td>
</tr>
<tr>
<td>Model 3: ( dr(t) = \kappa(X_t)[\alpha - r(t)]dt + \sigma(X_t)^2r(t)dW(t) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Month</td>
<td>-0.4118</td>
<td>0.3592</td>
<td>-0.3205</td>
<td>0.4094</td>
</tr>
<tr>
<td>1 Year</td>
<td>-0.8881</td>
<td>1.2558</td>
<td>-1.9432</td>
<td>3.4012</td>
</tr>
<tr>
<td>2 Year</td>
<td>-1.1255</td>
<td>1.7106</td>
<td>-2.8431</td>
<td>5.0770</td>
</tr>
<tr>
<td>5 Year</td>
<td>-0.6273</td>
<td>0.4848</td>
<td>-0.4073</td>
<td>0.3610</td>
</tr>
<tr>
<td>Model 4: ( dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)^2r(t)dW(t) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Month</td>
<td>-0.1659</td>
<td>0.0704</td>
<td>-0.0434</td>
<td>0.0354</td>
</tr>
<tr>
<td>1 Year</td>
<td>-0.3812</td>
<td>0.2711</td>
<td>-0.2459</td>
<td>0.2891</td>
</tr>
<tr>
<td>2 Year</td>
<td>-0.5654</td>
<td>0.4625</td>
<td>-0.4464</td>
<td>0.5269</td>
</tr>
<tr>
<td>5 Year</td>
<td>-0.3965</td>
<td>0.2185</td>
<td>-0.1404</td>
<td>0.1061</td>
</tr>
<tr>
<td>Model 5: ( dr(t) = \kappa(X_t)[\alpha(X_t) - r(t)]dt + \sigma(X_t)^2r(t)dW(t) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Month</td>
<td>-0.4264</td>
<td>0.3635</td>
<td>-0.3260</td>
<td>0.3926</td>
</tr>
<tr>
<td>1 Year</td>
<td>-0.8930</td>
<td>1.2511</td>
<td>-1.9041</td>
<td>3.2248</td>
</tr>
<tr>
<td>2 Year</td>
<td>-1.1031</td>
<td>1.6432</td>
<td>-2.6718</td>
<td>4.6273</td>
</tr>
<tr>
<td>5 Year</td>
<td>-0.5979</td>
<td>0.4449</td>
<td>-0.3614</td>
<td>0.3102</td>
</tr>
</tbody>
</table>
Table 10. Kolmogorov-Smirnov Tests

<table>
<thead>
<tr>
<th>Model</th>
<th>6 Months</th>
<th>1 Year</th>
<th>2 Year</th>
<th>5 Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.1912</td>
<td>0.2647</td>
<td>0.2206</td>
<td>0.1176</td>
</tr>
<tr>
<td></td>
<td>(0.1665)</td>
<td>(0.0171)</td>
<td>(0.0731)</td>
<td>(0.7344)</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.0735</td>
<td>0.1324</td>
<td>0.1176</td>
<td>0.0735</td>
</tr>
<tr>
<td></td>
<td>(0.9929)</td>
<td>(0.5907)</td>
<td>(0.7344)</td>
<td>(0.9929)</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.2206</td>
<td>0.3088</td>
<td>0.3382</td>
<td>0.1765</td>
</tr>
<tr>
<td></td>
<td>(0.0731)</td>
<td>(0.0031)</td>
<td>(0.0008)</td>
<td>(0.2402)</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.1176</td>
<td>0.1765</td>
<td>0.1912</td>
<td>0.1176</td>
</tr>
<tr>
<td></td>
<td>(0.7344)</td>
<td>(0.2402)</td>
<td>(0.1665)</td>
<td>(0.7344)</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.2206</td>
<td>0.3088</td>
<td>0.3382</td>
<td>0.1765</td>
</tr>
<tr>
<td></td>
<td>(0.0731)</td>
<td>(0.0031)</td>
<td>(0.0008)</td>
<td>(0.2402)</td>
</tr>
</tbody>
</table>

Notes: The figures in parentheses are p-values