The effects of different parameterizations of Markov-switching in a CIR model of bond pricing

Journal Article

Citation:

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

Studies in Nonlinear Dynamics & Econometrics

Volume 13, *Issue* 1 2009 *Article* 1

The Effects of Different Parameterizations of Markov-Switching in a CIR Model of Bond Pricing

John Driffill* Turalay Kenc†
Martin Sola‡ Fabio Spagnolo**

*Birkbeck College, University of London, j.driffill@bbk.ac.uk
†Bradford University School of Management, t.kenc@bradford.ac.uk
‡University of London and Universidad Torcuato Di Tella, msola@econ.bbk.ac.uk
**Brunel University, fabio.spagnolo@brunel.ac.uk

The Effects of Different Parameterizations of Markov-Switching in a CIR Model of Bond Pricing*

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

Abstract

We examine several discrete-time versions of the Cox, Ingersoll and Ross (CIR) model for the term structure, in which the short rate is subject to discrete shifts. Our empirical analysis suggests that careful consideration of which parameters of the short-term interest rate equation that are allowed to be switched is crucial. Ignoring this issue may result in a parameterization that produces no improvement (in terms of bond pricing) relative to the standard CIR model, even when there are clear breaks in the data.

*We thank David Barr, Keith Cuthbertson, Demian Pouzo, Marzia Raybaudi, Daniel Thornton, and 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 correspondences 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 a process subject to structural breaks is to assume that the breaks follow a Markov chain as in (Hamilton, 1988, 1989). Although an extensive literature uses this approach, few papers systematically study the specification of the switching regression. Questions such as: *(i)* which parameters are allowed to switch [see Hall and Sola (1993)]; *(ii)* how many states should be allowed for [see (Hansen, 1992, 1996), Garcia (1998), and (Psaradakis and Spagnolo, 2003, 2006)]; *(iii)* how many lags should be included in the switching regression [see Kapetanios (2001)] have attracted comparatively little attention. The importance of each of these questions varies with the application at hand.[1] The correct specification of the switching process is crucial when forecasting, or for applications that involve rational expectations. Different specifications typically imply different forecasts and pricing equations, and consequently affect conclusions about the validity of any theoretical model.

Short and long-term interest rates have been characterized as a stochastic process subject to regime switches [see, for example, Hamilton (1988), Sola and Driffill (1994), Garcia and Perron (1996), Gray (1996), Dahlquist and Gray (2000), Landen (2000), (Ang and Bekaert, 2002a,b), Bansal and Zhou (2002), Smith (2002), Evans (2003), and Dai and Singleton (2003)]. In particular, Gray (1996) showed that a version of the Cox, Ingersoll, and Ross (1985) (CIR) model with time-varying parameters provides an appropriate characterization of US short-term interest rate data. While some of the papers that use the CIR model only allow the volatility of the short-term interest rate to switch [e.g., Naik and Lee (1997)], others, such as Bansal and Zhou (2002), allow all the parameters to switch [see also Dahlquist and Gray (2000) and (Ang and Bekaert, 2002a,b)]. However, none of these papers evaluate the performance of different specifications, either in terms of fit, or in terms of real-time one-step-ahead bond pricing. This paper attempts to fill this gap by evaluating how different parameterizations of the switching process affect bond prices. Therefore, we do not ask, as most of the literature does, which of the many competing models best fit the data in-sample; but, for a given model, what are the effects of allowing all the parameters of the exogenous driving equation to switch (as is standard these days in the literature) on the one step ahead bond prices.[2]

We rank the different versions of the CIR model in terms of their ability to generate prices close to the data. Our approach is based on recursively estimating the different parameterizations of the switching CIR process, and using the results to price bonds of different maturities. In this way we generate a series of prices which are then compared with the actual prices.[3]

These results are then compared with those obtained using standard likelihood ratio tests and complexity-penalized likelihood criteria.[4] We find that the models which provide

[1]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.

[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 speculate that the point raised in this paper is equally important (or probably more important given the nature of the driving process) for other more complex affine switching models. We explain in detail in the text why we think that this will be the case.

[4]Such methods have enjoyed much popularity in statistics as a means of choosing among competing

the best fit (and those versions are not rejected by the data) do not necessarily provide the best bond prices. This result seem to suggest that the tendency in the literature to estimate more general and complicated models (with time varying transitions and adding more factors) may simply improve the fit while worsening the one step ahead pricing (which is the main interest of the practitioner) and the forecasting performance of the model [see for example Diebold and Li (2006)].

The main results of the paper are that simpler specifications such as a Markov-switching CIR (MS-CIR) short rate with only regime-dependent volatility and with regime-dependent volatility and long-run interest rate produce better bond prices than those obtained using parameterizations with no regime switching, models where all the parameters are allowed to switch, and parameterizations with both regime-dependent volatility and regime-dependent speed of adjustment. We also find that the pricing gains of Markov-switching models diminish the further away from the break the bond price is evaluated, to the extent that, eventually, the CIR outperforms the MS-CIR model. Indeed, on the basis of criteria which evaluate the ability of the models to correctly predict turning points (i.e. whether rates are rising or falling), the 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.

The plan of the paper is as follows. In section 2 we present the benchmark model and its extensions which are used to evaluate the empirical issues. Section 3 considers using bond pricing as a model selection criterion for the interest rate. Section 4 summarizes and concludes.

2 Term Structure Models

In this section, we present the Cox, Ingersoll and Ross (CIR) benchmark model that is used to evaluate the alternative empirical issues addressed in the paper. We modify the original CIR model as in Bansal and Zhou (2002) to allow the short-term interest rate to switch between regimes. The benchmark model and its extensions are presented below. Notice that Bansal and Zhou (2002) seem to suggest, that a two factor model has a better fit in sample. Nevertheless, it is not obvious that we can draw any conclusions from their results for our paper since we not only use a different sample period (our data goes up to 1998 instead of 1995) but also, and most importantly, use a different sample frequency. Notice also that there are many differences in emphasis between this paper and that of Bansal and Zhou (2002). First, we study the relative importance of different assumptions which are common in the literature, rather than purposing a new model to explain the term structure. Second, we mostly focus in the out of sample performance of the models rather than trying to explain the model that better prices the past. Third, we the compare full and real sample performance of the different bond prices.

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 (1997).

2.1 The Benchmark Cox, Ingersoll and Ross (CIR) Model

We first consider the benchmark CIR model in which a single factor x, typically associated with the short rate r, follows a mean-reverting square root process. The discrete-time version of the CIR process for the single factor is written as

$$x_{t+1} - x_t = \kappa[\theta - x_t] + \sigma\sqrt{x_t}u_{t+1}, \tag{1}$$

with $\{u_{t+1}\}$ distributed normally, independently, with mean zero and variance one. The factor reverts to a long-term mean value θ. The parameter κ determines the adjustment speed of x towards the long-term mean, and $\sigma^2 x$ is the variance of the unexpected changes in the factor. The term σ is the local volatility and serves as a scaling parameter. The pricing kernel (stochastic discount factor), M, for a discrete time version of the CIR model is

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

We refer to λ as the *market price of factor risk*, since it determines the covariance between shocks to M and x, and thus the risk characteristics of bonds and related assets. Note that $E_t[M_{t+1}] = \exp(-r_t^f)$, where r_t^f is the one-period risk-free rate. We assume that, for every τ, the price of a maturity τ-bond has the form:

$$P_t^\tau = \exp[-A_\tau - B_\tau x_t]. \tag{3}$$

2.2 Regime Shifts

We account for regime switches by assuming that the parameters $\kappa(s_t)$, $\theta(s_t)$ and $\sigma(s_t)$ in eq. (1) take different values in different regimes s_t. We model s_t as a two-state Markov process which takes values of either 0 (regime 0) or 1 (regime 1). The switch between the regimes is governed by a Markov chain with a transition probability matrix $\Pi = (\pi_{ij})$:

$$\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}, \tag{4}$$

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 π_{01}. Similarly, π_{ii} 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.[5]

The Markov-switching mean-reverting square root process (MS-CIR) can be written as follows,

$$x_{t+1} - x_t = \kappa(s_{t+1})[\theta(s_{t+1}) - x_t] + \sigma(s_{t+1})\sqrt{x_t}u_{t+1}. \tag{5}$$

Following Bansal and Zhou (2002) we model the market price of random risk as regime dependent: $\lambda(s_{t+1})$. The pricing kernel therefore needs to be adjusted for regime shifts as follows

$$M_{t+1}(s_{t+1}) = \exp\left[-r_t^f - \left(\frac{\lambda(s_{t+1})}{\sigma(s_{t+1})}\right)^2 \frac{x_t}{2} - \frac{\lambda(s_{t+1})}{\sigma(s_{t+1})}\sqrt{x_t}u_{t+1}\right]. \tag{6}$$

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

2.3 Bond Pricing

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

$$P_t^\tau(s_t) \ = \ \exp\left[-A_\tau(s_t) \ - \ B_\tau(s_t)x_t\right], \tag{7}$$

where A and B are deterministic functions.

To ensure that the bond prices satisfy the no-arbitrage condition we use the fundamental pricing equation

$$P_t^\tau(s_t) = E_t[M_{t+1}(s_{t+1})P_{t+1}^{\tau-1}(s_{t+1})]. \tag{8}$$

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

$$\log P_t^\tau(s_t) = E_t[\log M_{t+1}(s_{t+1}) + \log P_{t+1}^{\tau-1}(s_{t+1})] + \tfrac{1}{2}Var_t[\log M_{t+1}(s_{t+1}) + \log P_{t+1}^{\tau-1}(s_{t+1})]. \tag{9}$$

This equation is then used to obtain the constants A and B, using equations (1), (2) and (3) for the single-regime model and equations (5), (6) and (7) for the switching-regime model. The corresponding solutions are provided, respectively, in appendices A and B. Once the constants A and B are obtained, bond yields are calculated as follows:

$$y_t^\tau(s_t) = -\frac{\log P_t^\tau(s_t)}{\tau} = \frac{A_\tau(s_t)}{\tau} + \frac{B_\tau(s_t)}{\tau}x_t \qquad i = 0, 1. \tag{10}$$

3 Markov-Switching Cox, Ingersoll and Ross (MS-CIR) Models with Switching Market Price of Factor Risk

In this section we inquire whether a common assumption made in the literature, that all the parameters of the instantaneous interest rate are allowed to switch between regimes, is important for bond pricing. We establish the relative performance of the pricing model under different assumptions about which parameters of the short rate are allowed to switch.

In principle overparameterized models might overfit the data and have a poor out-of-sample performance. Since pricing is intrinsically a forecasting exercise (because long term rates are, using the appropriate kernel, some kind of discounted average of the future expected short-term interest rate), we speculate that an overfitted model might also produce 'bad' bond prices (i.e., bond prices with big errors).

When using different versions of the single-factor Markov-switching Cox, Ingersoll and Ross model (depending on the assumed switching parameterization) to price bonds, an estimate of the market price of factor risk is required. The parameters λ_i, $(i = 0, 1)$, measuring the market price of factor risk, are estimated from the data. This strategy is based on a common assumption in the literature where the bond prices are observed with errors for some maturities [see for example Pearson and Sun (1994)]. This allows to jointly estimate the parameters λ_i along with the other parameters of the model. The yields with measurement error are given by:

$$y_t^\tau = \frac{A_\tau}{\tau} + \frac{B_\tau}{\tau}x_t + \varepsilon_t.$$

In estimating λ_i we assume that the yields on bonds with maturities 6 months and 5 years are measured with error. In this paper we do not impose the assumption that, for the other maturities under consideration (the 1, 2 and 10 years yields), the bonds are exactly observed (priced), but we use those maturities to evaluate the pricing performance of the alternative parameterizations by comparing the prices generated by the model with the actual price. Implicitly we get a measure of how strong is that assumption.

The assumption that some maturities are observed without error, and therefore are exactly priced, has contributed to the increasingly common use of highly parameterized models (i.e. models with several factors, models where all the parameters are allowed to switch and/or models with time varying probabilities). This is because, under the assumption that some maturities are observed without error, only complex models can fit the data in the sample. This strategy is usually defended on the grounds that it rules out arbitrage opportunities. Nevertheless we argue that this argument might be misleading because: *i*) the ex-ante pricing (or out-of-sample forecasting) performance of those highly parameterized models is usually very poor; *ii*) some of the maturities that are priced without error are, most of the time, synthetic and constructed from coupon paying bonds (that is, the data are by construction only an approximation).

3.1 Comparison Based on Goodness of Fit

For the estimation of the parameters of the model we use the 3 month T-Bill yield as a proxy of the instantaneous rate.[6] We use quarterly data to avoid the potential serial correlation which would be induced by the existence of overlapping expectations whenever the sampling frequency is higher than the maturity of the short term interest rate. The five models specified in *table 1* are estimated for the period 1964:1–1998:4, using the 3, 6 month bills and 5 years bond. We use the 1, 2, and 10 year bonds for the evaluation of the models.

The estimation of the different models for the short-term interest rate is 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 $\Theta = \{\kappa_0, \kappa_1, \theta_0, \theta_1, \sigma_0, \sigma_1, \lambda_0, \lambda_1, \Sigma_0, \Sigma_1\}$, subject to the constraint that $p = P(s_{t+1} = 1|s_t = 1)$ and $q = P(s_{t+1} = 0|s_t = 0)$ lie in the open unit interval (see Appendix C).[7]

In *table 2*, we report Gaussian standard pseudo-maximum likelihood (S–PML) estimates of the parameters along with the corresponding asymptotic standard errors.[8] Given the nature of the maximizing algorithm, we need to classify the regimes, not only in terms of the parameters of the switching CIR model, but also in terms of the state dependent variance-covariance matrix of the maturities priced with error. We find that the variances (for both maturities) of the pricing equations for the maturities observed with error in state 0 are higher than those variances in state 1, $\{\sigma^2_{0(6m)} > \sigma^2_{1(6m)}, \sigma^2_{0(5y)} > \sigma^2_{1(5y)}\}$ (below we offer an explanation for this finding). We find, for all models, that state 1 is more persistent that state 0, $\{\kappa_0 < \kappa_1\}$; that the volatility of the short term interest rate is higher in state 1 than in state 0, $\{\sigma_1 > \sigma_0\}$; and that (except for *model 5*) the long run value is higher in state 0 than in state 1,$\{\theta_0 > \theta_1\}$.[9] At this stage it should be clear that when many

[6]The data used in this paper are available on the web page associated with Duffee (2002).

[7]Σ_0 and Σ_1 are the variance covariance matrices of the pricing errors for the maturities assumed to be observed with error in state 0 and 1 respectively.

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

[9]In the estimation and pricing the interest rates are expressed in quarterly basis (instead of in annual basis) to avoid complicated transformations of the parameters when expressed in annual basis. We then do

pararameters are allowed to switch, even the definition of the regimes is cumbersome. This is aggravated by the fact that the variance-covariance matrix of the pricing error equations also is regime-dependent.[10]

In *figure 1* we plot all the maturities of Duffee (2002) along with the estimated filter probabilities. As explained above, the separation of the filter mostly associates regime 0 (regime 1) with: *i*) high (low) pricing errors (see the estimates of the variance-covariance matrix of the pricing error equation in state 0 (state 1) presented in *table 2*) and *ii*) high (low) variance of the short term interest rates and low (high) price of the risk (see *table 2*). From the top panel of *figure 1*, we can see that small pricing errors are associated with periods where the different interest rates are close. For those periods we expect more accurate prices and smaller pricing errors. On the other hand, the periods that the filter associates with state 0 are those where the different maturities are relatively more separated (the spreads are bigger) and therefore the pricing errors incurred by the different models are bigger.[11] Note that it seems that state 1 is broadly associated with periods when the interest rates, for all the maturities, increase and state 0 with the interest rate decrease.

In *table 2* we present the estimated switching CIR models. It shows that the hypothesis that *model 4* and *model 2* are valid simplifications of *model 5* (the general model) are rejected [the likelihood ratio test (LR) statistics are 6.46, distributed $\chi^2(1)$, and 6.50, distributed $\chi^2(2)$, respectively]. On the other hand, the null hypothesis that *model 3* is a valid reduction of the general model, is not rejected [the likelihood ratio test statistic is 2.96, 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 a clear cut indication of which model should be preferred in sample.[12] In order to establish whether these results are sensitive to the sample specifications we recursively estimate the five models described in *table 1* (starting from 1964:1-1991:2 and sequentially enlarging the sample up to 1998:4) and calculate, for each sample enlargement, the different complexity-penalized likelihood measures.[13]

In *table 3* we report results of recursive goodness of fit criteria and indicate periods during which each model is selected.[14] On the basis of the AIC criterion, *model 5* is preferred for the whole sample (1991:3 to 1998:4). On the other hand, using the SIC only *models 3* and *4* are selected, while the HQ criterion, with only the exceptions of short periods of time, always selects *model 5*. These results are further corroborated by *table 4*, which shows complexity-penalized likelihood cumulative measures, capturing both the time series and the cross section dimension. More specifically, while *model 5* is preferred on the basis of the

---

the pricing comparisons (we convert the generated data and the actual data) in annual basis. This implies that paramerters such as the long run value should be approximately 4 times bigger when expressed in annual bases than the values reported in *table 2*.

[10]In Bansal and Zhou (2002), the probabilities of the regimes are a direct function of the implied pricing errors (consistent with the model). Here they are also functions of the driving short term interest rate (see Appendix C).

[11]We can associate state 1 with periods when the term structure is "relatively flat" and state 0 with periods when it is not. The task of pricing seems to be easier in the first case and therefore it produces smaller pricing errors.

[12]Notice that for our models, the goodness of fit criterion is based on the joint estimation of the bond equations and the short-term interest rates.

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

[14]See also Swanson (1998) for a similar approach.

Table 1: Estimated Models

Models of the Short Term Interest Rate
Model 1: No regime switching.
$x_{t+1} - x_t = \kappa[\theta - x_t] + \sigma\sqrt{x_t}u_{t+1}$
Model 2: Regime switching in volatility.
$x_{t+1} - x_t = \kappa[\theta - x_t] + \sigma(s_{t+1})\sqrt{x_t}u_{t+1}$
Model 3: Regime switching in volatility and adjustment speed.
$x_{t+1} - x_t = \kappa(s_{t+1})[\theta - x_t] + \sigma(s_{t+1})\sqrt{x_t}u_{t+1}$
Model 4: Regime switching in volatility and long-run rate.
$x_{t+1} - x_t = \kappa[\theta(s_{t+1}) - x_t] + \sigma(s_{t+1})\sqrt{x_t}u_{t+1}$
Model 5: Regime switching in all parameters.
$x_{t+1} - x_t = \kappa(s_{t+1})[\theta(s_{t+1}) - x_t] + \sigma(s_{t+1})\sqrt{x_t}u_{t+1}$
Specifications of the Market Price of Factor Risk λ
Estimated market price of factor risk ($\lambda_{s_{t+1}}$) using 6m and 1y yields.

AIC and HQ criteria, *model 2* outperform the competing models on the basis of the SIC.

Table 2: Parameter Estimates of Models — 1964:1-1998:4

	Model 1	Model 2	Model 3	Model 4	Model 5
θ_0	0.016591 (0.004879)	–	0.015391 (0.003100)	0.017309 (0.005578)	0.013485 (0.002410)
σ_0	0.018060 (0.001119)	0.017872 (0.001014)	0.017785 (0.001095)	0.017914 (0.001038)	0.017587 (0.001069)
κ_0	0.041443 (0.0133914)	–	0.083249 (0.029663)	–	0.093959 (0.021740)
λ_0	-0.024980 (0.012353)	-0.056341 (0.019630)	-0.066637 (0.024080)	-0.053655 (0.020706)	-0.077668 (0.019342)
θ_1	–	0.016471 (0.004445)	–	0.015590 (0.038494)	0.051590 (0.028198)
σ_1	–	0.137030 (0.013975)	0.132365 (0.011857)	0.134704 (0.012431)	0.132606 (0.008745)
κ_1	–	0.051323 (0.015445)	0.041028 (0.018538)	0.052369 (0.016584)	0.012240 (0.019915)
λ_1	–	-0.085218 (0.018479)	-0.087160 (0.014369)	-0.087489 (0.0225985)	-0.058737 (0.020068)
p	–	0.961467 (0.049159)	0.961528 (0.046397)	0.968131 (0.042771)	0.960638 (0.026426)
q	–	0.857613 (0.037403)	0.855226 (0.040546)	0.857881 (0.387862)	0.856034 (0.278974)
$\sigma^2_{0(6m)}$	$5.9e^{-7}$ $(0.7e^{-7})$	$4.3e^{-7}$ $(0.7e^{-7})$	$4.3e^{-7}$ $(0.7e^{-7})$	$4.2e^{-7}$ $(0.7e^{-7})$	$4.3e^{-7}$ $(0.7e^{-7})$
$\sigma^2_{0(5y)}$	$6.1e^{-6}$ $(0.1e^{-6})$	$4.1e^{-6}$ $(0.1e^{-6})$	$4.1e^{-6}$ $(0.1e^{-6})$	$4.1e^{-6}$ $(0.1e^{-6})$	$4.2e^{-6}$ $(0.1e^{-6})$
$\sigma_{0(6m,5y)}$	$8.5e^{-7}$ $(1.1e^{-8})$	$9.4e^{-7}$ $(7.3e^{-7})$	$9.3e^{-7}$ $(7.5e^{-7})$	$9.4e^{-7}$ $(7.5e^{-7})$	$9.4e^{-7}$ $(7.1e^{-7})$
$\sigma^2_{1(6m)}$	–	$3.0e^{-7}$ $(0.3e^{-7})$	$3.0e^{-7}$ $(0.3e^{-7})$	$3.0e^{-7}$ $(0.3e^{-7})$	$3.0e^{-7}$ $(0.5e^{-7})$
$\sigma^2_{1(5y)}$	–	$2.0e^{-6}$ $(0.9e^{-6})$	$1.9e^{-6}$ $(0.8e^{-6})$	$2.0e^{-6}$ $(0.9e^{-6})$	$1.9e^{-6}$ $(0.7e^{-6})$
$\sigma_{1(6m,5y)}$	–	$3.9e^{-7}$ $(6.4e^{-7})$	$3.9e^{-7}$ $(6.1e^{-7})$	$3.9e^{-7}$ $(6.4e^{-7})$	$3.9e^{-7}$ $(4.3e^{-7})$
Log L	1787.30	1873.23	1875.00	1873.25	1876.48
AIC	-3560.61	-3718.46	-3720.01	-3716.55	**-3720.97**
SIC	-3540.01	**-3677.28**	-3675.88	-3672.43	-3673.90
HQ	-3552.24	-3701.73	**-3702.07**	-3698.62	-3701.84

Notes: The table reports the Markov-switching maximum likelihood estimates for the **estimated market price of factor risk models** (Models b, Table 1). The estimates of the parameter λ are obtained under the assumption that the 6 months and 5 years bonds are observed with error. This equation is jointly estimated along with the short term interest rate equation. The figures in parentheses are asymptotic standard errors.

Table 3: Recursive AIC, SIC, HQ

	AIC	SIC	HQ
Model 1			
Model 2		1991:3-1993:1	1997.4-1998.2
		1993:4-1994:4	
		1995:3-1998:4	
Model 3		1993:2-1993:3	1991:3-1992:1
		1995:1-1995:2	1997:2
			1998:3-1998:4
Model 4			1997:1-1998:4
Model 5	1991:3-1998:4		1992:2-1997:1
			1997:3

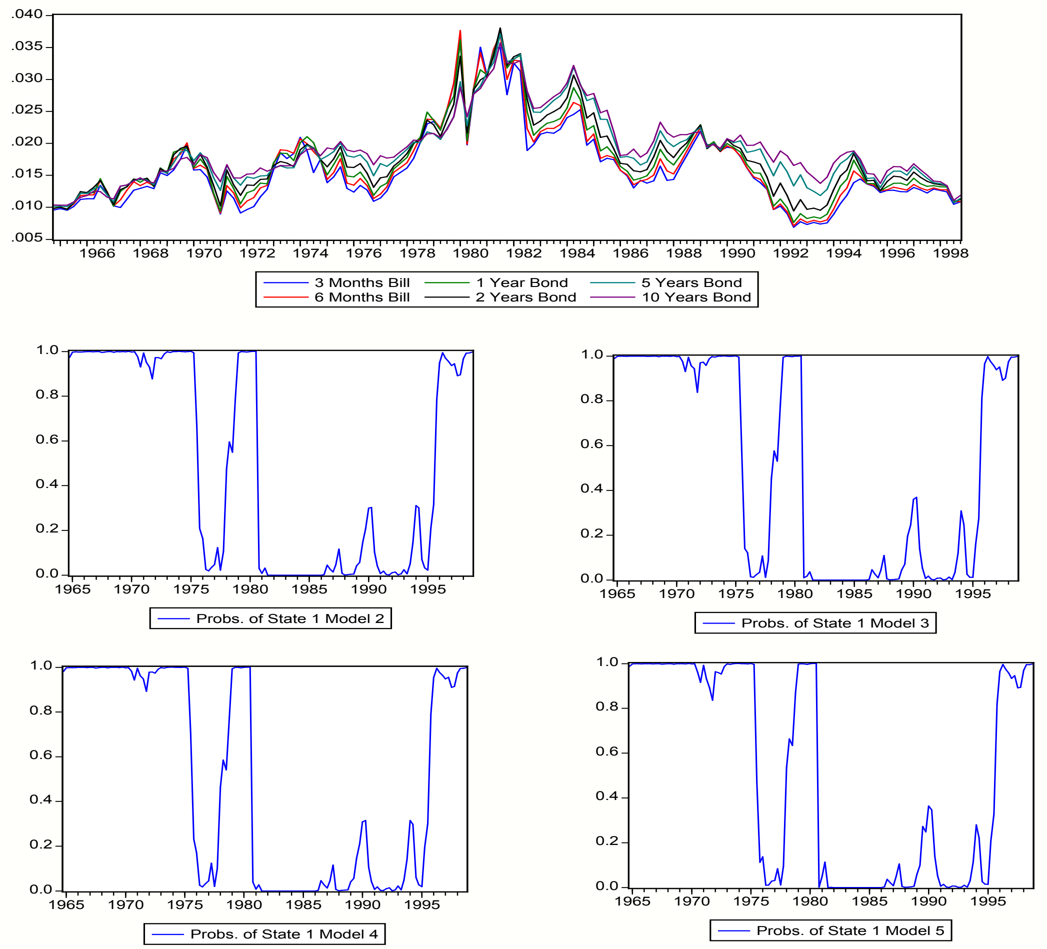
Note: The reported results are obtained in the following way:
(i) recursively estimate each of the Models described in Table 1
(starting from 1964:1-1991:2 and sequentially enlarging the sample up to 1998:4);
(ii) calculate, for each sample, the different complexity-penalized likelihood measures;
(iii) indicate periods during which each model is selected.

Table 4: Cumulative Recursive AIC, SIC and HQ

	AIC	SIC	HQ
Model 1	-89733.580	-89139.298	-89492.195
Model 2	-98907.021	**-97718.456**	-98424.251
Model 3	-98969.961	-97696.499	-98452.707
Model 4	-98851.139	-97577.676	-98333.885
Model 5	**-99009.363**	-97651.003	**-98457.626**

Note: Results are complexity-penalized likelihood cumulative measures.
These are obtained in the following way:
(i) recursively estimate each of the Models described in Table 1
(starting from 1964:1-1991:2 and sequentially enlarging the sample up to 1998:4);
(ii) calculate, for each sample, the different complexity-penalized likelihood measures;
(iii) compute the cumulative complexity-penalized likelihood measures for the enlarged sample.

Figure 1: Interest Rates and Probabilities Generated By the Models



4 Using Bond Pricing as Selection Criteria for the Interest Rate

In this section we assess the relative pricing performance of the different models under consideration to judge whether the standard assumption in the literature of allowing all the parameters of the model to switch is of economic importance (that is it affects the ex-ante pricing performance of the model). Given that pricing is intrinsically a forecasting exercise, we propose to use bond pricing as a selection criterion for the instantaneous interest rate. In other words, we use the information contained in the term structure to decide which parameterization of the short term interest rate produces the best bond prices. The main differences with the approach followed by papers such as that of Bansal and Zhou (2002) are that: *i*) the pricing is carried out recursively and ex-ante (see explanation below); *ii*) instead of pricing exactly the maturities that are not assumed to be priced with error, we leave them out of the estimation procedure and use them to assess the relative pricing performance of the different models.

To clarify the importance of this distinction, note that it is common practice to evaluate the pricing models using parameters which are obtained for the full sample and this is commonly done under the assumption that the public knows the true parameters of the model. Alternatively, in this paper we consider a framework where at each point in time prices are computed with the best available estimates of the parameters of the model (a real-time pricing approach).

Our approach is based on recursively estimating the models, using the observations from 1964:1-1991:2 to start the pricing exercise and sequentially enlarging the sample up to 1998:4 (our evaluation will therefore be based on a total of 30 sample points).[15] In other words, a yield curve, $-\frac{1}{\tau-t}ln(P_t^\tau(r_t, s_t))$, can be constructed by recursively estimating jointly the pricing equation and the instantaneous interest rate, using information up to time $t = t_1, ...T-1, T$. This produces a series of $T - t_1$ long-term interest rates for each maturity and estimated model. We then compare the actual and generated yields (for the maturities left out of the estimating procedure). This exercise is carried out thinking of the situation where a practitioner wants to price a long term bond at time t and cannot use information on the price of those bonds which are not yet priced (i.e. we recursively estimate the models using the estimates of the parameters obtained at time $t-1$ to price the bond at time t). The pricing (and estimation) is carried out recursively and the one step ahead prices at time t are computed as $\frac{A_\tau^{t-1}(s_{t-1})}{2} + \frac{B_\tau^{t-1}(s_{t-1})}{2}x_t$, where A_τ^{t-1} and B_τ^{t-1} are obtained using the estimates obtained at t-1 (since they use information of the long term bonds). For the one step ahead pricing we use the short term interest rates at time t (which are observed and assumed exogenous) to price the bonds at time t. In this way, we do not use information about the contemporaneous long yields to price them. We therefore refer to our approach as real time recursive one step ahead pricing (see Appendix C).

4.1 Comparison Based on Bond Pricing

We evaluate the relative performance of the different models using traditional accuracy measures, such as the RMSE, and by assessing their ability to correctly identify turning points (i.e. whether the rates are rising or falling regardless of the accuracy with which the magnitude of the change is predicted) using the so-called confusion rate and the procedure proposed by Pesaran and Timmermann (1992).[16] 4[17]

[15] We estimate the models as described in *table 1* and obtain bond yields for 1, 2 and 10 year (the 6 months and 5 years are used to obtain an estimate of the market price of factor risks λ_i).

[16] This evaluation method is particularly useful in situations where directional predictions are the focus of the analysis, as is the case, for instance, when we are trying to forecast the future price movements of asset prices.

[17] Let Δx_t be the actual change of interest rate and $\Delta\widehat{x}_t$ the predicted one. The evaluation is based on the following two criteria:

1) Consider the following 2×2 contingency table

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

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

Summarizing, we attempt to use all the information contained in our generated prices to assess which of the models has best predictive power. For each model we report: *i)* the relative mean square error (RMSE) of the difference between the generated yields and the actual data for each maturity, *ii)* the sum of the RMSE for all the maturities (which captures both the time series and the cross section dimension), *iii)* the confusion rate, which is used to measure whether our models correctly predict whether rates are rising or falling (i.e. the percentage of times the direction of the change in the yields is not correctly predicted), *iv)* the Pesaran and Timmermann (1992) statistic to formally test the success ratio, and *v)* the number of times each model outperforms the others on the basis of the RMSE.

Figure 2 shows the actual and the generated prices for the different maturities. We find that all the pricing models perform better in predicting the shorter maturities than in predicting the longer maturities.[18]

Table 5: Pricing Performance Results: Relative Mean Square Error, Confusion Rates and Pesaran-Timmermann tests

	Maturity					Total
	6 Month	1 Year	2 Year	5 Year	10 Year	
	Relative Mean Square Error					
Model 1	–	0.007756	0.010131	–	0.032819	0.050706
Model 2	–	**0.004287**	**0.009334**	–	0.022068	**0.035689**
Model 3	–	0.005558	0.011438	–	0.021748	0.038743
Model 4	–	0.004379	0.009518	–	0.022269	0.036167
Model 5	–	0.005413	0.011164	–	**0.021480**	0.038056
	Confusion Rates					
Model 1	–	0.10 (0.00)	0.10 (0.00)	–	0.13 (0.00)	**0.11**
Model 2	–	0.10 (0.00)	0.10 (0.00)	–	0.27 (0.00)	0.16
Model 3	–	0.10 (0.00)	0.17 (0.00)	–	0.30 (0.00)	0.19
Model 4	–	0.10 (0.00)	0.10 (0.00)	–	0.30 (0.00)	0.17
Model 5	–	0.10 (0.00)	0.13 (0.00)	–	0.30 (0.00)	0.18

Note: The values are Relative Mean Square Errors of the difference between the generated yields and the actual data for each maturity. In brackets are the (asymptotic) P-value of the success-ratio test statistic.

2) Consider the quantities:

$$l_t = \begin{cases} 1 & \text{if } \Delta x_t \leqslant 0, \\ -1 & \text{otherwise,} \end{cases} \qquad \text{and} \quad \widehat{l}_t = \begin{cases} 1 & \text{if } \Delta \widehat{x}_t \leqslant 0, \\ -1 & \text{otherwise,} \end{cases}$$

for $t = 1, \ldots, T$ (with T equals to the number of predicted yields). Then, the success ratio, i.e. the fraction of times the direction of changes of the yields are correctly predicted, is given by $S = (1/T)\sum_{t=1}^{T} \mathbb{I}(l_t \widehat{l}_t > 0)$, where $\mathbb{I}(\cdot)$ is the indicator function. If the model had no power in predicting the changes, $\{l_t\}$ and $\{\widehat{l}_t\}$ would be independent and the success ratio would be given by $S^* = q\widehat{q} + (1-q)(1-\widehat{q})$, where $q = (1/T)\sum_{t=1}^{T} \mathbb{I}(l_t = 1)$ and $\widehat{q} = (1/T)\sum_{t=1}^{T} \mathbb{I}(\widehat{l}_t = 1)$. Hence, we test whether the difference between S and S^* is statistically significant by using the statistic $\mathrm{SR} = (S - S^*)/(\sqrt{(1/T)\omega})$, where $\omega = S^*\{1 - S^*\} - q(1-q)(2\widehat{q} - 1)^2 - \widehat{q}(1-\widehat{q})(2q-1)^2$. Under the null hypothesis that actual and predicted changes are independent, SR has a standard normal asymptotic distribution.

[18] To evaluate how close are the generated prices to the actual ones, we exclude the two maturities (6 months and 5 years) which are used in the estimation process.

Figure 2: Pricing Performance of the Different Models Over Time

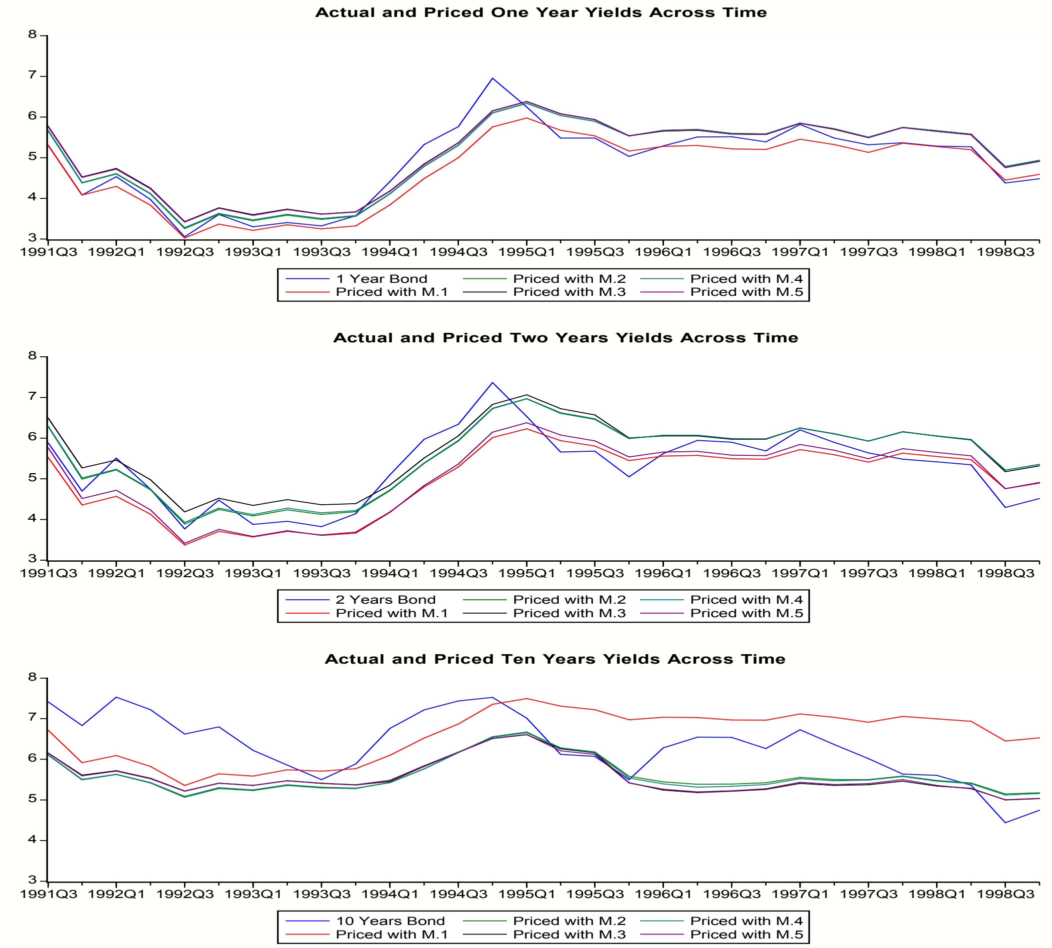


Table 5 reports RMSE values, confusion rates and the Pesaran and Timmermann test results for the 1 year, 2 year and 10 year maturities. We find that *model 2* outperforms the competing models in terms of producing prices closer to the actual data for the whole term structure. On the basis of the individual maturities, however, *model 5* achieves the smallest RMSE for the 10 year, while for the remaining maturities *model 2* significantly outperforms the competing models. Interestingly, a comparison between the models on the basis of confusion rates, shows that a 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. In fact *model 1* wrongly predicts the direction of the change 11% of time while *model 5* does it 18% of the time. This might imply that any improvement of *model 5* over *model 1* in terms of fit might be undone by its poor predictive performance, in terms of one-step ahead pricing. Finally, p-values of the Pesaran and Timmermann (1992) tests show that the null hypothesis that actual and predicted changes are independent is strongly rejected by the data for all maturities and models.

Table 6 reports the proportion of the times that each model achieves the smallest RMSE over the 30 sample points (1991:3-1998:4). This is calculated on the basis of the individual maturities (1, 2 and 10 years) and the sum of them. We find that *model 1* outperforms the alternative switching specifications 60% of the time, while *model 5* only outperforms the competing models 5% of the time.

It is very informative to compare the results presented in *table 6* with those presented in *table 5* (where we look at the average pricing errors). We find for the 10 year rate that the smallest RMSE is achieved by *model 5*, while when we evaluate the performance in terms of the number of periods with the smallest pricing errors, we find that *model 5* only outperforms the competing models 10% of the time and that *model 1* achieves the smallest RMSE 63% of the time.

Table 6: Pricing Performance Results: Percentage of the periods where each model outperforms the rival models using the Relative Mean Square (Pricing) Error

	Maturity					
	6 Month	1 Year	2 Year	5 Year	10 Year	Total
	Relative Mean Square Error over Time					
Model 1	–	0.57	0.50	–	0.63	0.57
Model 2	–	0.10	0.17	–	0.13	0.13
Model 3	–	0.27	0.27	–	0.07	0.20
Model 4	–	0.03	0.03	–	0.07	0.05
Model 5	–	0.03	0.03	–	0.10	0.05

Note: The reported dates are for the models with lowest Relative Mean Square Errors of the difference between the generated yields and the actual data. The entries are the percentage of time each models achieve the smallest RMSE over the sample size (1991:3-1998:4).

To summarize: the linear model seems to be more successful for pricing bonds over time (i.e. it outperforms the switching models most of the time), while, switching models seem to be more successful on average because they outperform the linear models around the breaks in the data. In fact *figure 2* shows that (given that our sample includes two changes in regime: in 1991:3-1995:3 and 1995:4-1998:4) the switching models seem to be useful for pricing 10-year bonds immediately after the break, but their pricing performance deteriorates the further away from the break we evaluate the models.

The poor performance of *model 5* (over time), compared with *model 1,* highlights the fact that attempting to correctly specify the switching model is crucial (especially when the model is used to produce one step ahead prices). This exercise suggests how important it 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 that do not allow for regime switching, even in cases where there are clear breaks in the data.

5 Conclusions

This paper provides an analysis of several regime-switching characterizations of the Cox, Ingersoll and Ross 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. Our approach is based on recursively estimating Markov-switching models for the short-term interest rate and generating bond yields which are then compared with actual yields. We find that the results obtained for the whole sample do not coincide with those obtained using different pricing strategies. These results illustrate that, for the one-factor model analyzed in the paper, Markov-switching specifications provide the best in-sample fit but not necessarily the best ex-ante one step ahead prices.

The main results of the paper can be summarized as follows: *(i)* for short and medium term maturities, simpler Markov-switching specifications produce better bond prices than those obtained using models where all the parameters are allowed to switch (and models with no regime switching); *(ii)* the pricing gains of Markov-switching models diminish the further away of the break the bond price is evaluated, to the extent that eventually the no regime switching model beats the Markov-switching model. Nevertheless, some of these findings should not be very surprising, since a similar phenomenon is found in the literature on forecasting with Markov-switching models. These results highlight the importance of paying special attention to the parameterizations of Markov-switching models.

Appendix A. Model 1: The Benchmark CIR Model

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

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

and

$$x_{t+1} - x_t = \kappa[\theta - x_t] + \sigma\sqrt{x_t}u_{t+1}. \tag{A.2}$$

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

$$\log P_t^\tau = E_t[\log M_{t+1} + \log P_{t+1}^{\tau-1}] + \tfrac{1}{2}Var_t[\log M_{t+1} + \log P_{t+1}^{\tau-1}]. \tag{A.3}$$

Using the following affine functional form for bond prices

$$P_t^\tau = \exp\left[-A_\tau - B_\tau x_t\right], \tag{A.4}$$

with the boundary condition

$$P_\tau^0 = 1,$$

we obtain the expressions P_t^τ and hence $P_{t+1}^{\tau-1}$ required in (A.3). Substituting them into (A.3) and using the fact that $E_t\left[M_{t+1}\right] = \exp[-r_t^f] = \exp[-x_t]$ yield

$$A_\tau + B_\tau x_t = [A_{\tau-1} + B_{\tau-1}\kappa\theta] + [1 + B_{\tau-1}(1-\kappa)]x_t - \tfrac{1}{2}B_{\tau-1}^2\sigma^2 x_t - B_{\tau-1}\lambda x_t.$$

The right side is obtained as follows:

$$\begin{aligned}\log M_{t+1} + \log P_{t+1}^{\tau-1} &= -x_t - \left(\tfrac{\lambda}{\sigma}\right)^2 \tfrac{x_t}{2} - \left(\tfrac{\lambda}{\sigma}\right)\sqrt{x_t}u_{t+1} - A_{\tau-1} - B_{\tau-1}x_{t+1} \\ &= -[A_{\tau-1} + B_{\tau-1}\kappa\theta] - [1 + \left(\tfrac{\lambda}{\sigma}\right)^2 \tfrac{1}{2} + B_{\tau-1}(1-\kappa)]x_t \\ &\quad -[\left(\tfrac{\lambda}{\sigma}\right) + B_{\tau-1}\sigma]\sqrt{x_t}u_{t+1}\end{aligned} \tag{A.5}$$

which has the conditional moments

$$E_t[\log M_{t+1} + \log P_{t+1}^{\tau-1}] = -[A_{\tau-1} + B_{\tau-1}\kappa\theta] - [1 + \left(\frac{\lambda}{\sigma}\right)^2 \tfrac{1}{2} + B_{\tau-1}(1-\kappa)]x_t,$$

and

$$Var_t[\log M_{t+1} + \log P_{t+1}^{\tau-1}] = [\frac{\lambda}{\sigma} + B_{\tau-1}\sigma]^2 x_t.$$

Separating the coefficients on the constant and on the terms in x in (A.5) gives us a set of difference equations for A_τ and B_τ

$$\begin{aligned} B_\tau &= 1 + (1-\kappa-\lambda)B_{\tau-1} - \tfrac{1}{2}B_{\tau-1}^2\sigma^2, \\ A_\tau &= A_{\tau-1} + B_{\tau-1}\kappa\theta. \end{aligned} \tag{A.6}$$

The boundary condition $P_\tau^0 = 1$ implies that

$$A_0 = B_0 = 0.$$

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

$$y_t^\tau = -\frac{\log P_t^\tau}{\tau} = \frac{A_\tau}{\tau} + \frac{B_\tau}{\tau}x_t.$$

Appendix B. The CIR Model with Regime Switching

Assuming that within regime s_{t+1} the evolution of the short rate under physical (historical) measure $\mathbb{P}$ follows the process (5) in the text

$$x_{t+1} - x_t = \kappa(s_{t+1})[\theta(s_{t+1}) - x_t] + \sigma(s_{t+1})\sqrt{x_t}u_{t+1}, \tag{B.1}$$

and that, the pricing kernel allowing for changes in regime takes the form

$$M_{t+1}(s_{t+1}) = \exp\left[-r_t^f - \left(\frac{\lambda(s_{t+1})}{\sigma(s_{t+1})}\right)^2 \frac{x_t}{2} - \frac{\lambda(s_{t+1})}{\sigma(s_{t+1})}\sqrt{x_t}u_{t+1}\right], \tag{B.2}$$

then, (zero-coupon) bond prices in regime $s_t = i$ are given by

$$P_t^\tau(s_t = i) = \exp\left[-A_\tau(i) - B_\tau(i)x_t\right], \tag{B.3}$$

where

$$\begin{aligned} A_\tau(i) &= \pi_{ii}(A_{\tau-1}(i) + B_{\tau-1}(i)\kappa_i\theta_i) + \pi_{ij}(A_{\tau-1}(j) + B_{\tau-1}(j)\kappa_j\theta_j) \qquad i \neq j, \\ B_\tau(i) &= \pi_{ii}\left((1-\kappa_i-\lambda_i)B_{\tau-1}(i) - \frac{B_{\tau-1}^2(i)}{2}\sigma_i^2 + 1\right) \\ &\quad + \pi_{ij}\left((1-\kappa_j-\lambda_j)B_{\tau-1}(j) - \frac{B_{\tau-1}^2(j)}{2}\sigma_j^2 + 1\right), \end{aligned}$$

with initial conditions $A_0(i) = 0$ and $B_0(i) = 0$.

Proof: Notice that when the underlying process is subject to regime shifts, the fundamental bond pricing equation (9) becomes

$$\begin{aligned} P_t^\tau(s_t = i) &= \sum_{j=0,1} \pi_{ij} E_t \left[M_{t+1}(s_{t+1}) P_{t+1}^{\tau-1}(s_{t+1}) | s_{t+1} = j \right] \\ &= \pi_{i0} E_t \left[M_{t+1}(j) P_{t+1}^{\tau-1}(j) | j = 0 \right] + \pi_{i1} E_t \left[M_{t+1}(j) P_{t+1}^{\tau-1}(j) | j = 1 \right]. \end{aligned} \tag{B.4}$$

Then we can calculate the following relationships:

i) Conditional on $s_t = 0$ we can write

$$P_t^\tau(s_t = 0) = \pi_{00} E_t \left[M_{t+1}(0) P_{t+1}^{\tau-1}(0) \right] + \pi_{01} E_t \left[M_{t+1}(1) P_{t+1}^{\tau-1}(1) \right].$$

ii) Conditional on $s_t = 1$ we can write

$$P_t^\tau(s_t = 1) = \pi_{10} E_t \left[M_{t+1}(0) P_{t+1}^{\tau-1}(0) \right] + \pi_{11} E_t \left[M_{t+1}(1) P_{t+1}^{\tau-1}(1) \right]. \tag{B.5}$$

Notice that under the informational assumptions of Bansal and Zhou (2002)

$$E_t \left[M_{t+1}(0) P_{t+1}^{\tau-1}(0) \right] = E_t \left(\exp \left[-r_t^f - \left(\frac{\lambda_0}{\sigma_0} \right)^2 \frac{x_t}{2} - \frac{\lambda_0}{\sigma_0} \sqrt{x_t} u_{t+1} - A_{\tau-1}(0) - B_{\tau-1}(0) x_{t+1} \right] \right)$$

and

$$E_t \left[M_{t+1}(1) P_{t+1}^{\tau-1}(1) \right] = E_t \left(\exp \left[-r_t^f - \left(\frac{\lambda_1}{\sigma_1} \right)^2 \frac{x_t}{2} - \frac{\lambda_1}{\sigma_1} \sqrt{x_t} u_{t+1} - A_{\tau-1}(1) - B_{\tau-1}(1) x_{t+1} \right] \right),$$

which allows us to express the pricing equation (B.4) as

$$\begin{gathered} \exp\left[-A_\tau(0) - B_\tau(0) x_t \right] = \\ \pi_{00} E_t \left(\exp \left[\begin{array}{c} -r_t^f - \left(\frac{\lambda_0}{\sigma_0} \right)^2 \frac{x_t}{2} - \frac{\lambda_0}{\sigma_0} \sqrt{x_t} u_{t+1} \\ -A_{\tau-1}(0) - B_{\tau-1}(0) x_{t+1} \end{array} \right] \right) \\ + \pi_{01} E_t \left(\exp \left[\begin{array}{c} -r_t^f - \left(\frac{\lambda_1}{\sigma_1} \right)^2 \frac{x_t}{2} - \frac{\lambda_1}{\sigma_1} \sqrt{x_t} u_{t+1} \\ -A_{\tau-1}(1) - B_{\tau-1}(1) x_{t+1} \end{array} \right] \right), \end{gathered} \tag{B.6}$$

and

$$\begin{gathered} \exp\left[-A_\tau(1) - B_\tau(1) x_t \right] = \\ \pi_{10} E_t \left(\exp \left[\begin{array}{c} -r_t^f - \left(\frac{\lambda_0}{\sigma_0} \right)^2 \frac{x_t}{2} - \frac{\lambda_0}{\sigma_0} \sqrt{x_t} u_{t+1} \\ -A_{\tau-1}(0) - B_{\tau-1}(0) x_{t+1} \end{array} \right] \right) \\ + \pi_{11} E_t \left(\exp \left[\begin{array}{c} -r_t^f - \left(\frac{\lambda_1}{\sigma_1} \right)^2 \frac{x_t}{2} - \frac{\lambda_1}{\sigma_1} \sqrt{x_t} u_{t+1} \\ -A_{\tau-1}(1) - B_{\tau-1}(1) x_{t+1} \end{array} \right] \right). \end{gathered} \tag{B.7}$$

To arrive to the final result we notice that

$$E_t\left(\exp\left[-r_t^f-\left(\frac{\lambda_0}{\sigma_0}\right)^2\frac{x_t}{2}-\frac{\lambda_0}{\sigma_0}\sqrt{x_t}u_{t+1}-A_{\tau-1}(0)-B_{\tau-1}(0)x_{t+1}\right]\right)=$$
$$\exp(-r_t^f-\left(\frac{\lambda_0}{\sigma_0}\right)^2\frac{x_t}{2}-A_{\tau-1}(0))E_t\left(\exp\left[-\frac{\lambda_0}{\sigma_0}\sqrt{x_t}u_{t+1}-B_{\tau-1}(0)x_{t+1}\right]\right)$$

and that

$$E_t\left(\exp\left[-\frac{\lambda_0}{\sigma_0}\sqrt{x_t}u_{t+1}-B_{\tau-1}(0)x_{t+1}\right]\right)$$
$$=\exp\left(\left(\frac{\lambda_0}{\sigma_0}\right)^2\frac{x_t}{2}-B_{\tau-1}(0)(x_t+\kappa_0[\theta_0-x_t])+\frac{B_{\tau-1}^2(0)}{2}\sigma_0^2x_t+B_{\tau-1}(0)\lambda_0x_t\right)$$

This last result holds since

$$E_t\exp\left[-\frac{\lambda_0}{\sigma_0}\sqrt{x_t}u_{t+1}\right]=\exp(\left(\frac{\lambda_0}{\sigma_0}\right)^2\frac{x_t}{2})\text{ where }u_{t+1}\tilde{}N(0,1),$$

$$E_t\left(\exp\left[-B_{\tau-1}(0)x_{t+1}\right]\right)=\exp\left(-B_{\tau-1}(0)(x_t+\kappa_0[\theta_0-x_t])+\frac{B_{\tau-1}^2(0)}{2}\sigma_0^2x_t\right),$$

and that the cross term that enters in the variance is $B_{\tau-1}(0)\lambda_0x_t$.

Putting all this results together we obtain that

$$E_t\left(\exp\left[-r_t^f-\frac{{\lambda_0}^2}{2}x_t-\lambda_0\sqrt{x_t}u_{t+1}-A_{\tau-1}(0)-B_{\tau-1}(0)x_{t+1}\right]\right)=$$
$$\exp(-r_t^f-A_{\tau-1}(0)-B_{\tau-1}(0)(x_t+\kappa_0[\theta_0-x_t])+\frac{B_{\tau-1}^2(0)}{2}\sigma_0^2x_t+B_{\tau-1}(0)\lambda_0x_t)$$

Using the log-linear approximation $\exp^x\approx 1+x$ as in Bansal and Zhou (2002) and the fact that $x_t=r_t^f$, we get the following pricing relationships:

i) Conditional on the current regime $s_t=0$,

$$[-A_\tau(0)-B_\tau(0)x_t]=$$
$$\pi_{00}(-x_t-A_{\tau-1}(0)-B_{\tau-1}(0)(x_t+\kappa_0[\theta_0-x_t])+\frac{B_{\tau-1}^2(0)}{2}\sigma_0^2x_t+B_{\tau-1}(0)\lambda_0x_t)$$
$$+\pi_{01}(-x_t-A_{\tau-1}(1)-B_{\tau-1}(1)(x_t+\kappa_1[\theta_1-x_t])+\frac{B_{\tau-1}^2(1)}{2}\sigma_1^2x_t+B_{\tau-1}(1)\lambda_1x_t).\quad\text{(B.8)}$$

ii) Conditional on the current regime $s_t = 1$,

$$[-A_\tau(1) - B_\tau(1)x_t] =$$
$$\pi_{10}(-x_t - A_{\tau-1}(0) - B_{\tau-1}(0)(x_t + \kappa_0[\theta_0 - x_t]) + \frac{B^2_{\tau-1}(0)}{2}\sigma_0^2 x_t + B_{\tau-1}(0)\lambda_0 x_t)$$
$$+\pi_{11}(-x_t - A_{\tau-1}(1) - B_{\tau-1}(1)(x_t + \kappa_1[\theta_1 - x_t]) + \frac{B^2_{\tau-1}(1)}{2}\sigma_1^2 x_t + B_{\tau-1}(1)\lambda_1 x_t). \tag{B.9}$$

Finally equating the constant terms and the terms in x_t we obtain that

$$\begin{aligned}
A_\tau(0) &= \pi_{00}(A_{\tau-1}(0) + B_{\tau-1}(0)\kappa_0\theta_0) + \pi_{01}(A_{\tau-1}(1) + B_{\tau-1}(1)\kappa_1\theta_1) \\
B_\tau(0) &= \pi_{00}\left((1 - \kappa_0 - \lambda_0)B_{\tau-1}(0) - \frac{B^2_{\tau-1}(0)}{2}\sigma_0^2 + 1\right) \\
&\quad +\pi_{01}\left((1 - \kappa_1 - \lambda_1)B_{\tau-1}(1) - \frac{B^2_{\tau-1}(1)}{2}\sigma_1^2 + 1\right)
\end{aligned}$$

$$\begin{aligned}
A_\tau(1) &= \pi_{10}(A_{\tau-1}(0) + B_{\tau-1}(0)\kappa_0\theta_0 + \pi_{11}(A_{\tau-1}(1) + B_{\tau-1}(1)\kappa_1\theta_1) \\
B_\tau(1) &= \pi_{10}\left((1 - \kappa_0 - \lambda_0)B_{\tau-1}(0) - \frac{B^2_{\tau-1}(0)}{2}\sigma_0^2 + 1\right) \\
&\quad +\pi_{11}\left((1 - \kappa_1 - \lambda_1)B_{\tau-1}(1) - \frac{B^2_{\tau-1}(1)}{2}\sigma_1^2 + 1\right)
\end{aligned}$$

Appendix C. Maximum Likelihood Estimation

The estimates of the regime-swtching models are obtained using procedures which are identical to those described by (Hamilton, 1988, 1989), except that in this case the short term interest rate and the yields of the two maturities observed with error (the six months bill and the 5 years bond) depend on the state of the economy. The density of the data y_t conditional on the state s_t and the history of the system can be written as

$$\begin{aligned}
P(y_t|s_t, y_{t-1}, ..., y_1 \quad &: \quad \Theta) = \frac{1}{(2\pi)^{.5}\sigma_{s_t}\sqrt{x_{t-1}}}\exp(-(\sigma^2_{s_t}x_{t-1})^{-1}(\Delta x_t - \kappa(s_t)[\theta(s_t) - x_{t-1}])^2) \\
&\times\frac{1}{2\pi|\Sigma_{s_t}|^{1/2}}\exp(-\frac{1}{2}u_t'\Sigma^{-1}_{s_t}u_t),
\end{aligned}$$

where y_t is a 3×1 vector containing the 3months T-bill, x_t, the 6 month T-bill, y_t^2, and the 5 years bond, y_t^{20}, where
$u_t = \begin{bmatrix} y_t^2 - \frac{A_2(s_t)}{2} + \frac{B_2(s_t)}{2}x_t \\ y_t^{20} - \frac{A_{20}(s_t)}{2} + \frac{B_{20}(s_t)}{2}x_t \end{bmatrix}$, $\Sigma_0 = \begin{bmatrix} \sigma^2_{0(6m)} & \sigma_{0(6m,5y)} \\ \sigma_{0(6m,5y)} & \sigma^2_{0(5y)} \end{bmatrix}$,
$\Sigma_1 = \begin{bmatrix} \sigma^2_{1(6m)} & \sigma_{1(6m,5y)} \\ \sigma_{1(6m,5y)} & \sigma^2_{1(5y)} \end{bmatrix}$, and $A_\tau(s_t)$ and $B_\tau(s_t)$ are generated as in Appendix B, where $\Theta = \{\kappa_0, \kappa_1, \theta_0, \theta_1, \sigma_0, \sigma_1, \lambda_0, \lambda_1, \Sigma_0, \Sigma_1\}$.

The pricing (and estimation) is carried out recursively and the one step ahead prices are computed as $\frac{A_\tau^{t-1}(s_{t-1})}{2} + \frac{B_\tau^{t-1}(s_{t-1})}{2} x_t$, where A_τ^{t-1} and B_τ^{t-1} are obtained using the estimates obtained at $t-1$ (since they use information of the long term bonds). For the one step ahead pricing we use the short term interest rates at time t (which are observed and assumed exogenous) to price the bonds at time t. In this way we do not use information about the contemporaneous long yields to price them.

References

Ang, A. and G. Bekaert (2002a): "Regime switches in interest rates," *Journal of Business and Economic Statistics*, 20, 163–182.

Ang, A. and G. Bekaert (2002b): "Short rate nonlinearities and regime switches," *Journal of Economic Dynamics and Control*, 26, 1243–1274.

Bansal, R. and H. Zhou (2002): "Term structure of interest rates with regime shifts," *Journal of Finance*, 57, 1997–2043.

Bossaerts, P. and P. Hillion (1999): "Implementing statistical criteria to select return forecasting models: What do we learn?" *Review of Financial Studies*, 12, 405428.

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.

Dai, Q. and K. Singleton (2003): "Term structure dynamics in theory and reality," *Review of Financial Studies*, 16, 631–678.

Diebold, F. and C. Li (2006): "Forecasting the term structure of government bond yields," *Journal of Econometrics*, 130, 337–364.

Duffee, G. R. (2002): "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, 405–443.

Evans, M. (2003): "Regime shifts, risk and the term structure," *Economic Journal*, 113, 345–389.

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," *Review of Economics and Statistics*, 78, 111–125.

Granger, C., M. King, and H. White (1995): "Comments on testing economic theories and the use of model selection criteria," *Journal of Econometrics*, 67, 172187.

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.

Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–384.

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.

Hansen, B. E. (1996): "Erratum: The likelihood ratio test under non-standard conditions: Testing the markov switching model of gnp," *Journal of Applied Econometrics*, 11, 195–198.

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. (1992): "Maximum-likelihood estimation for hidden markov-models," *Stochastic Processes and their Applications*, 40, 127143.

Leroux, B. and M. Puterman (1992): "Maximum penalized likelihood estimation for independent and markov-dependent poisson mixtures," *Biometrics*, 48, 545558.

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.

Pearson, N. and T. Sun (1994): "Exploiting the conditional density in estimating the term structure: An application to the cox, ingersoll, and ross model," *Journal of Finance*, 49, 1279–1304.

Pesaran, M. H. and A. Timmermann (1992): "A simple nonparametric test of predictive performance," *Journal of Business and Economic Statistics*, 10, 461–465.

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.

Psaradakis, Z. and N. Spagnolo (2006): "Joint determination of the state dimension and autoregressive order for models with markov regime switching," *Journal of Time Series Analysis*, 27, 753–756.

Ryden, T. (1997): "On recursive estimation for hidden markov models," *Stochastic Processes and their Applications*, 66, 79–96.

Sin, C.-Y. and H. White (1996): “Information criteria for selecting possibly misspecified parametric models,” *Journal of Econometrics*, 71, 207225.

Smith, D. R. (2002): “Markov-switching and stochastic volatility in short-term interest rates,” *Journal of Business and Economic Statistics*, 20, 183–197.

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.

Swanson, N. R. (1998): “Money and output through a rolling window,” *Journal of Monetary Economics*, 41, 455–473.