Consistent fitting of one-factor models to interest rate data

L.C.G. Rogers [1] * and Wolfgang Stummer [2] *

[1] School of Mathematical Sciences
University of Bath
Bath BA2 7AY
UK
and
[2] Department of Finance
University of Ulm
Helmholtzstrasse 18
D-89081 Ulm, Germany
email: ws@mathematik.uni-ulm.de
Tel. +49 731 50-23599, Fax +49 731 50-23950

January 12, 2016

We carry out a full maximum-likelihood fit of the popular single-factor Vasicek and Cox-Ingersoll-Ross models to term-structure data from the US and UK. This contrasts with the usual practice of performing a day-by-day fit. The day-by-day fitting gives average errors in yields of the order of 2–3 basis points, whereas with the full maximum-likelihood fit we found an average error in yield of 3–8 basis points.

Keywords: term structure of interest rates, Cox-Ingersoll-Ross model, Vasicek model

JEL classification: E 43, C 13

*Research supported by SERC grants numbers GR/H52023 and GR/K00448

0. Introduction

It is well known that the topics of interest rate and term structure modeling are very important for actuarial issues, especially for managing ”interest rate risk” and for valuing insurance assets and liabilities.

Some recent general discussions along this line are given e.g. in Tilley (1988), Santomereo & Babbel (1997), Reitano (1997) and Girard (2000). In Ang & Sherris (1997) one can find a nice overview on several different actuarial problems which involve term structure models – annuity products, insurance products with interest-rate-sensitive lapses, property and casualty insurance company outstanding claims reserves, accrued liabilities of defined-benefit pension funds.

Let us also mention some articles which deal with more specific (rather than general) situations of interest to actuaries: The actuarial immunization problem in face of a (non-flat) term structure of interest rates was investigated by Boyle (1978, 1980), Albrecht (1985) and Beekman & Shiu (1988); the latter derive a duration formula for the poular Cox, Ingersoll & Ross (1985) model (hereafter, CIR). In the paper of De Schepper, Goovaerts & Delbaen (1992) some interest rate models for annuities are introduced. A method for the estimation of the yield curve by means of interpolation can be found in Delbaen & Lorimier (1992); see also the follow-up paper of Corradi (1996). Pedersen & Shiu (1994) explain how to use term structure models in order to evaluate rollover options for a Guaranteed Investment Contract (GIC), which is a popular product sold by insurance companies in North America. Sherris (1994) uses a certain interest rate model in order to evaluate loan contracts which are similar to insurance policies with a surrender option. Bühlmann (1995) explains how to use stochastic interest rates in life insurance. The paper of Boyle (1995) analyzes risk-based capital for financial institutions with the help of the Vasicek model (1977) of stochastic interest rates. In the light of promises of guaranteed interest rates on long-term (life) insurance products, Deelstra & Delbaen (1995a, 1995b) use an extension of the CIR model and study the long-term return; for a discrete-time approximation of their model, see Deelstra & Delbaen (1998). A discussion on yield rates for pure discount bonds with maturities of up to 100 years and their impact for evaluating life annuities can be found in Carriere (1999). Another recent actuarial paper which deals with long-term returns is Yao (1999). The use of the Duffie & Kan (1996) term structure model for maximizing expected lifetime utility (subject to budget constraints) is explained in Boyle & Yang (1997).

It is beyond the scope of this paper to give a complete overview of the vast literature on interest rate models in actuarial sciences. Some nice surveys can be found in the papers of Vetzal (1994), Ang & Sherris (1997) and the monographs of Panjer et al. (1998) and Rolski, Schmidli, Schmidt & Teugels (1999). For further references in a general financial context, see below.

Thus an important issue in insurance and finance is to find adequate models for the movement of the interest-rate term structure. While the irresistible rise of arbitrage pricing theory has relegated the earliest primitive models to their true historical position (see, for example, the article of Cox, Ingersoll & Ross (1981)), it has not yet resulted in any one obviously "right" model for the term structure, although there are several which, like the geometric Brownian motion model for share prices, are good enough to be getting on with. Following the influential papers of Vasicek (1977) and of Cox, Ingersoll & Ross (1985), many have offered extensions and generalisations of the models presented there; Hull & White (1990) and Jamshidian (1995) consider time-dependent versions, and among the papers which introduce multifactor extensions of the models of Vasicek and Cox, Ingersoll & Ross we mention Beaglehole & Tenney (1991), El Karoui, Myneni & Viswanathan (1992), Duffie & Kan (1996), Constantinides (1992), Richard (1978), Longstaff & Schwartz (1992), Jamshidian (1996), as well as Cox, Ingersoll & Ross (1985) themselves. For recent surveys of some of the various models, see e.g. the papers of Rogers (1995a) and Back (1996) as well as the monographs of Anderson et al. (1996), Baxter & Rennie (1996), Rebonato (1998), Duffie (1996), Lamberton & Lapeyre (1996), Musiela & Rutkowski (1997), Bingham & Kiesel (1998), Kwok (1998).

These extensions of the simple models naturally result in a wider class (and therefore a better fit to data), but suffer the attendant disadvantages; computations on such models are more complicated, and one encounters problems of overparametrisation. The principal aim of this paper is to explore to what extent the simple, classical one-factor models of Vasicek and Cox, Ingersoll & Ross can fit bond prices. This will be done in consistent econometric way, which we call *full* maximum likelihood method (FML). This will be compared with more crude data analysis methods. By analysing a number of datasets from the British and US markets (which differ from those used in the literature described in Section 1 below), we conclude that the fits are remarkably good.

The plan of the rest of the paper is as follows. A survey of previous related empirical work as well as an outline of the proposed new econometric method will be given in Section 1. In Section 2, we describe the data sets used, and the analyses performed on them. Section 3 reports the results of the fits, and Section 4 contains our conclusions.

1. An overview of previous empirical work

The econometric problems of fitting the classical models to data are considerable. If one tries to fit the basic models of Vasicek or CIR to cross-sectional data on N days by a *full* maximum likelihood fit, we end up with a maximisation over the 3 parameters (which are kept the same for each observation day) and the N values of the short rate, which will typically be a large number of variables. Previous econometric analyses have avoided this problem in a variety of ways.

One way is simply to ignore the issue, and fit a different model to each day's cross-sectional data, and this is the approach taken by Brown & Dybvig (1986), Barone, Cuoco & Zautzik (1991), Brown & Schaefer (1994), De Munnik & Schotman (1994) and Moriconi (1995), for example. In fact, Brown & Dybvig use monthly data from the CRSP tape on US treasury security prices from 1952 to 1983. For each of the 373 days in the study, they separately computed the least-squares fit to the bond prices (rather than the log bond prices we have used). Their estimation did not constrain the parameters to be non-negative, so they occasionally obtained negative estimates for the variance. The implied long rates moved around substantially (whereas both the Vasicek and CIR models predict a constant long rate and in general – see Dybvig, Ingersoll & Ross (1996) – the long rate is a non-decreasing function of time). Brown & Schaefer (1994) conduct a detailed and informative analysis of weekly UK Government index-linked bonds from 1984 to 1989, which they attempt to fit with a one-factor CIR model separately to each day's data (again, they use the bond prices themselves). The frequency of their data is weekly. Among the other conclusions of Brown & Schaefer, the yield curve is very well fitted by the one-factor CIR model within each cross-section; the function to be minimised is very flat in certain directions in parameter space. They also find that the long end of the yield curve is much more stable than with nominal bonds. De Munnik & Schotman (1994) obtain similar conclusions as Brown & Schaefer, but for the Dutch Government bond market from 1989 to 1990 (with maturities of up to 10 years). Barone, Cuoco & Zautzik (1991) estimate the CIR model from daily data on Italian Treasury bonds (BTP's) from 1983-1990. They conclude that the CIR model appears to be a good explanation of the observed variability of the bond prices. Moriconi (1995) investigates the Italian Government securities market from 1990-1992 via the CIR model. He observes that the corresponding one-day cross-sectional estimation procedure produces a fairly good fit for two (namly BTP's and BTO's) out of the four main classes of Italian Treasury bonds.

With these single-day fitting methods, one should then at least consider the issue of parameter stability; for example, Brown & Schaefer find that the estimates of the parameters are very unstable from day to day. As we shall see, this instability is not surprising, but does not necessarily lead us to reject the model. In contrast, Barone, Cuoco & Zautzik (1991) conclude that their daily estimated parameters appear to be sufficiently stable over time.

Another way to avoid the large number of variables needed for a full maximum-likelihood fit is the approach of Gibbons & Ramaswamy (1993), who fit the model by the generalised method of moments (GMM). While this methodology is widely used in econometrics, it has, general speaking, some disadvantages. The first is that the choice of functions whose sample moments are to be estimated is essentially arbitrary, even though natural choices can commonly be found. Secondly, the distributional properties of the estimates are only known asymptotically. Thirdly, the validity of the GMM approach is based on an assumption of stationarity for the underlying short-rate process. This assumption needs to be checked in any particular application. Indeed, inspection of the graphs of real returns on T-bills given in this paper do not suggest that this is likely. Nevertheless, Gibbons & Ramaswamy conclude that the one-factor CIR model does fit their short-term US T-bill data (from 1964 to 1983 with maturities 1, 3, 6, 12 months) reasonably well.

A third way to avoid the large number of variables is used by Pearson & Sun (1994), who investigate monthly prices from 1971 to 1986 of T-Bills and notes with maturities from 13 weeks to 10 years, taken from the abovementioned CRSP tape. They assume that the 13-week and 26-week T-bills are observed without error, and fit a two-factor CIR model to the data. This assumption means that once the values of the parameters are chosen, the values of the two stochastic factors can be computed from the 13-week and 26-week yields, so that the minimisation takes place over just the parameters of the model. One disadvantage of this approach is that it will be impossible to guarantee that the stochastic factors implied by the 13-week and 26-week yields will remain non-negative (as they must be in this model), and indeed Pearson & Sun reject the model largely on these grounds.

Let us also mention two other econometric analyses here, although their aims are quite different from those of the present study. Stambaugh (1988) analyses excess returns in terms of predictors based on today's yield curve, and finds that two or three factors are needed for satisfactory prediction. Chan, Karolyi, Longstaff & Sanders (1992) compare a large family of parametric models including the Vasicek and CIR model, and find evidence to prefer a different dependence of volatility on the level of interest rates. In more detail, Chan et. al investigate diffusion models for the spot-rate process of the form

$$dr_t = \sigma r_t^\gamma dW_t + (\alpha + \beta r_t)dt \ .$$

They discretize this stochastic differential equation in order to obtain a time series for r_t, and use the Generalized Method of Moments to estimate the parameters directly from this time series (rather than from cross-sectional data). They find that a model with $\gamma = 1.50$ fits the time series better than Vasicek or CIR. There are two points with the analysis conducted in this paper which from a theoretical point of view could turn out to be a little problematic. The first is that the asymptotics of the GMM estimators are only justified on the assumption that the underlying process is stationary and ergodic, and this can only happen if $\{\gamma = 1/2,\ 2\alpha > \sigma^2,\ \beta < 0\}$, or $\{1/2 < \gamma < 1,\ \alpha > 0,\ \beta \leq 0\}$, or $\{\gamma = 1,\ \alpha > 0,\ 2\beta/\sigma^2 < 1\}$, or $\{\gamma > 1,\ \alpha = 0,\ \beta > 0\}$, or $\{\gamma > 1,\ \alpha > 0\}$. The best estimated models in Chan et al. turn out to be ergodic except for the CIR VR model (which corresponds to the case $\{\gamma = 1.5,\ \alpha = 0,\ \beta = 0\}$). The second point concerns the data, which is monthly data on one-month T-Bill yields from the CRSP tapes. Quite apart from the question of whether the one-month T-Bill is a good proxy for r, the gap between observations may be too large to be able to pick up the volatility adequately. For example, if the mean-reversion parameter β were moderately large, and if r_t were also quite large compared to α/β, then r_{t+1} would have been pulled back in towards α/β, and so $r_{t+1} - r_t$ would be quite large (and negative); if r_t were small, then r_{t+1} is not going to be much bigger than r_t, so $r_{t+1} - r_t$ would be fairly small. The net effect is that the large inter-observation gap together with mean reversion would produce an effect similar to a value of $\gamma > 1$, especially if one crudely discretises the abovementioned SDE, as Chan et al. do. Accordingly, it seems that the conclusions drawn by Chan et al. should be investigated with a further more penetrating analysis.

It is beyond the scope of this paper to give a complete overview of the literature which deals with the implementation of the Vasicek and the CIR models; for further works, see the papers and books mentioned in Section 1 and the references therein.

What we propose in this paper is to carry out a *full* maximum likelihood fit (FML) of the one-factor Vasicek and CIR models to each of four bundles of cross-sectional data sets. To our knowledge, this is the first time that such an analysis has been carried out; since the models of Vasicek and CIR are intrinsically time-homogeneous, this is the only wholly consistent way to fit. For comparison purposes, we shall also perform other fits, which are detailed below. We find that by restricting the parameters of the model to be the same every day (while allowing the short rate to change) the fit is (of course) worsened, but that the goodness of fit is comparable to that obtained by allowing different parameters to be fitted each day. More precisely, if we fit a different one-factor model to each day's data, we get an average error of 2–3 basis points for each fitted yield, but if we demand that the model does not change from day to day, the average error gets increased by a factor of 2–3. Since the yield curve may be estimated from the prices of coupon-bearing bonds quoted to the nearest 1/32 out of 100 GBP, the errors inherent in the data will be of the order of a few basis points. Thus any model which could fit the data to within 4 basis points most of the time is doing reasonably well. The simple-minded one-factor models can achieve such a good fit, and what is of interest is that restricting the parameters to be constant does not make the fit much worse.

2. Datasets and analyses

Let $(r_t)_{t\in \mathbb{R}^+}$ denote the *short rate process* (i.e., the process of the instantaneous rate of interest), and let $P(t,T)$ denote the price at time t of a (zero-coupon) bond which delivers 1 with certainty at later time T; arbitrage pricing theory tells us that

$$P(t,T) = E_t\left[\exp\left(-\int_t^T r_u du\right)\right],$$

where E_t denotes conditional expectation with respect to a risk-neutral measure, given all information available at time t. To fix our terminology, we shall define the *return* $R(t,T)$ at time t on the zero-coupon bond which matures at $T > t$ by

$$R(t,T) \equiv -\log P(t,T),$$

and the *yield* to be

$$Y(t,T) \equiv R(t,T)/(T-t).$$

The *yield curve* at time t is the function $\tau \mapsto Y(t,t+\tau)$ defined for all $\tau > 0$.

The datasets used (which we shall refer to by the letters introduced here) were the following:

(A) The New York money-market rates reported in the *Financial Times* for the period 3rd May 1993 to 26th May 1993 were recorded. The range of maturities used was 1 month, 2 months, 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 7 years, 10 years. [The data in the *Financial Times* was quoted in terms of the yields $Y(t_i, t_i+\tau_j)$, to the nearest basis point.]

(B) The New York money-market rates reported in the *Financial Times*, taking one day per month for the period January 1991 to February 1993. The day chosen was the second Tuesday of the month in order to avoid "weekend and beginning of the month" effects.

(C) The yields for maturities 1 month, 3 months, 6 months, 1 year, 2 years, 3 years, 4 years, 5 years, 7 years and 10 years on each trading day t_i from 24th December 1992 to 25th May 1993 were extracted from data on British interest rates kindly supplied to us by Simon Babbs. These yields were estimated from the prices of (coupon bearing) UK government bonds.

(D) This dataset was from the same source as (C), but relates to US interest rates. We extracted the same maturities for the same days.

All the datasets have the structure of a rectangular array $(R_i^j)_{i=1,...,N,\ j=1,...,M}$ of returns, where N denotes the number of days on which returns were recorded, and M denotes the number of maturities recorded each day. We write R_i for the vector $(R_i^j)_{j=1}^M$ which is usually referred to as the cross-section (of returns, in our context) observed on day t_i. Thus, we deal with N cross-sections. The analyses performed on each dataset were the following.

Analysis I Considering $R_1, ..., R_N$ as a sample from some multivariate normal distribution, the sample covariance matrix was computed as a first crude attempt to understand the data.

As part of this analysis, the structure of the eigenspaces was investigated in the following way: there is a natural basis for $I\!R^M$ in the context of cross-sections, namely the vectors $u_1, ..., u_M$ for which the j-th coordinate is given by

$$u_i^j = (\tau_j)^{i-1} \quad (j = 1, ..., M)$$

where the τ_j are the maturities. If V_k is the subspace spanned by $u_1, ..., u_k$, then the eigenvectors can be projected onto the subspaces $V_1, V_2, ...$ in order to understand the structure of the sample covariance matrix. This method differs from that of Dybvig, who does a principal components analysis of the *changes* in returns.

Analysis II For each cross-section, e.g. at day t_i, we fitted the model $\widetilde{R}_i^j(\omega) = a\tau_j + b(1 - e^{-\gamma\tau_j}) + \varepsilon_{ij}(\omega)$ to the observations R_i^j, where the ε_{ij} are independent, $N(0, 1)$ distributed random variables. This amounts to solving the problem

$$\min_{a,b,\gamma} \sum_{j=1}^{M} (R_i^j - a\tau_j - b(1 - e^{-\gamma\tau_j}))^2,$$

keeping a and γ non-negative. This is not perhaps as arbitrary as might at first sight appear; this is exactly the form of returns which one obtains if one takes the Vasicek (or CIR) model with $\sigma = 0$.

Before we describe further analyses, let us first give a quick review of the underlying theory. The simple one-factor models of Vasicek and Cox, Ingersoll and Ross define the short-rate process r as the solution of a stochastic differential equation

$$dr_t = \sigma\ (r_t)^\theta\ dW_t\ +\ (\alpha - \beta' r_t)\ dt \tag{2.0}$$

under the real-world (probability) measure Q', where α, β', σ are non-negative constants (the parameters of the model) and the power $\theta = 0$ in the Vasicek model, $\theta = 1/2$ in the CIR model. It is widely accepted that nominal interest rates should never be negative, and therefore a model (such as the Vasicek model) which allows rates to go negative is not valid. Such models are widely used in practice because of their tractability, and therefore we have included it in the study. Rogers (1996) investigates some of the possible snags which can arise, but we shall ignore them here; such a Gaussian model is usually an adequate first attempt.

It is well known that in arbitrage-free pricing system there is a measure Q (called the risk-neutral measure) which is equivalent to the real-world measure Q' such that the zero coupon bond prices can be expressed as

$$P(t,T) \;:=\; EQ_t\left[\exp\left(-\int_t^T r_u du\right)\right], \tag{2.1}$$

where EQ_t denotes the conditional expectation with respect to the risk-neutral measure Q, given all information available at time t. For simplicity, in this paper we only consider measures

$$Q \;:=\; Z_{0,T^*} \;*\; Q'$$

having density

$$Z_{0,T^*} \;:=\; exp\left(\int_0^{T^*} \lambda_u \, dW_u - \frac{1}{2}\int_0^{T^*} \lambda_u \, du\right)$$

with respect to Q', where the function λ (called market price of risk, see below) has the following particular form:

$$\lambda_u \;:=\; \frac{(\beta' - \beta)\; r_u}{\sigma}, \quad \beta > 0, \qquad \textit{in the Vasicek model},$$

$$\lambda_u \;:=\; \frac{(\beta' - \beta)\; r_u}{\sigma\;\sqrt{r_u}} \;=\; \frac{(\beta' - \beta)\;\sqrt{r_u}}{\sigma}, \quad \beta > 0, \qquad \textit{in the CIR model}.$$

Here, T^* denotes the maximum maturity at the last day of the observation period.

By the well-known Girsanov theorem (see e.g. the original articles of Girsanov (1960) and Maruyama (1954, 1955); further explanatory material can be found e.g. in the books of Liptser & Shiryayev (1977), Rogers & Williams (1987), Karatzas & Shreve (1991) and e.g. in the papers of Stummer (1990, 1993, 1999)) one gets the dynamics of the short rate process r_t under the risk neutral measure Q:

$$dr_t \;=\; \sigma\;(r_t)^\theta\; d\overline{W_t} \;+\; (\alpha - \beta r_t)\; dt \tag{2.2}$$

with $\overline{W_t} := W_t - \int_0^t \lambda_u \, du$, which is a Brownian motion under Q. The power θ remains the same as in (2.0).

Remark: Notice that, in both the CIR and Vasicek models, the solution laws Q of (2.0) and Q' of (2.2), which correspond to different values of β, are locally equivalent. The same is true in the Vasicek model (i.e. $\theta = 0$) if we would also use α' (rather than α) in (2.0); the corresponding market price of risk would be $\lambda_u := \big((\alpha' - \alpha) - (\beta' - \beta)\, r_u\big)/\sigma$. However, in the CIR model (i.e. $\theta = 1/2$), if one uses (2.0) with α', then one still gets a (unique strong) solution measure $\overline{\overline{Q'}}$ of (2.0), but it is, in general, not locally equivalent to the (unique strong) solution measure Q of (2.2) anymore. The reason is related to the discussion in Section 5 of Cox, Ingersoll & Ross (1985), which in turn hinges on properties of squared Bessel processes. To explain more fully, if one restricts the dimension (defined to be $4\alpha/\sigma^2$) of the BESQ processes to be always at least 2, then for a starting short rate level $r_0 > 0$ the laws of the process for different dimension are locally equivalent. Indeed, as is presented in Exercise 1.22 on p. 419 of Revuz & Yor (1991), there is a simple expression for the density martingale. However, for starting short rate level $r_0 = 0$, the laws for different dimension are mutually singular; the escape rates from the short rate level 0 are different. If the dimension is allowed to fall below 0, the short rate level 0 becomes accessible for the process r_t, and so the laws cannot be locally equivalent.

For simplicity, we have kept the parameter α the same in (2.0) and (2.2) (in both the Vasicek and CIR contexts).

From (2.1) and (2.2) one gets the following explicit formulae for the return under the risk-neutral measure Q:

$$R(t,T) = r_t\, B(T-t) + A(T-t) , \qquad (0 \le t \le T),$$

where in the case of the Vasicek model

$$A(\tau) = \left(\frac{\alpha}{\beta} - \frac{\sigma^2}{2\beta^2}\right)\tau - \frac{\alpha}{\beta^2}\,(1 - e^{-\beta\tau}) + \frac{\sigma^2}{4\beta^3}\,(3 - 4e^{-\beta\tau} + e^{-2\beta\tau}) ,$$

$$B(\tau) = (1 - e^{-\beta\tau})/\beta ,$$

and in the case of the CIR model

$$A(\tau) = -\frac{2\alpha}{\sigma^2}\,\log\left(\frac{\gamma\, e^{\beta\tau/2}}{\gamma\,\cosh\gamma\tau + \frac{1}{2}\,\beta\,\sinh\gamma\tau}\right) ,$$

$$B(\tau) \;=\; \frac{\sinh \gamma\tau}{\gamma\;\cosh \gamma\tau \;+\; \frac{1}{2}\,\beta\;\sinh \gamma\tau}\;,$$

with $\gamma \equiv \frac{1}{2}\sqrt{\beta^2 + 2\sigma^2}$.

From this, one can immediately see two reasons why the Vasicek and CIR models are nice: the return has an explicit formula, and it is an affine function of the short rate.

Another feature of the Vasicek and CIR models is that the long rate $\lim_{T\to\infty} \frac{1}{T} R(0,T)$ exists under the risk-neutral measure Q and is equal to

$$\begin{array}{ll} \dfrac{\alpha}{\beta} - \dfrac{\sigma^2}{2\beta^2} & \qquad \textit{in the Vasicek model}, \\ \dfrac{2\alpha}{\beta + 2\gamma} & \qquad \textit{in the CIR model}, \end{array}$$

a constant in each case.

After this preparatory excursion in term structure theory, let us now describe further analyses:

Analysis III The next procedure we used was to fit the Vasicek and CIR models *day by day* to the cross-section of returns; that is, we solved for each $i = 1, ..., N$ the problem

$$\min_{\alpha,\;\beta,\;\sigma^2,\;r} \sum_{j=1}^{M} (R_i^j - A(\tau_j) - rB(\tau_j))^2,$$

keeping α, β and σ^2 non-negative (and the short rate r also non-negative in the CIR case). Hence, for every day i we obtain different estimates.

Analysis IV This was a least-squares fit of the two models to all N cross-sections of returns, keeping the parameters α, β and σ the *same* for all observation days; we solved

$$\min_{\alpha,\;\beta,\;\sigma^2,\;r_i} \sum_{i=1}^{N} \sum_{j=1}^{M} (R_i^j - A(\tau_j) - r_i B(\tau_j))^2,$$

keeping α, β and σ^2 non-negative (and r_i also non-negative in the CIR case).

Let us now explain the FML method which was already mentioned above:

Analysis V We performed a *full* maximum-likelihood fit to all N cross-sections. To describe this, we model the observed returns by

$$R_i^j = A(\tau_j) + r(t_i)B(\tau_j) + \varepsilon_{ij},$$

where the ε_{ij} are independent $N(0, v)$ random variables. If $p_t(\cdot,\cdot)$ denotes the transition density of the diffusion (which depends implicitly on the parameters α, β' and σ), we solved the problem

$$\max_{\alpha,\ \beta,\ \beta',\ \sigma^2,\ v,\ r_i} \sum_{i=2}^{N} \log p_\delta(r_{i-1}, r_i) - \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\{\log v + (R_i^j - A(\tau_j) - r_i B(\tau_j))^2/v\}$$

where δ is the spacing between times of observation. Note that the mean reversion parameter of the diffusion is not assumed to be the same β as appears in the bond price formula, since that relates to the risk-neutral measure, and the β' of the actual diffusion of the short rate in the real-world measure may be different.

Analysis VI In Analyses III and IV, there appeared to be considerable indeterminacy in the best parameters. As we shall explain, this was essentially to be expected; however, we investigated this further by considering what happened when we reduced the number of parameters by one, by fixing a value for the long rate, and then carrying out Analysis IV. The results helped to reconcile the models of Brennan & Schwartz (1979) and Schaefer & Schwartz (1987) with the result of Dybvig, Ingersoll & Ross (1996) that the long rate is a non-decreasing process.

Analysis VII One-factor models imply that the changes in returns of zero-coupon bonds of different maturities are instantaneously perfectly correlated. For the Vasicek and CIR models, we even have that if one forms

$$\delta_i^j \equiv R_{i+1}^j - R_i^j, \quad i = 1, ..., N-1, \quad j = 1, ...M,$$

then $\delta_i^j = (r_{i+1} - r_i)B(\tau_j)$. Of course such a phenomenon cannot occur in practice, but if one allowed for the possibility that the observed changes were modelled instead as

$$\delta_i^j = w_i u_j + \varepsilon_{ij} \tag{0.1}$$

where the ε_{ij} are independent $N(0, v)$ and w, u are two vectors, we computed the likelihood-ratio test statistic for testing this hypothesis against a more general hypothesis on δ_i^j. More simply, we investigated the *signs* of the δ_i^j; if the one-factor Vasicek or CIR models were true, then even with some quantisation of the observations one should have that for each i *either* $\delta_i^j \geq 0$ for all j *or* $\delta_i^j \leq 0$ for all j.

3. Results of the analyses

(I) The results are presented in Table 1. The columns record (respectively) the label of the data set, the ratio of the eigenvalue to the largest eigenvalue, and the squared length of the eigenvector accounted for by projection into the subspaces V_1, V_2, V_3 defined in Section 2. The rows relate to the three eigenvectors of largest eigenvalue.

The overwhelming conclusions from this are that (i) *the sample covariance matrix is almost of rank 1* (the top eigenvalue is always much larger than all the others); (ii) the jth eigenvector is largely made up of its projection into V_j, but not into V_{j-1}. In words, this means that *the displacements of the R_i from the mean value $\bar{R}$ are mainly due to adding a constant to the yield curve, a little due to adding a linear function to the yield curve, and even less to addition of a quadratic function to the yield curve.* This is consistent with the findings of Dybvig (1997) based on the CRSP data.

(II) The residual sum of squares was typically of the order of ????????? quote in terms of bp per maturity ??????????? 10^{-5} for each day (the outcomes for datasets A-D are summarised in box-and-whisker plots in Figure 1. Any observation which exceeds the 75-percentile by more than 1.5 times the interquartile range is marked separately.) This corresponds to an average daily error in each fitted yield of around 2–4 basis points. It is remarkable that the quality of fit in this deterministic case is almost as good as in the day-by-day fits of Analysis III.

Bearing in mind bid-ask spreads, and the fact that most parts of the yield curve have to be deduced from the prices of coupon-bearing bonds (which might be quoted to the nearest 1/32 of a pound in a price of about £100), there is an intrinsic measurement error of the order of 3–4 basis points.

The estimate of a corresponds to the long rate, and for each of the datasets this varied little from day to day. The estimates of b were always negative.

(III) As with Analysis II, we present in Figure 1 box-and-whisker plots of the residual sums-of-squares from fitting the CIR and Vasicek models to datasets A-D. Notice the improvement in the quality of the fit in each case, usually more pronounced for Vasicek than for CIR. Table 2 also summarizes the fits of III, giving the average daily errors (in bp). The Analysis IV and V errors are given for comparison.

In this instance, the estimated values of the parameters are meaningful, and we display in Figure 2 on common axes the daily estimates of r from the CIR and Vasicek models, for datasets A-D. While the estimates look broadly similar, it is not uncommon for them to differ by 30 basis points on any particular day. There appears to be no consistent sign for the difference in the estimates for the two models.

Information on the fluctuations in the three parameter estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^2$ from day to day is summarised in Table 3, with the estimates from Analysis IV given for comparison. In most cases, the interquartile range from Analysis III covers the estimate from Analysis IV, though often the range is quite large, underlining the instability of parameter estimates. The parameter instability in Analysis III is consistent with e.g. the abovementioned findings of Brown & Schaefer (1994) on index-linked bonds.

We also display, in Figure 3, the implied long rates calculated from the estimated values of α, β, σ in each of the two models, for each of the datasets A-D. What is striking is the relative stability of th estimated long-rate in the CIR model compared with the estimated long-rate in the Vasicek model; commonly, the estimates in the latter were negative (which is manifestly absurd!), and frequently enormously negative. In Figure 3, any negative estimate below -68% is marked by a "Z" and not joined to the rest of the graph. Long rates as low as $-10^9\%$ were obtained !

(IV) The box-and-whisker plots in Figure 1 show that the quality of fit for dataset A (May 1993) is scarcely affected, and the Vasicek model fits better than CIR. For datasets B-D, the fit is considerably worse. This suggests that a fixed model is not explaining the term structure on a longer time-scale so well as on a shorter one, which seems reasonable; the remarkable thing is that it is still comparable, bearing in mind that in Analysis III we have $3(D-1)$ more parameters, where D is the number of days being fitted! This is explained by the results of Analysis I; we have seen that in each of the data sets

$$R_i \doteq \bar{R} + \varepsilon_i v_1 + \varepsilon_i' v_2 + \varepsilon_i'' v_3$$

where $\varepsilon_i \gg \varepsilon_i' \gg \varepsilon_i''$ and v_1, v_2 and v_3 are the top three eigenvectors of the sample covariance matrix. Also, v_1 corresponds to a parallel shift of the yield curve. So if we could choose parameters α, β, σ^2 such that for some θ

$$\bar{R} \approx (A(\tau_j; \alpha, \beta, \sigma^2) + \theta B(\tau_j; \beta, \sigma^2))_{j=1}^m \text{ and } v_1 \approx (B(\tau_j; \beta, \sigma^2))_{j=1}^m,$$

then the returns over many days can be well fitted by the same model.

When using the least-squares fitting algorithm, we printed out the current values of the variables and the objective function every 5 or 10 calls to the function. Convergence happened quite quickly (generally more quickly for the larger datasets), and in all except dataset C, the estimate of β in the CIR model appeared to be converging to zero. It was noticeable that towards the end of the optimisation procedure, the value of the objective function changed very little, though the values of the parameter may have been changing substantially; this is to be expected, because near the minimum the derivative is almost zero.

We show in Figure 4 the best estimated values of r_i for the Vasicek and CIR models, plotted on the same axes as Figure 2. Figures 2 and 4 are broadly similar, but also plainly different. This is not surprising in view of the evidence in Figure 1 for the poorer fit of Analysis IV compared to Analysis III. Figures 5 and 6 demonstrate clearly how good "poorer" is in this context; we have plotted the theoretical curve for $-\log P(0,\tau)$ and marked on the observed values also. The day chosen was one where the residual sum of squares for both models was (slightly above) average. The closeness of both fits is remarkable.

The three parameter estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^2$, which by construction are held constant for each data set, are shown in Table 3.

(V) Analysis V was performed on data sets A, B, C, D, and the results are summarised in Tables 2 and 4. As is clear from the last two columns of Table 2, including the transition density in the fit made a negligible difference to the residual sum-of-squares. Therefore, the Figures 1, 4, 5, 6 and 7 look (nearly 100%) the same in the case of Analysis V (instead of IV).

(VI) By fixing the value of the long rate, and then carrying out Analysis IV, we obtained useful information on the stability of estimated long rates. As the graphs in Figure 7 show, it was possible for quite different values of long rate to give very similar values of residual sums-of-squares, and this was more pronounced for the Vasicek model. This empirical fact saves the models of Brennan & Schwartz (1979), and of Schaefer & Schwartz (1987), from the theoretical objection of the result of Dybvig, Ingersoll & Ross (1996) that the long-rate is non-decreasing; this may be so, but the degree of uncertainty in the estimates of the long rate is found empirically to be so great that modelling the long rate (or a proxy) as a diffusion is by no means rediculous.

(VII) As is explained in the Appendix, the minimal value of tr $(X-Y)(X-Y)^T$ over rank–r $N \times M$ matrices Y is the sum of the r smallest eigenvalues of X^TX. Thus the form of the LR test statistic of $H_0 : X$ is rank r against $H_1 : X$ is rank

$r+1$ would be

$$L_r = NM \log \left(\sum_{i=r+1}^{M} \lambda_i / \sum_{j=r+2}^{M} \lambda_j \right) .$$

and for each r this was found to be much larger than the number of degrees of freedom $(N+M-2r-1)$, suggesting that *no* finite-rank restriction applies.

By considering the *signs* of $R^j_{i+1} - R^j_i$, one has a way to test the one-factor Vasicek and CIR models; for each i, we counted the number of j for which $R^j_{i+1} - R^j_i \geq 0$, which would always be either 0 or 10 if the model was exactly true. For each of the datasets A–D, on at least half of the days the count lay in between 0 and 10.

4. Summary and discussion

We summarise here the main results of this investigation (listed without order of importance):

(i) The measurement error in the quoted yields was of the order of 1–3 basis points. A day-by-day fit gave errors of the order of 2–3 basis points, whereas fitting with the same model for all days resulted in errors of the order of 3–8 basis points per yield. Thus, restricting the parameters to be constant does not make the fit essentially worse.

(ii) Throughout all (relevant) analyses, the quality of fit is nearly as good for the CIR model as for the Vasicek; stability of long rates (in the case of Analysis III) was much better. So it seems one loses little, and gains non-negativity of short rates, by using the CIR model.

(iii) The less involved method IV is closely as good as the theoretically clean, more involved full maximum-likelihood estimation (FML): There is almost no difference in the quality of the fit.

(iv) Substantially different values of long rate (if fixed) can lead to very similar values of residual sums-of-squares; the Vasicek model estimates show this effect much more extremely than those of the CIR model.

(v) Daily *changes* in the estimates of r were typically around 5–10 basis points, sometimes 15–20; this is comparable with the error of the fit. We found that the single-factor Vasicek and CIR models did not explain changes in log bond prices, and this is the reason.

(vi) The co-movement of different yields which is predicted by the one-factor models does not happen in practice.

(vii) The movements of the yield curve are largely explained by parallel displacements of the entire curve, with the addition of linear perturbations playing a smaller rôle, and quadratic and higher-order perturbations being only faintly present. Thus the vector $R_i = (R(t_i, t_i + \tau_j))_{j=1}^M$ essentially lies in a subspace of dimension at most 3.

From (i) and (vii), it is clear that the one-factor models are often not in themselves good enough to fit the data to within observational error, though if we were to go to two- or three-factors we should have a chance to fit the data well (note, however, that with three factors, we are beginning to have more parameters than data points.) The worsened fit when we insist on the same parameters for the data set rules out these simple one-factor models; allowing the parameters to change slowly from day to day (as in Rogers & Zane (1997), for example) one might be more successful, but since the fit of Analysis III is already not particularly good, it is not worth trying in this framework. The complete failure of the models to explain co-movements of different yields means that they could not be used for any option on spreads, for example.

Let us finally mention some possible future refinements of the methods which were used in this paper (for the sake of brevity, this is not complete):

One suggestion is to take a time-varying version of the basic model, as is done by Hull & White (1990). This has the virtue of allowing an exact fit to the yield curve on any given day, by choosing α to be a deterministic function of time. In terms of $(\alpha_t)_{t\geq 0}, \beta, \sigma$, the zero-coupon bond prices for the Vasicek model are given by

$$\begin{aligned} -\log P(t,T) &= r_t \frac{1-e^{-\beta\tau}}{\beta} - \frac{\sigma^2}{4\beta^3}\left[2\beta\tau - 3 + 4e^{-\beta\tau} - e^{-2\beta\tau}\right] \\ &+ \int_t^T du \int_t^u ds\; e^{-\beta(u-s)}\alpha_s, \end{aligned} \tag{0.2}$$

where $\tau \equiv T - t \geq 0$; see, for example, Section 2 of Rogers (1995a). From this, we see that $0 < t < \tau$, the change in the maturity–τ log bond price from time 0 to time t is

$$-\log P(t,t+\tau) + \log P(0,\tau) = (r_t - r_0)\frac{1-e^{-\beta\tau}}{\beta} + \int_0^\tau du \int_0^u ds\; e^{-\beta(u-s)}(\alpha_{t+s} - \alpha_s)$$

which does not necessarily have the same sign at all points of the yield curve. Thus the feature of the Vasicek and CIR models which was addressed in Analysis VII is not a problem for the time-dependent version of the model.

For such time-dependent versions of Vasicek and CIR, the volatility of the yield of maturity τ is always decreasing with τ, which is not invariably observed in data; indeed, a humped term-structure of volatility seems more common as activity in very short-dated loans is small.

As a point of modelling style, it seems preferable to try representing α as

$$\alpha(t) = \sum_{i=1}^{K} a_i e^{-\lambda_i t}, \tag{0.3}$$

because then if we fit the yield curve again tomorrow, it *is* possible that the model we obtain will be exactly the same as today's, which would not be possible if we took (for example) α to be piecewise constant.

Another extension of the basic models is to consider multi-factor generalisations, as has been done by Longstaff & Schwartz (1992), Duffie & Kan (1996), Pearson & Sun (1994), Rogers & Zane (1997) among many others. The fits obtained are of course much better and are typically of the order of the observational error.

Apart from different modelling assumptions, there is scope for improved fitting if one fits to better data. The yield curve data is typically already derived, i.e. extracted from coupon bond prices; except for a some cases (e.g. US Treasury strips), one does not have direct evidence for the prices of zero-coupon bonds. If one used the prices of liquid interest-rate derivatives for the fitting procedure, then one would expect that data to be more reliable, and the whole estimation procedure would be a lot cleaner. In fact, pricing and calibration are very closely linked; a subroutine which prices a particular derivative within a particular model can be used *either* as part of a fitting procedure, *or else* can be used to give a price once the model has been fitted.

Another area where much remains to be done is in the analysis of term-structure models in *many* countries at once. The theory of this turns out to be surprisingly simple if viewed in the correct way (see Rogers (1995b)), and fitting investigations by Rogers & Zane (1997) are encouraging.

Acknowledgement

We are very grateful to Simon Babbs for letting us use some of his yield-curve data in this study. ....

Appendix

The test described in Analysis VI rests on the following attractive result. Though this must appear in print somewhere, the proof is not long, so we include it for completeness.

PROPOSITION 1 *Let X be an $N \times M$ matrix, $N \geq M$, and let $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_M \geq 0$ be the eigenvalues of $X^T X$. Then for $1 \leq r \leq M$*

$$\min\{tr(X - AB)(X - AB)^T : A \text{ is } N \times r, \quad B \text{ is } r \times M\}$$

$$= \sum_{i=r+1}^{M} \lambda_i.$$

Proof. Firstly, observe that the minimum is not changed if we replace X by RXS for any $R \in O(N)$, $S \in O(M)$, where $O(N)$ denotes the set of all orthogonal $N \times N$ matrices; so we may assume that both $X^T X$ and XX^T are diagonal non-negative definite. Thus we shall have

$$X^T X = \Lambda \equiv \text{diag}(\lambda_1, ..., \lambda_M).$$

Moreover, if for any rectangular matrix Z we have $ZZ^T v = \lambda v$, with $\lambda \neq 0$, $v \neq 0$, then $(Z^T Z)Z^T v = \lambda Z^T v$, so λ is also an eigenvalue of $Z^T Z$. Hence we see that

$$XX^T = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}$$

(at least after suitably interchanging rows of X). Now suppose that the minimum is attained at $A = A_*, \ B = B_*$. By considering a perturbation $A_* + Y$ for small Y, we see that

$$B_*(X - A_* B_*)^T = 0$$

and similarly that

$$A_*^T (X - A_* B_*) = 0.$$

Notice that for any invertible $r \times r$ matrix U, $(A_* U, \ U^{-1} B_*)$ will also minimise the objective function, so we shall suppose U chosen to make $U^T A_*^T A_* U = I$ and write A_* in place of $A_* U$. So we have

$$(X - A_* B_*)B_*^T = 0$$
$$A_*^T (X - A_* B_*) = 0$$

from which $B_* = A_*^T X$, and

$$tr(X - A_* B_*)(X - A_* B_*)^T = tr(I - A_* A_*^T) XX^T (I - A_* A_*^T).$$

Thus, since $P = I - A_* A_*^T$ is an orthogonal projection, of rank $N - r$, we are led to consider the problem

$$\begin{aligned} & \min\{trPXX^TP : \quad P \text{ is } N \times N, \text{ rank } N-r \text{ orthogonal projection}\} \\ = \; & \min\{trPXX^T : \quad P \text{ is } N \times N, \text{ rank } N-r \text{ orthogonal projection}\}. \end{aligned}$$

Now XX^T is a diagonal matrix, so we are attempting to minimise

$$\sum_{i=1}^{N} \lambda_i p_{ii}$$

(where $\lambda_j = 0$ for $M < j \leq N$) with P any rank $N - r$ orthogonal projection. However, note that $P = P^2 = PP^T = P^TP$, so $1 \geq p_{ii} \geq 0$ for all i, and $\sum p_{ii} = N - r$, so the minimum must be at least $\sum_{i=r+1}^{N} \lambda_i$. This lower bound can be achieved by taking

$$A_* = \begin{pmatrix} I_r \\ 0 \end{pmatrix}$$

and $B_* = A_*^T X$.

References

Albrecht, P. (1985). A note on immunization under a general stochastic equlilibrium model of the term structure. Insurance: Mathematics and Economics 4, 239–244.

Anderson, N., Breedon, F., Deacon, M., Derry, A., Murphy, G. (1996). Estimating and Interpreting the Yield Curve. John Wiley & Sons, Chichester.

Ang, A., Sherris, M. (1997). Interest rate risk management: developments in interest rate term structure modeling for risk management and valuation of interest-rate-dependent cash flows. North American Actuarial Journal 1, No.2, 1–26.

Back, K. (1996). Yield curve models: a mathematical review. In: Nelkin, I. (Ed), Option Embedded Bonds: Price Analysis, Credit Risk, and Investment Strategies, Irwin, Homewood IL, pp. 3–36.

Barone, E., Cuoco, D., Zautzik,E. (1991). Term structure estimation using the Cox, Ingersoll, and Ross model: The case of Italian Treasury bonds. The Journal of Fixed Income 1, 87–95.

Baxter, M., Rennie, A. (1996). Financial Calculus. Cambridge University Press, Cambridge.

Beekman, J.A., Shiu, E.S.W. (1988). Stochastic models for bond prices, function space integrals and immunization theory. Insurance: Mathematics and Economics 7, 163–173.

Bingham, N.H., Kiesel, R. (1998). Risk-neutral Valuation. Springer, London.

Beaglehole, D.R., Tenney, M.S. (1991). General solution of some interest rate-contingent claim pricing equations. The Journal of Fixed Income 1, 69–83.

Brennan, M.J., Schwartz, E.S. (1979). A continuous time approach to the pricing of bonds. Journal of Banking and Finnance 3, 133–155.

Brown, S.J., Dybvig, P.H. (1986). The empirical implications of the Cox-Ingersoll-Ross theory of the term structure of interest rates. The Journal of Finance 41, 617–630.

Brown, R.H., Schaefer, S.M. (1994). The term structure of real interest rates and the Cox, Ingersoll and Ross model. Journal of Financial Economics 35, 3–42.

Boyle, P.P. (1978). Immunization under stochastic models of the term structure. Journal of the Institute of Actuaries 105, 177–187.

Boyle, P.P. (1980). Recent models of the term structure of interest rates with actuarial applications. Transactions of the 21st International Congress of Actuaries, Vol 4, 95–104.

Boyle, P.P. (1995). Risk-based capital for financial institutions. In: Ottaviani, G. (Ed), Financial Risk in Insurance, Springer, Berlin, pp. 47–62.

Boyle, P.P., Yang, H. (1997). Asset allocation with time variation in expected returns. Insurance: Mathematics and Economics 21, 201–218.

Bühlmann, H. (1995). Life insurance with stochastic interest rates. In: Ottaviani, G. (Ed), Financial Risk in Insurance, Springer, Berlin, pp. 1–24.

Carriere, J.F. (1999). Long-term yield rates for actuarial valuations. To appear in North American Actuarial Journal 3, No.3.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B. (1992). An empirical comparison of alternative models of the short-term interest rate. The Journal of Finance 47, 1209–1227.

Constantinides, G.M. (1992). A theory of the nominal term structure of interest rates. The Review of Financial Studies 5, 531–552.

Corradi, C. (1996). On the estimation of smooth forward rate curves from a finite number of observations: a comment. Insurance: Mathematics and Economics 18, 115–117.

Cox, J.C., Ingersoll, J.E., Ross, S.A. (1981). A re-examination of traditional hypotheses about the term structure of interest rates. The Journal of Finance 36, 769–798.

Cox, J.C., Ingersoll, J.E., Ross, S.A. (1985). A theory of the term structure of interest rates. Econometrica 53, 385–407.

Deelstra, G., Delbaen, F. (1995a). Long-term returns in stochastic interest rate models. Insurance: Mathematics and Economics 17, 163–169.

Deelstra, G., Delbaen, F. (1995b). Long-term returns in stochastic interest rate models: convergence in law. Stochastics and Stochastics Reports 55, 253–277.

Deelstra, G., Delbaen, F. (1998). Convergence of discretized stochastic (interest rate) processes with stochastic drift term. Applied Stochastic Models and Data Analysis 14, 77–84.

Delbaen, F. (1993). Consols in the CIR model. Mathematical Finance 3, 125–134.

Delbaen, F., Lorimier, S. (1992). Estimation of the yield curve and the forward rate curve starting from a finite number of observations. Insurance: Mathematics and Economics 11, 259–269.

De Munnik, J.F.J, Schotman, P.C. (1994). Cross-sectional versus time series estimation of term structure models: empirical results for the Dutch bond market. Journal of Banking and Finance 18, 997–1025.

De Schepper, A., Goovaerts, M., Delbaen, F. (1992). The Laplace transform of annuities certain with exponential time distribution. Insurance: Mathematics and Economics 11, 291–294.

Duffie, D. (1996). Dynamic Asset Pricing Theory, 2nd ed. Princeton University Press, Princeton.

Duffie, D., Kan, R. (1996). A yield factor model of interest rates. Mathematical Finance 6, 379–406.

Dybvig, P.H. (1997). Bond and bond option pricing based on the current term structure. In: Dempster, M. H., Pliska, S.R. (Eds), Mathematics of Derivative Securities, Cambridge University Press, pp. 271–293.

Dybvig, P.H., Ingersoll, J.E., Ross, S.A. (1996). Long forward and zero-coupon rates can never fall. Journal of Business 69, 1–25.

El Karoui, N., Myneni, R., Viswanathan, R. (1992). Arbitrage pricing and hedging of interest claims with state variables I: Theory. Stanford University and Université Paris VI working paper.

Gibbons, M.R., Ramaswamy, K. (1993). A test of the Cox, Ingersoll, and Ross model of the term structure. The Review of Financial Studies 6, 619–658.

Girard, L.N. (2000). Market value of insurance liabilities reconciling the actuarial appraisal and option pricing methods. To appear in North American Actuarial Journal 4, No.1.

Girsanov, I.V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory of Probability and its Applications 5, 285–301.

Hull, J., White, A. (1990). Pricing interest-rate derivative securities. The Review of Financial Studies 3, 573–592.

Jamshidian, F. (1995). A simple class of square-root interest rate models. Applied Mathematical Finance 2, 61–72.

Jamshidian, F. (1996). Bond, futures and option evaluation in the quadratic interest rate model. Applied Mathematical Finance 3, 93–115.

Karatzas, I., Shreve, S.E. (1991). Brownian Motion and Stochastic Calculus, 2nd ed. Springer, Berlin.

Kwok, Y.K. (1998). Mathematical Models of Financial Derivatives. Springer, Singapore.

Lamberton, D., Lapeyre, B. (1996). Introduction to Stochastic Calculus Applied to Finance. Chapman & Hall, London.

Liptser, R.S., Shiryayev, A.N. (1977). Statistics of Random Processes, Vol. I. Springer, New York.

Longstaff, F.A., Schwartz, E.S. (1992). Interest-rate volatility and the term structure: a two factor general equilibrium model. The Journal of Finance 47, 1259–1282.

Maruyama, G. (1954). On the transition probability functions of the Markov processes. Natural Science Reports of the Ochanomizu University 5, 10–20.

Maruyama, G. (1955). Continuous Markov processes and stochastic equations. Rendiconti del Circolo Matematico di Palermo, Serie II 4, 48–90.

Moriconi, F. (1995). Analyzing default-free bond markets by diffusion models. In: Ottaviani, G. (Ed), Financial Risk in Insurance, Springer, Berlin, pp. 25–46.

Musiela, M., Rutkowski, M. (1997). Martingale Methods in Financial Modelling. Springer, Berlin.

Panjer, H.H. (Ed.) (1998). Financial Economics - with Applications to Investments, Insurance and Pensions. The Actuarial Foundation, Schaumburg IL.

Pearson, N.D., Sun, T.-S. (1994). Exploiting the conditional density in estimating the term structure: an application to the Cox, Ingersoll, and Ross model. The Journal of Finance 47, 1279–1304.

Pedersen, H.W., Shiu, E.S.W. (1994). Evaluation of the GIC rollover option. Insurance: Mathematics and Economics 14, 117–127.

Revuz, D., Yor, M. (1991). Continuous Martingales and Brownian Motion. Springer, New York.

Rebonato, R. (1998). Interest-Rate Option Models: Understanding, Analysing and Using Models for Exotic Interest-Options, 2nd ed. John Wiley & Sons, Chichester.

Reitano, R.R. (1997). Two paradigms for the market value of liabilities. North American Actuarial Journal 1, No.4, 104–122. Discussion 122–137.

Richard, S.F. (1978). An arbitrage model of the term structure of interest rates. Journal of Financial Economics 6, 33–57.

Rogers, L.C.G. (1995a). Which model for the term-structure of interest rates should one use ? In: Davies, M.H.A., Duffie, D., Fleming, W.H., Shreve S.E. (Eds), Mathematical Finance, Springer, New York, pp. 93–115.

Rogers, L.C.G. (1995b). The potential approach to term-structure of interest rates and foreign exchange rates. Mathematical Finance 7, 157–176.

Rogers, L.C.G. (1996). Gaussian errors. RISK 9, 42–45.

cite ????? Rogers, L.C.G., Stummer, W. (1994). How well do one-factor models fit bond prices ? University of Bath Working Paper.

Rogers, L.C.G., Williams, D. (1987). Diffusions, Markov processes, and Martingales, Vol. II. John Wiley & Sons, Chichester.

Rogers, L.C.G., Zane, O. (1997). Fitting potential models to interest rate and foreign exchange rates. In: Hughston, L.P. (Ed), Vasicek and beyond, RISK Publications, London, pp. 327–342.

Rolski, T., Schmidli, H., Schmidt, V., Teugels, J. (1999). Stochastic Processes for Insurance and Finance. John Wiley & Sons, Chichester.

Santomero, A.M., Babbel, D.F. (1997). Financial Risk Management by Insurers: An Analysis of the Process. The Journal of Risk and Insurance 64, No. 2, 231–270.

Schaefer, S.M., Schwartz, E.S. (1987). Time-dependent variance and the pricing of bond options. The Journal of Finance 42, 1113–1128.

Sherris, M. (1994). A one-factor interest rate model and the valuation of loans with prepayment provisions. Transactions of the Society of Actuaries 46, 251–301.

Stambaugh, R.F. (1988). The information in forward rates: implications for models of the term structure. Journal of Financial Economics 21, 41–70.

Stummer, W. (1990). The Novikov and Entropy Conditions of Diffusion Processes with Singular Drift. PhD thesis, University of Zürich.

Stummer, W. (1993). The Novikov and entropy conditions of multidimensional diffusion processes with singular drift. Probability Theory and Related Fields 97, 515–542.

Stummer, W. (1999). On Novikov and arbitrage properties of multidimensional diffusion processes with exploding drift. To appear in Statistics and Probability Letters.

Tilley, J.A. (1988). The application of modern techniques to the investment of insurance and pension funds. Transactions of the 23th International Congress of Actuaries: Helsinki 1988, Volume R, 301–326.

Vasicek, O.A. (1977). An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.

Vetzal, K.R. (1994). A survey of stochastic continuous time models of the term structure of interest rates. Insurance: Mathematics and Economics 14, 139–161.

Yao, Y. (1999). Term structure models: a perspective from the long rate. To appear in North American Actuarial Journal 3, No.3.

Table 1. Structure of the eigenspaces in Analysis I

Data	e-value ratio	$\lvert P_1x\rvert^2$	$\lvert P_2x\rvert^2$	$\lvert P_3x\rvert^2$
	1.00000	0.977	0.992	0.993
A	0.00854	0.012	0.796	0.835
	0.00433	0.010	0.114	0.694
	1.00000	0.968	0.999	1.000
B	0.00505	0.027	0.753	0.944
	0.00119	0.004	0.237	0.774
	1.00000	0.979	0.998	1.000
C	0.00834	0.021	0.937	0.999
	0.00076	0.000	0.052	0.812
	1.00000	0.981	0.998	1.000
D	0.00944	0.019	0.901	0.994
	0.00040	0.000	0.094	0.813

Table 2. Comparison of quality of fit for Analyses III, IV and V via average daily errors

Dataset	Model	Average daily error (in bp)		
		Analysis III	Analysis IV	Analysis V
A	CIR	2.86	3.33	3.33
A	V	1.84	2.49	2.49
B	CIR	2.76	6.71	6.71
B	V	2.38	6.08	6.08
C	CIR	3.49	8.51	8.54
C	V	3.17	8.62	8.65
D	CIR	1.96	6.86	6.87
D	V	1.07	3.68	3.68

Table 3. Fluctuations of Analysis III – parameter estimates, compared with Analysis IV – param

		CIR				VAS		
		Analysis III			Analysis IV	Analysis III		
Dataset	Parameter	1st quart	Median	3rd quart		1st quart	Median	3rd qua
A	$\hat{\alpha}$	0.0150	0.0154	0.0159	0.0150	0.0139	0.0150	0.0160
A	$\hat{\beta}$	0.0000	0.0000	0.0000	0.0000	0.0299	0.0463	0.0624
A	$\hat{\sigma}^2$	0.0918	0.0966	0.1067	0.0929	0.0019	0.0022	0.0025
B	$\hat{\alpha}$	0.0132	0.0163	0.0212	0.0140	0.0146	0.0214	0.0259
B	$\hat{\beta}$	0.0000	0.0000	0.0000	0.0000	0.0476	0.1058	0.1621
B	$\hat{\sigma}^2$	0.0501	0.0834	0.1435	0.0619	0.0021	0.0035	0.0049
C	$\hat{\alpha}$	0.0092	0.0100	0.0113	0.0124	0.0101	0.0152	0.0195
C	$\hat{\beta}$	0.0000	0.0000	0.0000	0.0289	0.0000	0.0693	0.1469
C	$\hat{\sigma}^2$	0.0212	0.0266	0.0308	0.0303	0.0011	0.0013	0.0022
D	$\hat{\alpha}$	0.0157	0.0168	0.0194	0.0133	0.0155	0.0192	0.0252
D	$\hat{\beta}$	0.0000	0.0000	0.0000	0.0000	0.0426	0.1185	0.1415
D	$\hat{\sigma}^2$	0.0899	0.1050	0.1290	0.0613	0.0023	0.0028	0.0047

Table 4. Analysis V – parameter estimates

Dataset	Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\beta}'$	$\hat{\sigma}^2$	$\hat{v} \times 10^6$
A	CIR	0.0149	0.0000	0.0000	0.0919	2.09
A	V	0.0156	0.0550	0.8031	0.0024	1.17
B	CIR	0.0139	0.0000	0.6579	0.0607	8.49
B	V	0.0252	0.1636	2.2137	0.0044	6.97
C	CIR	0.0113	0.0187	0.3678	0.0273	14.89
C	V	0.0176	0.1136	0.8341	0.0018	15.29
D	CIR	0.0131	0.0000	1.1060	0.0590	9.63
D	V	0.0158	0.0434	2.4184	0.0024	2.76