

Modelling Bond Yield and Forward-Rate Curves for the Financial Times Actuaries British Government Securities Yield Indices

Andrew J.G. Cairns^{1 2}

Department of Actuarial Mathematics and Statistics,
Heriot-Watt University,
Edinburgh, EH14 4AS, United Kingdom

Abstract

This paper discusses the existing approach to gilt yield curves used by the Financial Times and suggested by Dobbie and Wilkie (1978, 1979). This approach splits bonds into high, medium and low coupon bands and fits separate yield curves to each. In recent years this method has been identified as susceptible to ‘catastrophic’ jumps when the least squares fit jumps from one set of parameters to another set of quite different values. This problem is a result of non-linearities in the least-squares formula which give rise to more than one local minimum. A desire to remove the risk of catastrophic changes prompted this research which is being carried out as part of the work of the Fixed Interest Indices Working Party.

Recent data has been analysed and it is demonstrated that catastrophic jumps could occur more frequently than originally thought.

A new approach suggested is the use of forward-rate curves rather than yield curves. Recent changes to the taxation of bonds means that the same forward-rate curve can be applied to both capital and income. The proposed curves appears to provide a significantly better fit than the present yield-curve formula. It is also thought that the risk of catastrophic jumps has been significantly reduced.

Keywords: yield curve; forward-rate curve; catastrophic jumps; least squares; maximum likelihood.

¹E-mail: A.Cairns@ma.hw.ac.uk

²WWW: <http://www.ma.hw.ac.uk/~andrewc/>

1 Introduction

This paper investigates two approaches to the modelling of fixed-interest bond prices and the term structure of interest rates. The work described here makes up part of that of the Fixed Interest Indices Working Party of the Institute of Actuaries and the Faculty of Actuaries. This working party was set up to consider how the fixed-interest indices in the Financial Times could be improved.

This particular investigation was stimulated by the need to find an approach which avoided certain problems identified in recent years which arise out of the use of the Dobbie and Wilkie (1978, 1979) yield-curve model. In their model parameters are estimated by minimizing a least-squares function. It is now known that the function which is being minimized can often have two or more local minima giving rise to unexpected jumps in parameter estimates from one day to the next. This problem is illustrated in Section 3.

The forward-rate curve approach described in this paper, it is hoped, avoids this problem but it also brings the actuarial profession more into line with current thinking in the field of financial economics. Furthermore, the new method became much more useful following the announcement in 1995 that the taxation of government bonds was to change.

1.1 Approaches to modelling

There are a number of different types of model which one can estimate: that is, yield curves model (strictly speaking, gross-redemption yield curves) (for example, Dobbie and Wilkie, 1978, 1979, and Section 2 in this paper); forward rates (Section 3 in this paper); zero-coupon bond yields (or spot rates); and par-yield curves. Generally these models are equivalent in the sense that if we have one model then we have full information about the others. The one exception is that if one fits a yield curve to medium-coupon stocks, say, this does not uniquely define the forward rate curve. The yield curve fitted instead depends upon the range of stocks which are included. These different approaches are described in detail in Anderson *et al.* (1996) and it is not intended to give descriptions here other than those given in Sections 2 and 3.

Typically such models are not intended to be used as stock pickers. Instead the intention is that they should be used to give a good indication of what interest rates are implicit in market prices. The data used are fixed interest gilt prices as published in the Financial Times. As with any curve fitting problem any

curve will fit well in the middle of the range of dates and less well at the end points. In particular, the curves which have been fitted do not appear to be particularly good at the short end (that is, with a remaining term of less than one year). This problem is exacerbated by the fact that yields on the shortest dated stocks are susceptible to rounding in the prices of stocks.

The model excludes certain stocks with special features, such as double-dated bonds and convertible bonds, which may give rise to prices which are apparently out of line with the market. So any term structure which is produced is giving us information about what interest rates are implicit in the prices of what one might describe as 'normal' stocks.

1.2 Changes in the Gilts Market

During the first few months of existence of the working party in 1995 the British government announced that it would be changing the taxation structure of the market. In particular, from 1996 income and capital gains would be taxed on the same basis, thus removing the advantages of low-coupon stocks for taxpayers.

One reason for this move was to allow the introduction of a zero-coupon bond market created by the 'stripping' of a limited number of existing coupon-paying bonds into a series of individual cashflows.

A knock-on effect of this was that the alternative, forward-rate curve model described in Section 3 of this paper became much more appropriate. The forward-rate model does not distinguish between income and capital and would not have given an adequate description of the market before the government's announcement. Instead, a two-dimensional model taking account of coupon as well as term was necessary (for example, see Dobbie and Wilkie, 1978, 1979, and Clarkson, 1979).

When gross redemption yields are plotted one can still see a small coupon effect. This, however, is due to the fact that the gross redemption yield is, in some sense, a weighted average of the forward rates. High-coupon stocks place more weight on lower durations so that if the forward-rate curve is falling then high-coupon stocks will have slightly higher gross redemption yields than low-coupon stocks with the same term. This ordering will be reversed if the forward-rate curve is rising.

The disappearance of the genuine coupon effect is apparent when gross redemption yields are plotted. Whereas previously there would have been a fairly wide band of yields at all terms, there is now quite a well defined curve

even before any model-fitting has taken place.

1.3 Plan

In Section 2 of this paper we consider the yield-curve approach described by Dobbie and Wilkie (1978, 1979). The discussion includes an analysis of recent data and illustrates the problem on having multiple minima.

Section 3 introduces the forward-rate curve model. The curve is a flexible model with four exponential terms and nine parameters in total. However, four of these parameters (the exponential rates) are fixed which reduces the risk of multiple solutions. Rather than take a least-squares approach, as is common in this field, here we describe fully a statistical model which underlies the prices of the stocks. This allows a rigorous process of estimation.

Section 4 gives the results from the fitting of this model to two sets of data (1988 to 1993, and 1995) and discusses various points arising from this. The analyses demonstrate that the model now provides a much better fit since the announcement of the changes in taxation.

Section 5 gives some discussion of the possibility of multiple solutions with the new model. It is thought that this problem has been removed and some discussion is given over to why it is thought that this should be the case.

2 The existing yield curve approach and problems

2.1 The current FT-Actuaries approach

Dobbie and Wilkie (1978, 1979) proposed the model

$$y(t) = A + Be^{-Ct} + De^{-Ft}$$

where $y(t)$ is the gross redemption yield on a bond with maturity date time t from now.

At the time (which has been the case up until 1995) there was a marked coupon effect whereby high-coupon gilts had higher gross yields than low-coupon gilts. (In 1978, high-coupon stocks consisted of those with a coupon of more than 7%). Such an effect existed because coupons were taxed as income whereas

capital gains were tax-free. Thus, low-coupon stocks were sought after by taxpayers. Non-tax-payers, such as pension funds were able to take advantage of the lower demand for high-coupon stocks and their resulting higher yields. It was therefore proposed that the yield curve be fitted to low-, medium- and high-coupon stocks separately. This was the approach adopted by the Joint Index Committee of the Institute of Actuaries and the Faculty of Actuaries and it has been in place since then.

(It was also argued by Clarkson (1978) around this time that not only did this coupon effect exist but that it was non-linear and convex.)

A least squares approach was adopted to the estimation of the parameters of this curve. Thus we have to minimize the function:

$$S(A, B, C, D, F) = \sum w_i [Y_i - y(t_i)]^2$$

where Y_i is the observed gross redemption yield on bond i with maturity t_i . The weights, used in the Dobbie and Wilkie (1978, 1979) approach were proportional to the market capitalization of each stock. In this analysis the w_i were all set to 1 and there was no subdivision of the stocks according to the size of coupon. Currently the calculations carried out for the Financial Times place an upper limit on C and F of 3, although such a restriction was not included in this analysis.

This model has been reformulated here in the following way:

$$y(t) = A + Be^{-Ct} + D \frac{e^{-Ft} - e^{-Ct}}{C - F}.$$

The purpose of this was to provide sensible limits as C and F get closer and closer together. Previously as C converged to F , estimates for B and D would tend to infinity. (In practice the expression would be rearranged if C and F get too close together.) With the reformulated model, the limit as C tends to F is

$$y(t) = A + Be^{-Ft} + Dte^{-Ft}$$

and the estimates for B and D do not diverge as before.

A similar curve has been used as a model for forward rates by Nelson and Siegel (1987) and Svensson (1994).

Now $y(t)$ is linear in A , B and D so that it is straightforward to write down the values of $\hat{A} = \hat{A}(C, F)$, \hat{B} and \hat{D} , given C and F . This leaves us with the problem of how to minimize the function

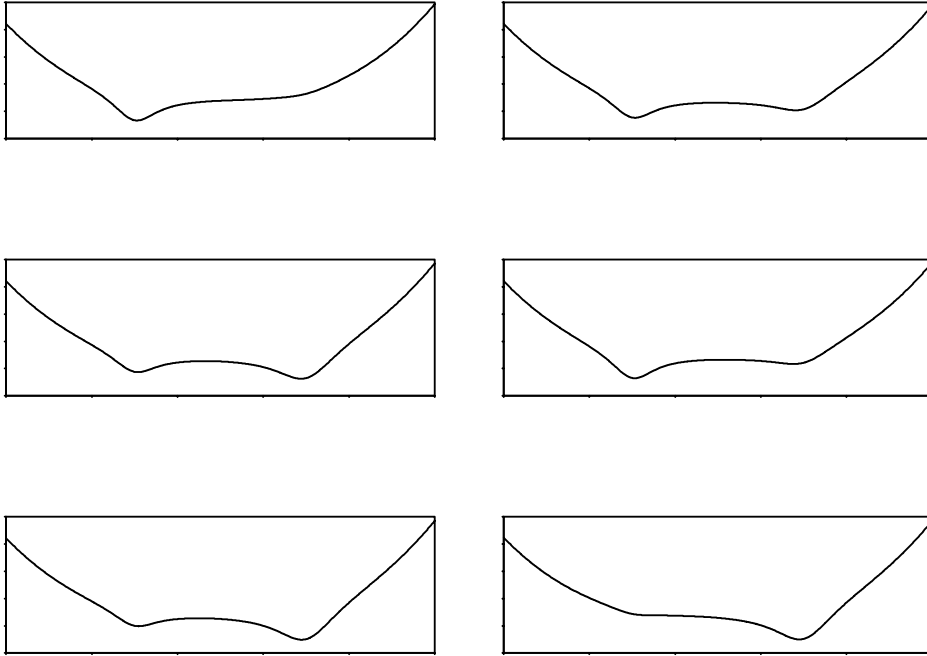


Figure 1: Example of a problem with the minimization algorithm. A catastrophic jump occurs between days 5 and 6.

$$\hat{S}(C, F) = S(\hat{A}, \hat{B}, C, \hat{D}, F)$$

It is now well known that $\hat{S}(C, F)$ can often have two or more local minima. This gave rise to occasional jumps from one minimum to another. Such a ‘catastrophic’ jump would sometimes go unnoticed. Occasionally, however, such a jump would occur during a period of relatively low volatility in prices, and the resulting change in the shape of the curve would be noticed by at least some practitioners. (However, it is not apparent that either of Nelson and Siegel (1987) and Svensson (1994) are aware of this problem.)

2.2 A simplified explanation

A stylized 2-dimensional version of this problem is illustrated in Figure 1. Suppose that we wish to minimize the function $S(t, x)$ with respect to x . This function varies randomly but continuously through time, t (measured in days). The sequence of plots in Figure 1 represents six snapshots of the evolution of this curve. On day 1, the curve has a unique minimum at A . On days 2 to 5 there are two minima at A and B , but the minimization algorithm starts at the previous minimum and stays at A (which varies slightly from one day to the next). On days 2 and 4 A represents the global minimum while on days 3 and 5 the chosen minimum at A is only a local minimum and not the global minimum, which is at B . On day 6 the algorithm jumps to what is now the unique minimum at B . This type of discontinuity between days 5 and 6 is often referred to as a ‘catastrophic’ jump and has a branch of mathematics devoted to it (for example, see Poston and Stewart, 1978, Chapter 5). It is not until day 6 when the algorithm finds the global minimum at B and this is because the local minimum at A has now disappeared. If the time sequence had been reversed then the solution to the algorithm would have been at B from time 6 down to time 2 before jumping to A at time 1 (the so-called ‘hysteresis’ effect). At the time of the catastrophic jump there may be an identifiable shift in the shape of the fitted yield curve, which would be particularly noticeable during quiet trading periods when actual yields had not shifted by any considerable margin.

In practice catastrophic jumps might occur more frequently than would strictly be the case from the above description. First, we only take a daily snapshot, so that the new starting point for minimization (that is, yesterday’s minimum) might be sufficiently far from today’s minimum that the optimization routine would find the other minimum rather than to continue to track the previous solution. Second, the minimization algorithm (the Nelder-Meade simplex algorithm) might step inadvertently outside the current minimum’s domain of attraction. Both cases have the result that the other minimum might be found despite that fact the previous minimum still exists.

2.3 Recent history

A sample period of 100 consecutive trading days during 1995 is given in Figure 2.

The program which generated this output took a different approach to the

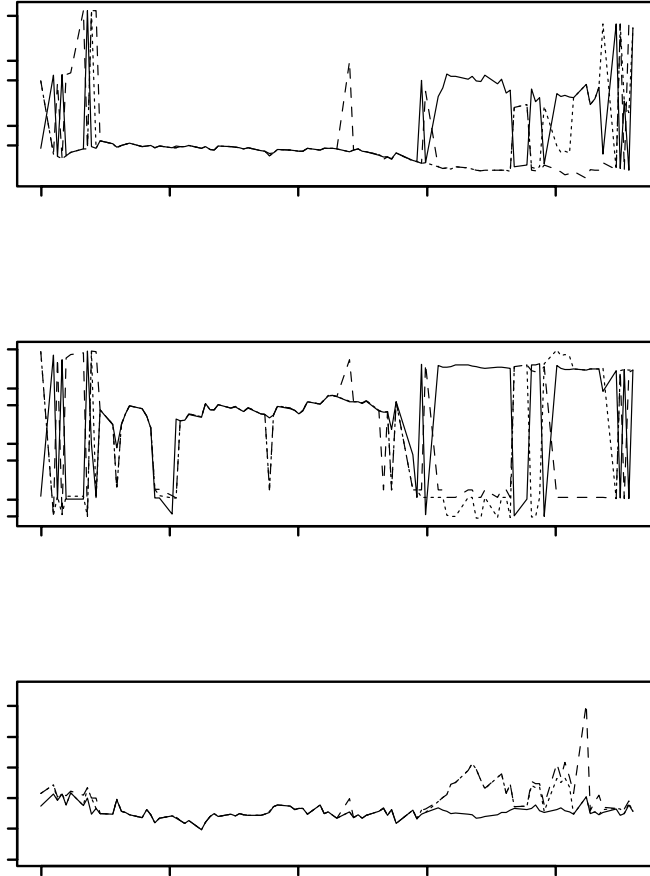


Figure 2: Evolution of the estimates for \hat{C} and \hat{F} , and of the Root Mean Squared Error (RMSE) in the estimate of the gross redemption yield during the period June to October 1995.

problem. First the data were not split into high- medium- and low-coupon stocks (for convenience). Second the weights (the w_i) were all set equal to 1. Third the program located and tracked as many local minima as possible. Different numbers of minima were observed at different times: for example, around the middle of July there appeared to be a unique minimum; around the end of July there were 2 minima; and around the middle of August there were three minima each giving noticeably different yield curves. In each of the plots in Figure 2 the solid line gives the values for the global minimum, the dotted line gives values for the second minimum and the dashed line values for the third minimum.

Figure 3 gives three yield curves on 27 September 1995 and a contour plot of $\hat{S}(C, F)$. P , Q and R were the three local minima identified on that date and mentioned above. The root mean squared errors are 18, 23 and 24 basis points, so each curve fits reasonably well. The best fit, P , succeeds by fitting the short end very well and adequately elsewhere. At terms 0 to 10 years Q and R are very similar and quite different from P . Between 10 and 15 years they are all similar, and after that P and R are similar and lower than Q . A catastrophic jump can occur when prices are relatively stable. From the example given here, such a jump would be noticeable.

Suppose, then, that we started at the global minimum at the start of June. There is always a minimum with \hat{C} in the range 0.1 to 0.5 and it is suspected that the original algorithm would have tracked this all the way through without any catastrophic jumps. However, it is necessary to track simultaneously \hat{F} to check for this continuity. The lowest values of \hat{F} , however, are slightly unreliable as many of the small values of F are combined with very small values of \hat{D} and so have very little impact on the shape of the yield curve. When this happens the minimization routine is not able to improve upon the previous estimate of F due to rounding errors.

If there were no catastrophic jumps then a potentially serious problem would have arisen during the early part of August. The old algorithm would have picked a value of \hat{C} of between 0.1 and 0.5: a local minimum. The global minimum during a good part of this period gave a much higher value for \hat{C} and, more importantly, a much better fit, and the algorithm would have completely missed this feature.

The combination of possible catastrophic jumps, of an algorithm which might track a poor local minimum, and of the recent changes in the taxation of gilts means that a fresh approach to the yield curve is now appropriate.

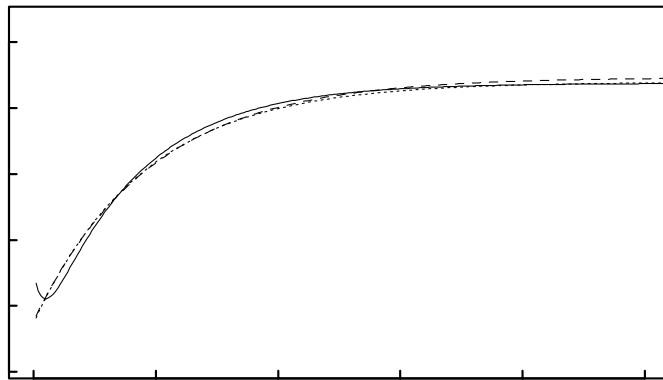
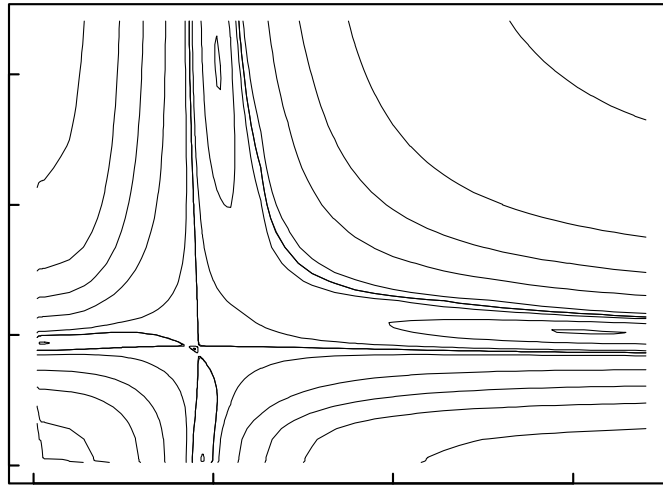


Figure 3: 27 September 1995. (Top) The function $\hat{S}(C, F)$ for this date showing the unusual feature of three local minima at P , Q and R (plus their reflections, P' and Q'). The locations of these local minima are typical for other dates. (Bottom) Three yield curves corresponding to the local minima at P (solid curve), Q (dotted curve) and R (dashed curve). The dots represent the gross redemption yields on the gilt market at the end of the day.

3 A forward-rate curve approach to the pricing of gilts

An alternative to the use of yield curves, and which is popular in the field of financial economics, is the use of forward-rate curves. In a traditional actuarial context the forward-rate at time s is sometimes referred to as the force of interest $\delta(s)$ (for example, see McCutcheon and Scott, 1986). The force of interest is, however, a one dimensional deterministic process whereas the forward-rate curve is two dimensional and evolves stochastically. This reflects the difference, for example, between an investment in an n -year zero-coupon bond and an investment in the short-term money market for n years.

In the Financial Economics literature forward-rate curves have been used directly or implicitly as a means of pricing derivative securities. For example, see Vasicek (1977), Heath, Jarrow and Morton (1992). Such models have necessarily been relatively simple in order to facilitate the pricing of derivative instruments. As a consequence the models often fit available data relatively poorly. The forward-rate curve used here is more complex and is not arbitrage-free, but it does provide a much better fit than the existing, simpler arbitrage-free models. With suitable modifications the present model could provide an arbitrage-free framework but this was not considered to be essential as it was not designed as a means of pricing derivatives: instead it is designed to give an indication of what interest rates are currently being implied by the market.

Let $f(t, s)$ be the forward rate curve observed at time t for $t \leq s < \infty$. That is, $f(t, s) = -\partial \log Z(t, s) / \partial s$ where $Z(t, s)$ is the price at time t of a zero-coupon bond maturing at time s .

The model for $f(t, s)$ used in this paper is:

$$\begin{aligned} f(t_0, t_0 + t) &= a(t_0) + \sum_{i=1}^4 b_i(t_0) e^{-c_i t} \\ &= a + \sum_{i=1}^4 b_i e^{-c_i t} \quad \text{for brevity.} \end{aligned}$$

If the value of c_i is small then the relevant value of b_i affects all durations whereas if c_i is large then the relevant value of b_i only affects the shortest durations.

Although this is a nine parameter model, the model is structured in such a way that c_1, c_2, c_3 and c_4 are regarded as fixed parameters and that we optimize

over a and the b_i only. As is described a choice of $c = (0.1, 0.2, 0.4, 0.8)$ was found to be appropriate for the period 1988 to 1995. With this set of values for c , a wide variety of shapes for the forward-rate curve can be generated by using different combinations of values for the b_i . In particular the forward-rate curve can have 0, 1, 2 or 3 turning points. Furthermore, b_1 will have more influence over the long term, while b_4 will have more influence over the short term.

This model was favoured over an alternative 5-parameter model

$$f(t_0, t_0 + t) = a(t_0) + \sum_{i=1}^2 b_i(t_0) e^{-c_i(t_0)t}$$

where optimization occurs over a , b_1 , b_2 , c_1 and c_2 . This preference arises out of the conjecture that the former model cannot give rise to catastrophic jumps while the latter is known to have multiple maxima in the likelihood function even for the simple case where there are only zero-coupon bonds (when the optimization problem becomes the same as that considered by Dobbie and Wilkie, 1978, and discussed in Section 2). A similar model for the forward-rate curve ($f(t_0, t_0 + t) = a(t_0) + b_1(t_0) \exp(-c_1(t_0)t) + b_2(t_0) \exp(-c_2(t_0)t)$) has been used by Nelson and Siegel (1987) and Svensson (1994), and this is, it is suspected, also subject to catastrophic jumps. Such models also only allow for a maximum of one turning point, so that the alternative 5 out of 9 parameter model has a much richer range of curves available for fitting the data.

The discount function at time t_0 for payments at time $t_0 + t$ is

$$\begin{aligned} v(t_0, t_0 + t) &= v(t) \text{ say, for convenience} \\ &= \exp \left[- \int_0^t f(t_0, t_0 + s) ds \right] \\ &= \exp \left[-at - \sum_{i=1}^4 \frac{b_i}{c_i} \{1 - \exp(-c_i t)\} \right] \end{aligned}$$

In the estimation problem we have N fixed interest stocks. Stock i has coupon g_i per annum per £1 nominal, with coupons payable at times $T_i = \{t_{i1}, \dots, t_{in_i}\}$ from time t_0 (so that, $0 < t_{i1} \leq 0.5$, $t_{ij+1} - t_{ij} = 0.5$ and $\tau_i = t_{in_i}$ is the time to redemption). Given a particular parameter set (a, b) ($b = (b_1, b_2, b_3, b_4)$) and corresponding forward rate curve $f(t_0, t_0 + t)$ the theoretical price of stock i is

$$\begin{aligned} \hat{P}_i &= \hat{P}_i(a, b) \\ &= \frac{g_i}{2} \sum_{t \in T_i} v(t) + v(\tau_i). \end{aligned}$$

Now the \hat{P}_i are theoretical prices, whereas the actual prices P_i are subject to errors. It is appropriate, therefore, to formulate some form of statistical model.

3.1 Statistical model

There are two sources of error taken into account here.

First, published prices are subject to rounding to the nearest $1/32$ per £100 nominal. This affects primarily the short-dated stocks and can lead to large errors in the gross redemption yield. Short-dated stocks in an appropriate model therefore will have standard deviations of about $1/32\%$ or, equivalently, the logarithm of the price has a standard deviation of $1/3200$ (and, for the sake of argument, a normal distribution). It can readily be argued that this standard deviation should be less than this, in the sense that the ‘true’ midmarket price should lie in a band of width $1/32$. The value taken here is more cautious and (as described below) simulates weightings which tend to zero as the term of a stock approaches zero.

Second, we postulate that the differences between the actual and expected yields on independent stocks are independent and identically distributed normal random variables with mean 0 and standard deviation σ . This part of the model is consistent with the statistical model which is contained implicitly in Dobbie and Wilkie’s (1978, 1979) paper.

In practice there will exist other minor features which could be added to this model. For example, each stock, i , could have a bias term, μ_i , to account for special characteristics, such as benchmark status or if it has special taxation concessions. Thus \hat{y}_i would represent the yield on a stock if it had no special characteristics. In the fitting of the forward rate curve described below, we have specifically excluded double-dated bonds and convertible bonds which have a clear bias attached to their prices which results from the built-in option. There may also be correlations between the error terms, $y_i - \hat{y}_i$, which have not been modelled here. In particular, when two ‘normal’ stocks have very similar redemption dates, if their respective errors are too far apart then this would provide something close to an arbitrage opportunity.

Now

$$\begin{aligned}
 y_i &\sim N(\hat{y}_i, \sigma^2) \\
 \Rightarrow \log P_i &\sim N\left(\log \hat{P}_i, \sigma^2 d_i^2\right) \quad (\text{approximately}) \\
 \text{where } d_i &= \text{duration of stock } i
 \end{aligned}$$

and the first source of error indicates that we would have $\log P_i \sim N(\log \hat{P}_i + 1/3200^2)$ (approximately).

Blending these two sources of error together we get the following model for a given parameter set (a, b) :

$$\begin{aligned} \log P_i &\sim N\left(\log \hat{P}_i, \sigma^2 d_i^2 + \frac{1}{3200^2}\right) \\ \text{where } d_i &= \text{duration of stock } i \\ &= \frac{\partial}{\partial y_i} \log P_i(y_i) \\ \text{and } y_i &= \text{the gross redemption yield on stock } i \end{aligned}$$

The log-likelihood function for the data is thus

$$\begin{aligned} g_1(P|a, b) &= -\frac{1}{2} \sum_{i=1}^N \left\{ \log(\sigma^2 d_i^2 / w_i) + \frac{(\log P_i - \log \hat{P}_i)^2}{\sigma^2 d_i^2 / w_i} \right\} \\ \text{where } w_i &= \frac{\sigma^2 d_i^2}{\sigma^2 d_i^2 + 1/3200^2} \end{aligned}$$

3.2 Equivalent least squares approach

It is well known that $\log x = (x - 1) - \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3 + \dots$ and that if x is close to 1 then $\log x \approx x - 1$. Therefore, since P_i and \hat{P}_i (per unit nominal) are close to 1 (that is, in the range 0.5 to 1.5) the log-likelihood function is approximately equal to

$$\tilde{g}_1(P|a, b) = -\frac{N}{2} \log(\sigma^2 d_i^2) + \frac{1}{2} \sum_{i=1}^N \log w_i - \frac{1}{2\sigma^2} \sum_{i=1}^N w_i \frac{(P_i - \hat{P}_i)^2}{d_i^2}.$$

Maximizing this function with respect to a and b is thus equivalent to minimizing the weighted least-squares function

$$\begin{aligned} S^2 &= \sum_{i=1}^N w_i \frac{(P_i - \hat{P}_i)^2}{d_i^2} \\ \text{where } w_i &= \frac{\sigma^2 d_i^2}{\sigma^2 d_i^2 + 1/3200^2} \end{aligned}$$

$$\begin{aligned}
&= \text{the weight attached to stock } i \\
&\rightarrow \begin{cases} 1 & \text{as } d_i \rightarrow \infty \\ 0 & \text{as } d_i \rightarrow 0 \end{cases}
\end{aligned}$$

This least squares approach is very similar that suggested by Phillips (personal communication) who uses $w_i = \min\{d_i, 1\}$. This approach is more familiar to many practitioners and it is reassuring to know that there is a sensible statistical model underlying it.

3.3 A Bayesian Approach

In the early stages of this investigation it was found occasionally that, for example, some predicted rates could be negative; or that the long term forward rate implied by fitted market prices (that is, estimated rather than actual prices) could tend to infinity. Such outcomes were considered to be undesirable: the first clearly leads to arbitrage possibilities and would not exist in practice; and in the second example it seems very unlikely that the market would expect interest rates to increase without bound.

It was therefore felt appropriate to introduce some way of constraining the solution to forward-rate curves which are bounded and non-negative.

One way to do this would be to maximize the log-likelihood function $g_1(P|a, b)$ subject to the constraint that $f(t_0, t_0 + t)$ is bounded and non-negative. This could still lead to solutions in which the forward rate was equal to zero for some value of t . From a subjective point of view this still seems to be unlikely since if $f(t_0, t_0 + t) = 0$ for some $t \geq 0$ we create an arbitrage opportunity of order dt . That is, we could go long in zero-coupon bonds with term t and short in zero-coupon bonds with term $t + dt$. The difference in price would be $o(dt)$ since $f(t_0, t_0 + t) = 0$ whereas investing the proceeds of the t -dated bond would create a profit of the order $r(t_0 + t).dt$. Since $r(s) > 0$ with probability greater than 0 this means we have a guaranteed profit of $O(dt)$.

We therefore adopt a Bayesian approach which incorporates this subjective viewpoint. This is done in the form of saying that before we look at the data we would say that it is more likely that the forward rate t years hence, $f(t_0, t_0 + t)$, is more likely to be, say, 4% than 2% which itself is more plausible than, say, a rate of 1%. This is done by introducing some prior distributions for the forward-rate curve.

Now we do not want the incorporation of this subjective view to swamp the data. The prior distribution was therefore chosen to reflect this view in a weak

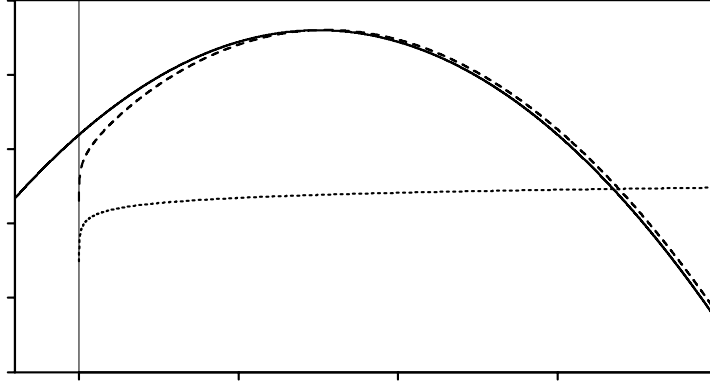


Figure 4: Possible log-likelihood (solid curve), log-prior density (dotted) and log-posterior density (dashed) curves for a parameter d . (For example, $d = f(t, t + 10)$.)

way. There are various ways of doing this, and the following prior density function for $\delta = f(t, t)$, $f(t, t + 10)$ and $f(t, \infty)$ was selected:

$$\begin{aligned} \text{prior-density} &= \exp[g_2(\delta)] = \frac{\beta^\alpha \delta^{\alpha-1} e^{-\beta\delta}}{\Gamma(\alpha)} \\ \Rightarrow \text{log-prior-density} &= g_2(\delta) = (\alpha - 1) \log \delta - \beta\delta + \text{constant} \end{aligned}$$

where $\alpha = 1.5$ and $\beta = 0.01$. β needs to be positive to ensure that we have a proper prior density function (that is, one which will integrate over its full range to 1), and to inhibit the possibility that the forward rate tends to plus-infinity. α must be greater than 1 to inhibit the forward rate from getting too close to zero. The prior distribution is very diffuse relative to the likelihood function so it has very little impact on the posterior distribution except when some points on the forward rate curve get too close to zero.

A stylized, one-dimensional representation of this is provided in Figure 4. The three curves are log-likelihood (solid curve), log-prior density (dotted curve) and log-posterior density (dashed curve) for some parameter d (for example $f(t_0, t_0)$). It can be seen that the main effect of the introduction of a prior

distribution is to ensure that the log-posterior tends to minus infinity as d tends to 0. The maximum log-likelihood occurs at $\hat{d} = 0.075$ while the maximum log-posterior occurs at about $\hat{d} = 0.078$. This shift is relatively small. In reality, however, the difference would be even smaller given that the log-likelihood will normally be concentrated over a much narrow range.

The log-posterior density is thus

$$g_3(a, b|P) = g_1(P|a, b) + g_2[f(t, t)] + g_2[f(t, t + 10)] + g_2[f(t, \infty)] + \text{constant}.$$

In practice the the estimates of parameter values were not found to be sensitive to the choice of values for α or β .

The parameter estimates are (\hat{a}, \hat{b}) where

$$g_3(\hat{a}, \hat{b}|P) = \max_{a, b} g_3(a, b|P)$$

The very diffuse nature of the prior distribution means that these estimates will be very close to the maximum likelihood estimates.

4 Results

Two sets of data were used: bi-monthly data running from 1988 to 1993 (35 data points), representing a typical period before the recent tax announcement; and 100 consecutive trading days running from June to October, 1995. The announcement early in 1995 that the taxation of gilts was to change means that yield curves can almost be drawn by hand in the 1995 dataset. Certain stocks were excluded from the fitting process: in particular, convertible stocks, those with optional redemption dates and all undated stocks (except for 4% Consols and 2.5% Treasury).

Using the 1988-1993 data (Figure 5), it was found that $\sigma = 0.001$ (i.e. 10 basis points in the yield) was about right. For the 1995 data (Figure 6) $\sigma = 0.0005$ was nearer the mark. This difference is not surprising as prior to 1995 there was a marked coupon effect which the present model for the the forward-rate curve was unable to take account of given that it does not distinguish between income and capital. Never-the-less, it was perhaps surprising that the value of $\sigma = 0.001$ was so low.

The effect of the choice of σ^2 was also considered. (It is intended that, like c , σ^2 should be fixed at the outset.) If there was no prior distribution then the value

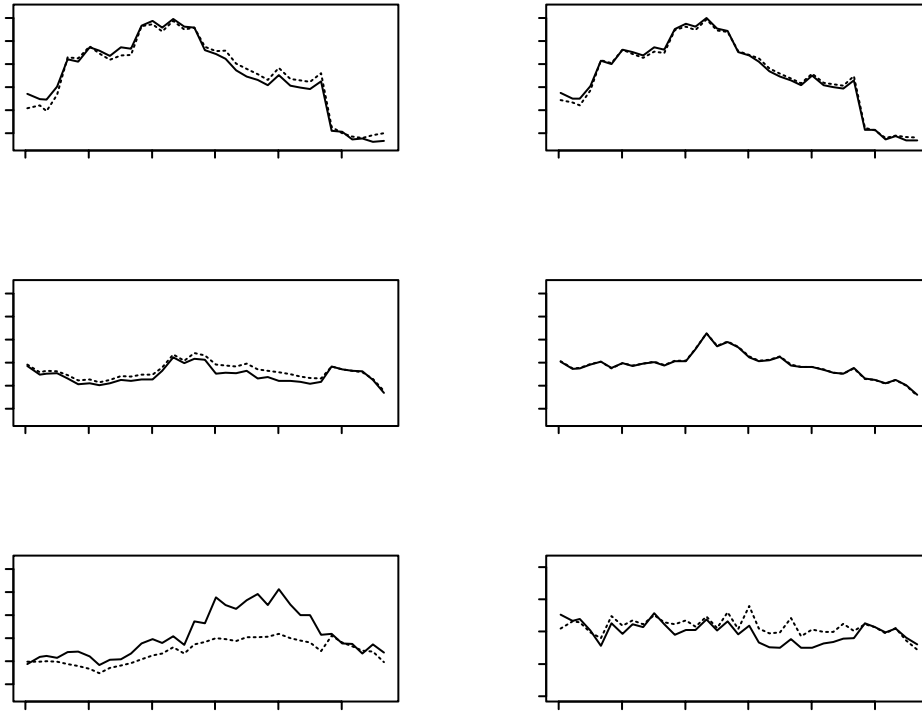


Figure 5: Evolution of various points on the forward-rate curve and on the yield curve during the period 1988 to 1993. The solid lines represent values for $c = (0.1, 0.2, 0.4, 0.8)$ and the dotted lines represent values for $c = (0.2, 0.4, 0.8, 1.6)$.

of σ^2 would have no effect on the estimates of a or b . If prior distributions are included, there is an effect on the estimates of a and b but this is very insignificant.

Results for 1988-1993 are plotted in Figure 5. (The current spot-rate $f(t, t)$ and the 3-month par-yield $y(0.25)$ are given to allow comparison with money-market rates and treasury-bill prices.)

It can be seen that there was considerable variation over this period, but this variation appeared to be relatively smooth. Possibly the least variation was observed in the 10-year forward rates and yields.

Several choices for the vector $c = (c_1, c_2, c_3, c_4)$ were considered. Of these,

$c = (0.1, 0.2, 0.4, 0.8)$ was found to be best over both the 1988-1993 and 1995 periods investigated. Other choices were only slightly worse and an example of one alternative is given in Figures 4 and 5 ($c = (0.2, 0.4, 0.8, 1.6)$). Although there was not an obvious deterioration in fit, it has already been noted by considering Figure 5 that the long-term forward rate is fairly sensitive to the choice of c . Although this gives some cause for concern, it is unlikely that this problem could be removed through the use of an alternative model, because there will always be a range of possible parameter sets which are almost as good as the optimum and which give a wide variety of long-term forward rates. Given that the 1995 dataset covers a much shorter period of time there is much less variability to observe in the evolution of the forward rate curve. However, there was a reasonable amount of randomness from one day to the next. Standard errors are considerably smaller than those in Figure 5 because the coupon effect had been almost entirely removed by this time.

4.1 Possible constraints

Two constraints on the forward-rate curve were considered:

- (a) that the estimated yield on a three-month zero-coupon bond be equal to the 3-month Treasury-Bill rate given in the FT London Money Rates table (that is, $\int_0^{0.25} f(t, t+s)ds = \log[1 + TB_{91}/4(t)]$ and TB_{91} is the mid-point 3-month Treasury-Bill rate);
- (b) that the yield, $\hat{y}(\infty)$, on an irredeemable stock, is equal to the mean of the gross redemption yields on 4% Consols and on 2.5% Treasury.

Inserting the constraint on 3-month zero-coupon bonds was found to increase the errors in the predicted yields by about 2 basis points during 1995. Treasury Bill data was not readily available for the period 1988 to 1993. It has been suggested that an alternative to the Treasury Bill rates is the 3-month Commercial Deposit rate minus 20 basis points, but this has yet to be tested.

Pegging down the long-term gross redemption yield was found to make very little difference to the goodness of fit implying that this constraint is acceptable leaving optimization over 4 rather than 5 parameters. To satisfy the constraint it is necessary to specify, for example, the values of b_1 , b_2 , b_3 and b_4 and solve numerically for the value of a which gives the correct yield on irredeemable stocks. This numerical step itself requires additional computational effort reducing the benefits of the reduction in the number of parameters to maximize over.

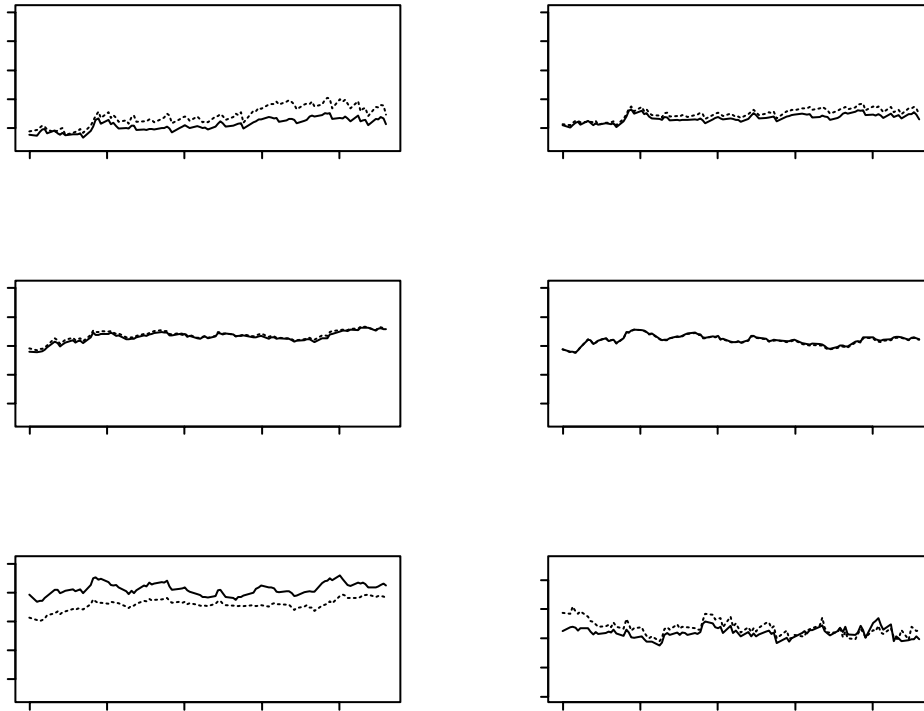


Figure 6: Evolution of various points on the forward-rate curve and on the yield curve during the period June to October 1995 (t is measured in complete months from 1 January 1995). The solid lines represent values for $c = (0.1, 0.2, 0.4, 0.8)$ and the dotted lines represent values for $c = (0.2, 0.4, 0.8, 1.6)$.

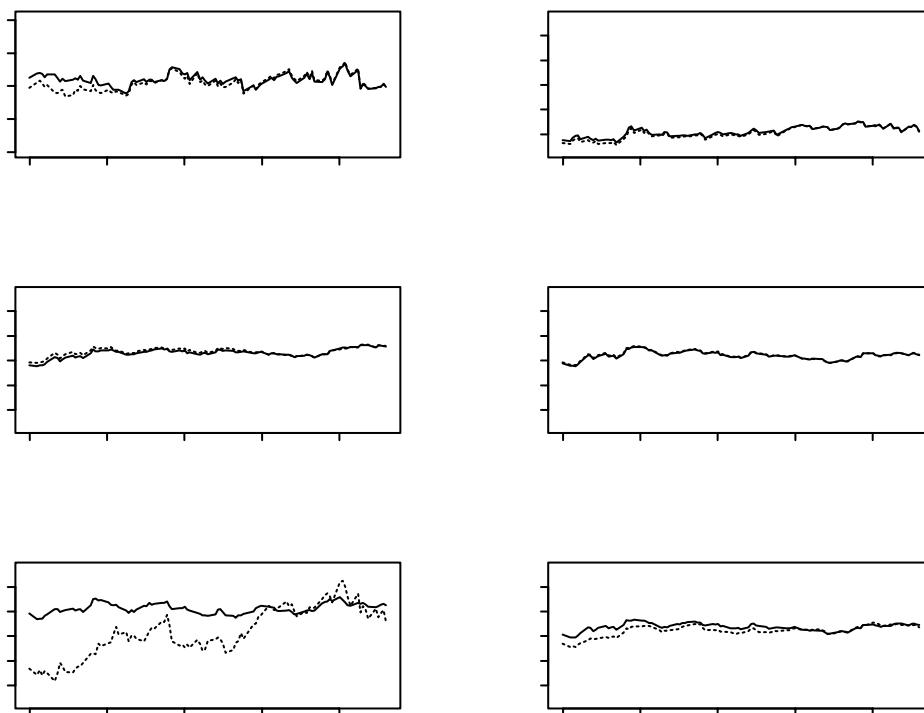


Figure 7: Evolution of various points on the forward-rate curve and on the yield curve during the period June to October 1995 (t is measured in months from 1 January 1995). Results with (solid line) and without (dotted line) irredeemable stocks.

It was generally felt amongst the members of the Fixed Interest Indices Working Party that even 4% Consols and 2.5% Treasury have unreliable prices which are out of line with the rest of the market. On that basis it would be appropriate to either ignore all irredeemable stocks or to specify a higher variance in the price of an irredeemable. The fitting exercise was therefore repeated, with all irredeemable stocks removed from the data. The results of this are plotted in Figure 7.

There are two points to make here. First, it can be seen that when the consols' yield is out of line with that of the more-frequently-traded long-dated stocks (for example, in June 1995) the fit of the model is improved. Furthermore, as is described below removing the irredeemables improves the performance of the model in the Runs Test. Second, we can observe that the long-term forward rate $f(t, t + \infty)$ is less stable when the irredeemable stocks are removed. This

indicates that there is some value added by the irredeemable stocks.

4.2 Other tests of goodness of fit

It has already been pointed out that Root Mean Squared Errors (RMSE) were found to be about 10 basis points (1988-1993 data, significant coupon effect) or 4 to 5 basis points (1995 data, no significant coupon effect) (Figures 4, 5 and 6). This represents a significant improvement over the existing yield-curve model without the necessity of adding further parameters. The RMSE in the case of the new statistical model is, in fact approximately equal to (but larger than) the derived RMSE for the estimated errors in the gross redemption yields. (In fact, if we excluded the $1/3200^2$ from the variance of $\log P_i$ in the current model then the RMSE plotted here would be equal to the RMSE in the fitted yields.)

The Runs Test was also applied to the results to test for independence of the residuals. In this test we consider the residuals $z_i = y_i - \hat{y}_i$, and let n_1 be the number of positive and n_2 be the number of negative residuals. If the estimated curve provides a poor fit then, we may, for example, observe that all long-dated stocks all have yields which are below their estimated values, and the number of runs would be quite low.

As an example, suppose $z = (1.1, 0.6, -0.3, -0.7, -0.2, 0.7, -0.1, 0.6)$. We first translate this into +'s and -'s: that is, $(++-- - + - +)$. A sequence of +'s (minimum length 1) then counts as one run. So then the number of runs here is 5.

If the residuals are independent with $Pr(z_i > 0) = n_1/(n_1 + n_2)$ (that is, the null hypothesis) then we can derive the distribution of the number of runs. If the observed value is *consistently* too low then the model may be considered to have provided a poor fit. If the number of runs calculated each day occasionally looks too low then this is not a problem, since random variation will dictate that this will happen occasionally.

The Runs test was applied to each of the 100 days included in the 1995 data and the results are plotted in Figure 8. The plot gives results for the two cases with (solid line) and without (dotted line) the irredeemable stocks. The plot gives percentage points under the null hypothesis for the observed number of runs using the Normal approximation to the number of runs. The following observations are made:

- As one might expect, there is a high degree of correlation from one day

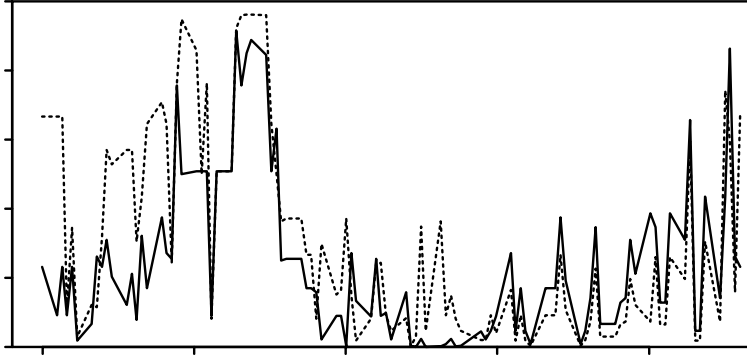


Figure 8: 1995 data. Approximate p-values (calculated separately for each day) for the numbers of runs based upon the null-hypothesis that residuals z_i are independently distributed with $Pr(z_i > 0) = n_1 / (n_1 + n_2)$.

to the next.

- There are periods when the numbers of runs is relatively low. The worst cases occurred when there were 10 runs against an expected number of 18.
- Generally, however, the numbers of runs were not so low as to give serious grounds for concern. The outcome is reasonably acceptable without being too good. With more experience we could find that the situation is better or worse but certainly, because of serial correlations, we have an inadequate set of data at the moment.
- Removing the irredeemables increases the number of runs indicating that their removal is worthwhile.

Residuals were also tested for correlation with the size of the coupon. The observed correlation coefficients for the 1988 to 1993 data were positive and very significant as we would expect. On-the-other-hand when we consider the 1995 data we find the the residuals showed no significant correlation with coupon, demonstrating that the imminent changes to the taxation of gilts has already been factored into the pricing structure of fixed interest bonds.

4.3 Comment on the choice of model

In section 3.5 it was remarked that the root-mean-squared-error with the new model hovers around 4 or 5 basis points. This was achieved with a 5-parameter model (the long-term forward rate plus $m = 4$ exponential terms). It was found that $m = 4$ was necessary to fit the data well on all dates examined. With $m = 2$ the RMSE was much higher at about 10 basis points during 1995. With $m = 3$ the RMSE only a little above that the $m = 4$ on average but there was some evidence that the differences could vary significantly. On some days the differences would be negligible, while on other days $m = 3$ would give a significantly worse fit. Since $m = 4$ was required for a good fit on a limited number of days it was felt that this should be used rather than $m = 3$.

5 Catastrophic Jumps

The new forward-rate model was investigated for the existence of multiple maxima. If more than one maximum exists then clearly there would still be the potential for catastrophic jumps in the future.

Using the new model the logarithms of the prices are not linear functions of the five parameters. If this was the case then we could say for sure that there will always be a single maximum because maximizing the likelihood would be equivalent to maximizing a quadratic function of the five parameters. This would be a result of the switch from a non-linear function to one which is linear in the 5 parameters a and the b_i . Since this is not the case it is difficult to say with certainty that there are no multiple maxima, but it is strongly suspected that this is the case. There are two reasons for why this might be.

First, a number of specific dates were considered on an individual basis. For each date 100 starting points for the optimization were chosen at random. In each of these cases the algorithm converged to the same maximum.

Second, one can argue that multiple maxima are less likely to exist than before on the following grounds. If the market was made up entirely of zero-coupon stocks then the logarithms of the prices would be linear functions of the parameters. Furthermore, if we ignored the prior distribution the the problem would reduce to maximizing a quadratic function. So in this special case it is known that there would be a unique maximum. In the earlier model, even in the zero-coupon bond case the Dobbie and Wilkie (1978, 1979) model would still be non-linear and we know that multiple minima exist in that particular example. As we move (in a continuous sense) away from this ideal position

to one which includes small coupons then, at first, continuity tells us that the likelihood surface will have more-or-less the same shape as before and, in particular, will have only one maximum. As we increase the significance of the coupons further this may cease to be the case. But we can say that at least there will be a unique maximum if there are only low-coupon (in some sense) stocks. What is not clear is whether or not this is true also for markets with higher coupons, or, if it multiple maxima do turn up, is this beyond reasonable coupon rates.

Further investigation will need to be done here.

Acknowledgements

The author would like to thank the various members of the Fixed Interest Indices Working Party. In particular, George Gwilt first suggested the use of the exponential class of forward-rate curves as an alternative to yield curves, while Geoff Chaplin, Keith Feldman and David Wilkie provided useful discussion of the results.

References

- Anderson, N., Breedon, F., Deacon, M., Derry, A. and Murphy, G. (1996) *Estimating and interpreting the yield curve*. Wiley, Chichester.
- Clarkson, R.S. (1979) A mathematical model for the gilt-edged market. *Journal of the Institute of Actuaries* **106**, 85-132.
- Dobbie, G.M. and Wilkie, A.D. (1978) The FT-Actuaries Fixed Interest Indices. *Journal of the Institute of Actuaries* **105**, 15-27.
- Dobbie, G.M. and Wilkie, A.D. (1979) The FT-Actuaries Fixed Interest Indices. *Transactions of the Faculty of Actuaries* **36**, 203-213.
- Heath, D., Jarrow, R. and Morton, A. (1992) Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. *Econometrica* **60**, 77-105.
- McCutcheon, J.J. and Scott, W.F. (1986) *An Introduction to the Mathematics of Finance*. Heinemann, Oxford.
- Nelson, C.R., and Siegel, A.F. (1987) Parsimonious modeling of yield curves. *Journal of Business* **60**, 473-489.
- Poston, T. and Stewart, I.M. (1978) *Catastrophe theory and its applications*. Pitman, London.
- Svensson, L.E.O. (1994) Estimating and interpreting forward interest rates: Sweden 1992-1994. *Working paper of the International Monetary Fund* **94.114**, 33 pages.
- Vasicek, O. (1977) An equilibrium characterization of the term structure. *J. Financial Economics* **5**, 177-188.