Stability of Descriptive Models for the Term Structure of Interest Rates with Application to German Market Data

Andrew J.G. Cairns[1 2] and Delme J. Pritchard
Department of Actuarial Mathematics and Statistics,
Heriot-Watt University,
Edinburgh, EH14 4AS,
United Kingdom
Fax: +44 131 451 3249

Abstract

This paper discusses the use of parametric models for the term structure of interest rates and their uses. The paper focuses initially on a potential problem which arises out of the use of certain models. In most cases the process of parameter estimation involves the minimization or maximization of a function (for example, least squares or maximum likelihood). In some cases this function can have a global minimum/maximum plus one or more local minima/maxima. As we progress through time this leads to a process under which parameter estimates and the fitted term structure can jump about in a way which is inconsistent with bond-price changes.

Here a number of models are identified as susceptible to this sort of problem. However, under one of the descriptive models (the restricted-exponential model) it is proved that the likelihood and Bayesian posterior functions have unique maxima: both in a zero-coupon bond market and in a low-coupon bond market. A counterexample shows that this result can break down for larger-coupon bond markets. An alternative Bayesian estimator in combination with the restricted-exponential model is shown to be free from the problem of catastrophic jumps in all coupon-bond markets.

This model has previously been applied to UK data (Cairns, 1998). Here, we consider its wider application in European bond markets. In particular, German market data from 1996 and 1997 is analysed using the restricted-exponential model. We find that the model gives a good description of the German market during this period.

Keywords: term-structure; multiple maxima; restricted-exponential model; likelihood function; posterior distribution; Bayesian estimator.

[1]E-mail: A.Cairns@ma.hw.ac.uk
[2]WWW: http://www.ma.hw.ac.uk/∼andrewc/

1 Introduction

1.1 Descriptive models

A descriptive model takes a snapshot of the bond market as it is today. Generally, there is no reference to price data from other dates. The sole aim is to get a good description of today's prices: that is, of the rates of interest which are implicit in today's prices.

A descriptive model, on its own, gives us no indication of how the term structure might change in the future. We know that there is randomness in the future but this sort of model does not describe this feature. The description of the dynamics of the term structure falls into the domain of arbitrage-free equilibrium and evolutionary models (for example, see Baxter and Rennie, 1996, Rebonato, 1996, or Jarrow, 1996) or more general actuarial and econometric models which are not necessarily arbitrage free but which pay more attention to past history (for example, see Wilkie, 1995, or Mills, 1993).

Descriptive models have a number of uses:

- They give us a broad picture of market rates of interest which are implied by market prices (and, in particular, if there is only a coupon-bond market) (for example, see Nelson & Siegel, 1987, Svensson, 1994, and Dalquist & Svensson, 1996).
- They can be used to price forward bond contracts.
- They can assist in the analysis of monetary policy (Dalquist & Svensson, 1996).
- A forward-rate curve can be used as part of the input to a model based on the Heath, Jarrow & Morton (1992) framework or the more recent positive interest framework of Flesaker & Hughston (1996). Of course, here, it is also necessary to specify a volatility structure (for example, see Jarrow, 1996). Once this has been added to the descriptive model we have a full model which describes the dynamics of the term structure. The input forward-rate curve is often recalibrated each day.
- They can be used in the construction of yield indices (Feldman *et al.*, 1998).
- Finally descriptive models provide sufficient information for us to get a precise market value of a non-profit insurance portfolio or to price, for example, annuity contracts.

1.2 Parametric models

This paper will concentrate on the use of parametric models.

The alternative to such models is spline graduation (for example, see McCulloch, 1971, 1975, Vasicek & Fong, 1982, Mastronikola, 1991, Deacon & Derry, 1994, Fisher *et al.*, 1995, or Waggoner, 1997). Parametric curves aim to give a parsimonious description of the term structure (Svensson, 1994) providing a broad picture which shows the main features of the term structure. Spline graduations aim to give a detailed and highly-parametrized picture of the market: warts and all. It is questionable, however, if all this detail is accurate. For example, splines can overcompensate for small but genuine differences between observed prices and underlying theoretical prices (for example, due to liquidity problems, small variations in taxation or, simply, the bid/offer spread). Furthermore, the results of a spline graduation seem to be sensitive to the number and the location of the knots (Dalquist & Svensson, 1996, and, for example, see Deacon & Derry, 1994, Figures 3.2 to 3.5). In terms of confidence intervals for rates of interest this is likely to result in a relatively wide band at all maturities (that is, a relatively wide margin of error). Parametric models, on the other hand, display a relatively wide margin of error only at the very short and the very long end of the maturity spectrum (Cairns, 1998). At all other maturities estimates of par yields, spot rates and forward rates are all relatively robust. Lorimier (1995) considered spline graduation which aimed to find the best curve in the sense of fitting prices perfectly while at the same time being as smooth as possible.

With parametric curves, the aim is to get a parsimonious description of the term structure: that is, we wish to use a model which captures as much of the detail of the market as possible with as few parameters as possible. These two requirements clearly conflict: it is always possible to improve the fit of the model by increasing the number of parameters. For an additional parameter to be worthwhile the improvement in fit must exceed a specified amount (for example, according to the Schwarz-Bayes Criterion discussed in Wei, 1990, and Cairns, 1995, 1998). Such curves avoid the lumpiness of spline models. However, they can still be biased at certain maturities if there are small heterogeneities in the market not accounted for in the model and if bonds with certain characteristics are clustered. For example, in the UK there is a cluster of strippable gilts at the long end of the market.

1.3 Existing parametric models

1.3.1 Gross redemption yields

Dobbie & Wilkie (1978) proposed the model which is currently used in the construction of the UK yield indices published in the Financial Times. It is a model for gross redemption yields: that is,

$$y(t,t+s) = b_0 + b_1 e^{-c_1 s} + b_2 e^{-c_2 s}$$

is the gross-redemption-yield curve at time t for a coupon bond maturing at time $t+s$. This curve is fitted to low, medium, and high-coupon bands separately to take account of the old, UK coupon effect. Since 1996, however, income and capital gains on UK gilts have been taxed on the same basis making this yield curve approach obsolete and forward-rate curves more relevant.

1.3.2 Forward-rate curves

Suppose that $f(t,t+s)$ is the instantaneous forward rate at time t for payments at time $t+s$.

Nelson & Siegel (1987) proposed the following curve:

$$f(t,t+s) = b_0 + (b_{10} + b_{11}s)e^{-c_1 s}.$$

The curve is of the form of a constant *plus* a polynomial-times-exponential term. It allows for a single hump or dip in the curve.

Svensson (1994) generalised this by adding a further polynomial-times-exponential term:

$$f(t,t+s) = b_0 + (b_{10} + b_{11}s)e^{-c_1 s} + b_{21}se^{-c_2 s}.$$

This curve can have up to two turning points.

Wiseman (1994) proposed another model of exponential type:

$$f(t,t+s) = b_0 + b_1 e^{-c_1 s} + \ldots + b_n e^{-c_n s}.$$

The order of the model n varies from one country to the next. The curve can have up to $n-1$ turning points.

Bjørk & Christensen (1997) generalised all of the previous forms of forward-rate curves by describing the *exponential-polynomial* class of curves:

$$f(t,t+s) = L_0(s) + \sum_{i=1}^{n} L_i(s)e^{-c_i s}$$

where each of $L_0(s), \ldots, L_n(s)$ is a polynomial in s:

$$L_i(s) = b_{i0} + b_{i1}s + \ldots + b_{ik_i}s^{k_i}.$$

In the *unrestricted* case all parameters (the b_{ij} and the c_i) are estimated. This is the case, for example, in Nelson & Siegel (1987), Svensson (1994) and Wiseman (1994).

1.4 Estimation

The parameters in a descriptive model are fitted by taking, first, a snapshot of the data. For example, this may give us a set of price data, P. For a given model, let ϕ be the set of parameters. For a given model and value for ϕ we have, for each i, a theoretical price $P_i(\phi)$ in addition to the observed price P_i. ϕ can be estimated by a number of means: for example, weighted least squares (Dobbie & Wilkie, 1978, Wiseman, 1994); maximum likelihood (Cairns, 1998); or Bayesian methods (Cairns, 1998). Least squares methods can be demonstrated to have a sound statistical basis (Cairns, 1998) but only the likelihood and Bayesian approaches can give a complete picture of the results. In particular, they give not only parameter estimates but also an indication of the level of parameter uncertainty and of the level of uncertainty in the estimates of various interest rates.

Let us take a specific example. Suppose that we are considering a zero-coupon bond market and that we wish to fit the Nelson & Siegel (1987) model using maximum likelihood.

For maximum likelihood we must specify a full statistical model. Here we assume for simplicity that the logarithm of the price of a zero-coupon bond has a Normal distribution with mean $\log \hat{P}_i$ and with a variance which depends upon the term to maturity:

$$\begin{aligned} \log P_i &\sim N\left(\log \hat{P}_i(\phi),\ \sigma^2(t_i)\right) \\ \text{where } t_i &= \text{term of stock } i. \end{aligned}$$

It is not necessary for us to specify the form of $\sigma^2(t)$ at this point. For possible definitions see Svensson (1994), Wiseman (1994) or Cairns (1998) or later in this paper (Section 4).

Now the form of the forward-rate curve in the Nelson & Siegel (1987) model means that $-\log \hat{P}_i(\phi)$ has the following simple form:

$$-\log \hat{P}_i(c_1, b) = b_0.t_i + b_{10}.\frac{(1-e^{-c_1 t_i})}{c_1} + b_{11}.\frac{1}{c_1^2}\left(1-(1+c_1 t_i)e^{-c_1 t_i}\right)$$

There are three components to this formula. Each component is of the form of a linear b-coefficient times a non-linear term involving t_i and c_1.

Because of this linearity in b it is easy to estimate $b = (b_0, b_{10}, b_{11})$ for this statistical model given a specific value of c_1.

We call the b's linear parameters and c_1 a non-linear parameter.

If we go back to the log-likelihood function this is a function of c_1 and b given P

$$l(c_1, b; P) = -\frac{1}{2}\sum_{i=1}^{n}\left\{\log[2\pi\sigma^2(t_i)] + \frac{(\log P_i - \log \hat{P}_i(c_1, b))^2}{\sigma^2(t_i)}\right\}$$

and we have to maximise this over c_1 and b. This function is quadratic in b so that, for a given value of c_1, the estimate $\hat{b}(c_1)$ is unique. For simplicity we write $\hat{l}(c_1)$ when the function has been maximised over b. $\hat{l}(c_1)$ is the profile log likelihood.

We then have to maximise this over c_1. This function has to be maximised numerically. Sometimes we have a problem with this because $\hat{l}$ occasionally can have more than one maximum: say one global maximum and another local maximum.

The same problem arises within a coupon-bond market. However, the so-called linear parameters, $\hat{b}(c_1)$, are no longer simple to estimate. Generally, $\hat{b}(c_1)$ is uniquely defined, while $\hat{l}(c_1)$ can still have more than one maximum.

In Figure 1 we give an example of this from the close of business on 31 May, 1995, in the UK coupon-bond market. The function $\hat{l}$ has a global maximum at about $c_1 = 0.6$ but it also has a local maximum at about $c_1 = 2.7$. The difference in likelihoods is not too large implying that the local maximum gives almost as good a fit as the global maximum.

1.5 Instability of parameter estimates

Let us think now about the consequences of having more than one maximum as we move through time.

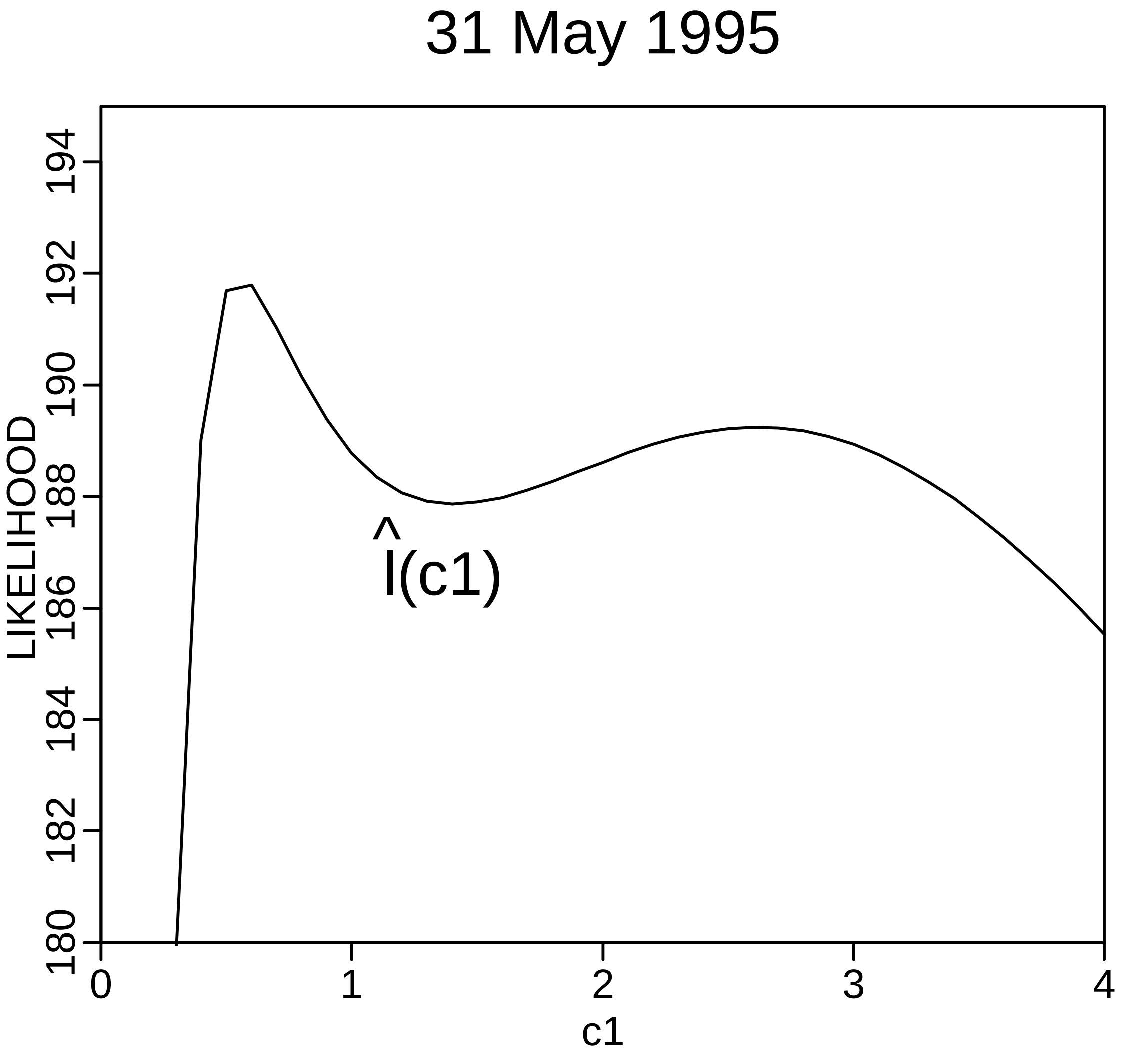


Figure 1: 31 May, 1995. UK coupon bond market. Profile log-likelihood function $\hat{l}(c_1)$ for the Nelson & Siegel model.

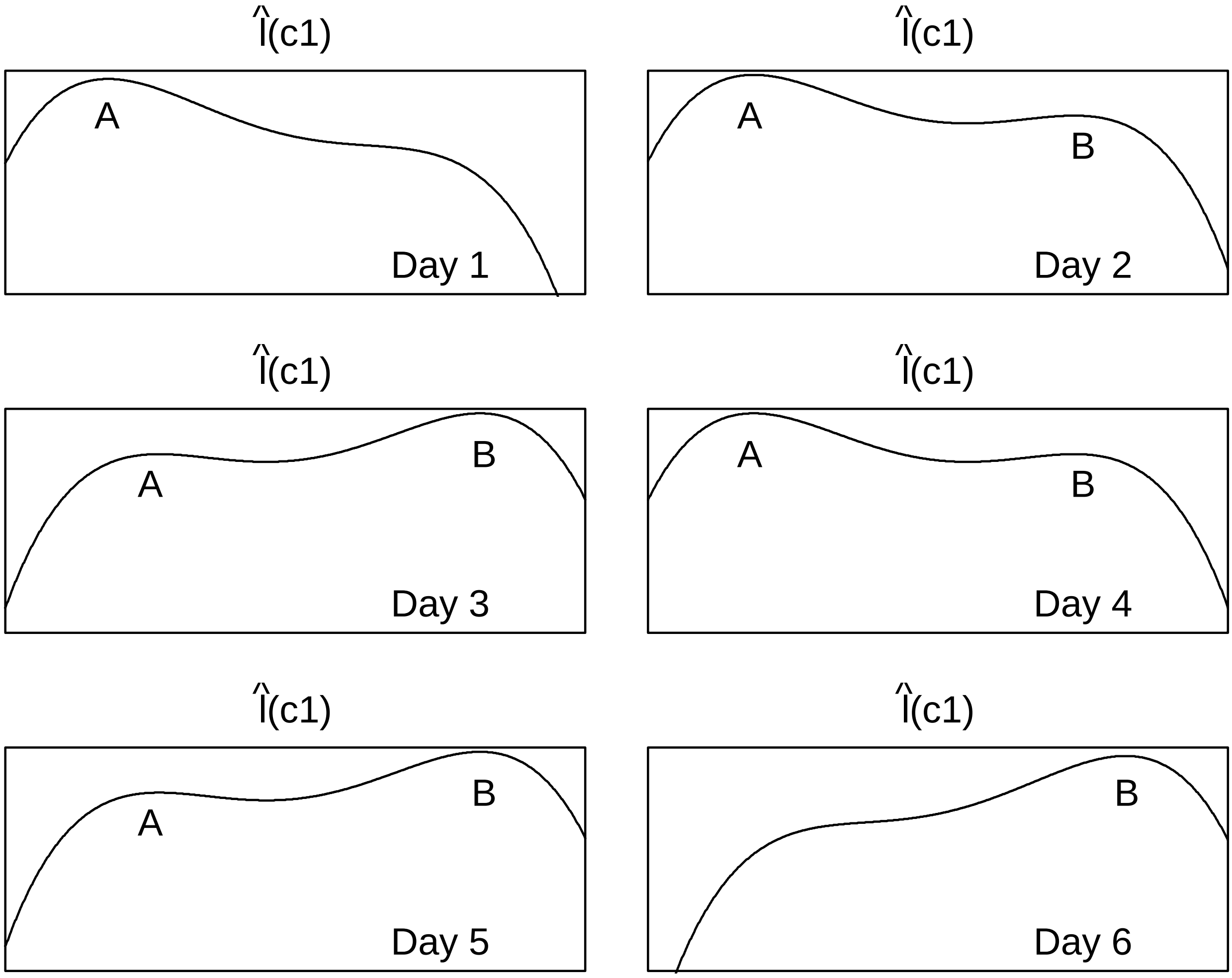


Figure 2: Possible development of the profile log-likelihood function through time.

Figure 2 gives a simple scenario of what might happen. As we go from one day to the next, the likelihood curve will gradually evolve. In particular, the two maxima will go up and down. From time to time one of them might disappear totally and from time to time the global maximum might jump from one location to the other as we see here. We start on day 1 with a unique maximum at A. Suppose also that the maximisation algorithm starts at yesterdays maximum and finds the local maximum. From days 2 to 5 there is a second maximum at B and indeed on days 3 and 5 this is the global maximum. The algorithm will continue until day 5, however, at maximum A. It is not until day 6 when the maximum at A disappears totally that the algorithm moves across to B. Other algorithms might jump more frequently, in particular if they are designed to find the global maximum.

What are the consequences of this problem?

We have identified that as we move from one day to the next the location of the maximum might jump. This is sometimes referred to as a *catastrophic* jump. When such a jump occurs, the size of the jump will typically be much larger than would be consistent with the corresponding changes in prices. For example, if prices follow a diffusion process then the parameter estimates should also follow a diffusion process and in particular should be a continuous process: this continuity will clearly be violated if there is a catastrophic jump.

If parameter estimates jump then a published yield index will also jump in an equally obvious way and the indices will start to lack credibility and fall into disuse. Equally if the curve is used as input to a Heath-Jarrow-Morton model with frequent recalibration it is essential that the recalibrated curves evolve in a way which is consistent with price changes. This is not the case if catastrophic jumps occur which will cause unexpected jumps, for example, in derivative prices.

The existence of more than one maximum can also lead to potential mispricing of such things as bonds, interest-rate derivatives or the pricing of annuities or other life insurance contracts.

All of the Dobbie & Wilkie (1978), Svensson (1994) and Wiseman (1994) models exhibit the same problem with multiple maxima (for example, see Cairns, 1998). In particular, this problem can arise in a zero-coupon bond market as well as in a coupon-bond market. In each case estimates for the linear, polynomial coefficients are unique and simple to derive while the multiple maxima show up in their profile log-likelihood functions. The problem arises on different dates, however, for different models and with varying degrees of magnitude.

In Figure 3 we return to the previous example. Here we have plotted forward-rate, spot-rate and par-yield curves for each of the two maxima in Figure 1. The largest differences occur between the two forward-rate curves. This is because these rates are the furthest from what we actually observe: which is coupon-bond prices. Par yields are closest to what we see on the market so the errors are smallest. The maximum difference here is about only 0.03% which does not sound very much. But if we take the issue of a long-dated stock with a duration, say, of 10 years, then this leads to an error of £3 million per £1 billion issued, which is not trivial. On other dates and for other models these errors can be bigger.

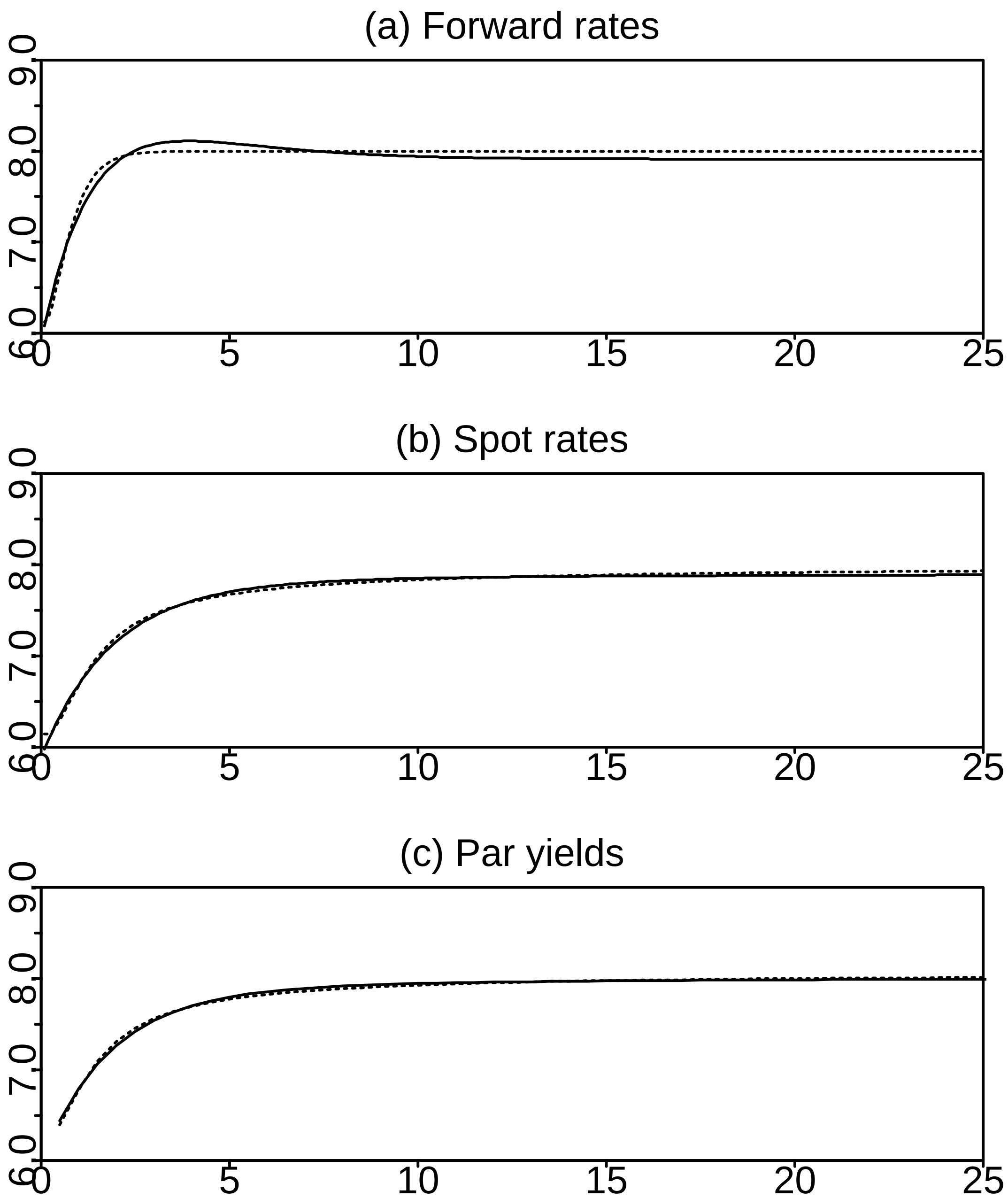


Figure 3: Forward rates, spot rates and par yields for the global maximum (solid line) and for the local maximum (dotted line).

2 The restricted-exponential family

Given the shortcomings of the models described in the previous section, an alternative model is therefore appropriate. Here we describe the the restricted-exponential family proposed by Cairns (1998) (see also Björk & Christensen, 1997). This is a simple family of forward-rate curves:

$$\begin{aligned} f(t,t+s) &= b_0 + b_1 e^{-c_1 s} + \ldots + b_m e^{-c_m s} \\ \phi &= (b,c) \end{aligned}$$

The curve is a sum of a constant plus m constant-times-exponential terms and is superficially the same as the Wiseman (1994) model. The parameter set ϕ is divided up into subsets of linear terms $b = (b_0, \ldots, b_m)$ and non-linear terms $c = (c_1, \ldots, c_m)$. This suggests that there is likely to be the same problem as before. However, the approach taken here is a bit different.

In the previous models we estimated all of the parameters: linear and non-linear. In contrast, here we only estimate a subset of the parameter set. Thus, we fix the exponential parameters c at the outset and at no future point do we estimate their values. At any point in time we only estimate the linear parameters b.

For consistency with the nomenclature in Section 1.3.2 we call this family of curves the *restricted-exponential* class.

2.1 Maximum-likelihood estimation

Proposition 2.1

Under the statistical model proposed in Section 1.4:

(a) In a zero-coupon bond market the resulting maximum-likelihood estimate $\hat{b}$ is unique.

(b) In a low-coupon bond market the log-likelihood function is concave within the required region containing all possible maxima. Hence, the maximum-likelihood estimate $\hat{b}$ is unique.

Proof: See Appendix A. In particular, we define what we mean by a low-coupon bond market.

(Lorimier (1995, Theorem 4.2) proved a similar result. The differences when compared with Proposition 2.1 are:

- the form of the forward-rate curve. Here it is a sum of exponentials, whereas Lorimier uses splines.
- the form of the objective function. Here the aim is to get as good a fit as possible with relatively few parameters. Lorimier aims to get as smooth a set of prices as possible given a perfect fit with enough knots.

These differences require a different style of proof but the flavour of the result is the same: that uniqueness holds for small coupons but it is not guaranteed for larger coupons.)

2.1.1 A counterexample for larger coupons

The likelihood and Bayesian posterior-density functions have been shown to have a unique maximum in a low-coupon bond market. Unfortunately, the result does not extend, at least theoretically, to markets with higher coupons as the following simple counterexample shows.

Suppose that we take a very simple case where $m = 1$: that is, $f(t,t+s) = b_0 + b_1 \exp(-c_1 s)$. Our market consists of two stocks:

Stock 1: annual coupon, rate g, term t_1 to maturity.

Stock 2: zero-coupon, term t_2 to maturity.

Stock 1 has an actual price of P_1 and a theoretical price given b_0, b_1 of $\hat{P}_1(b_0,b_1)$. Let L_1 be the set $\{(b_0,b_1) : \hat{P}_1(b_0,b_1) = P_1\}$. This is a downward sloping and *convex* curve (see Appendix A).

Stock 2 has an actual price of P_2 and a theoretical price of $\hat{P}_2(b_0,b_1)$. Let L_2 be the set $\{(b_0,b_1) : \hat{P}_2(b_0,b_1) = P_2\}$. L_2 is a straight line with a negative gradient.

Given the details of stock 1 it is possible to choose P_2 and t_2 such that L_1 and L_2 intersect in two points within the feasible region $B = \{b : f(t,t+s) \geq 0$ for all $s \geq 0\}$. (With $m = 2$, $B = \{b : b_0 \geq 0, b_0 + b_1 \geq 0\}$.) Let the points of intersection be (b_{10}, b_{b11}) and (b_{20}, b_{b21}). At each point the theoretical prices equal the observed prices so clearly the likelihood function will be maximised at both points of intersection. These maxima will also, of course, be of the same height.

Numerical example:

Suppose that $c_1 = 0.2$. For stock 1 we have $P_1 = 1$, $t_1 = 20$ and $g = 0.08$, and for stock 2 we have $P_2 = 0.352478$, $t_2 = 13.3562$.

There are two solutions in $b = (b_0, b_1)$: $b = (0.03, 0.137958)$ and $b = (0.11, -0.091620)$.

2.1.2 Actual experience

The experience of the UK gilts market suggests, however, that there is no problem with multiple maxima. The counterexample above just shows that we cannot rule out the possibility altogether. One reason for this is that the problem diminishes as the number of stocks increases due to the large sample properties of the likelihood function (for example, see Silvey, 1970). Thus, even if secondary maxima persist as the number of stocks increases, the global maximum will tend to the true parameter set (if the underlying model is the same as the one being considered here). If there were significant differences between the true forward-rate curve and the best-fitting restricted-exponential curve then secondary maxima could, in theory, persist. However, this is not what we see in practice.

2.2 Bayesian estimation

2.2.1 Maximum-posterior-density estimation

Suppose instead we wish to use Bayesian methods with a prior distribution $g(b)$ for b and a 0-1 loss function. Then the log-posterior density function is $g(b|P) = g(b) + l(b;P) +$ constant, and the Bayesian estimator is the mode of the posterior distribution. (The 0-1 loss function thus gives an estimator which gives the best fit consistent with the prior distribution.)

There are two principal reasons for using Bayesian methods:

- We have introduced a constraint that the forward-rate curve should be positive at all maturities, on the basis that the risk-free rate of interest, $r(t)$, will be non-negative at all times, t. If $f(t,t+s)$ is equal to zero for any value of s then this means that $r(t+s) = 0$ with probability 1. If we wish to exclude this possibility then we must require that $f(t,t+s)$ is strictly positive for all s. It is unreasonable to require that $f(t,t+s)$ has a minimum value higher than 0. However, if we use maximum likelihood and this finds that the maximum is at some b for which the forward-rate curve is negative at some maturities then the introduction of a constraint will mean that the forward-rate curve is still equal to 0 for some s. This problem can be avoided if we use Bayesian methods. In particular, if the prior density function, $g(b)$, tends to zero on the boundary of the feasible region (in which $f(t,t+s)$ remains positive) then the maximum of the posterior will give a strictly positive forward-rate curve at all maturities.

- Bayesian methods provide a coherent framework within which we can analyse parameter risk and construct confidence intervals for specified interest rates and so on.

Corollary 2.2

If the log-prior distribution function is concave then:

(a) In a zero-coupon bond market the resulting Bayesian estimate $\hat{b}$ is unique.

(b) In a low-coupon bond market the log-posterior density function is concave within required region containing all possible maxima. Hence, the Bayesian estimate $\hat{b}$ is unique.

2.2.2 Squared-error loss functions

Suppose instead that the loss function is of the form

$$L(b,\tilde{b}) = \int_0^\infty \nu(s)\{f(t,t+s;b) - f(t,t+s;\tilde{b})\}^2 ds,$$

where b is random with density equal to the posterior density function, $\nu(s)$ defines the weight attached to duration s, $\int_0^\infty \nu(s)ds < \infty$, $f(t,t+s;b) = b^T d'(s)$, and $d'(s)^T =$

$(1,\ \exp(-c_1 s),\ \ldots,\ \exp(-c_m s)\ )$. The best estimator is $\hat{b}$ which maximizes the posterior expectation of the loss function. Thus:

$$E[L(b,\hat{b})|P] = \inf_{\tilde{b}} E[L(b,\tilde{b})|P].$$

Now

$$\begin{aligned}
e(\tilde{b}) = E[L(b,\tilde{b})|P] &= E\left[\int_0^\infty \nu(s)\left(b^T d'(s) - \tilde{b}^T d'(s)\right)^2 ds\ |P\right] \\
&= E\left[\int_0^\infty \nu(s)(b-\tilde{b})^T d'(s)d'(s)^T(b-\tilde{b})ds\ |P\right] \\
\Rightarrow \frac{\partial e}{\partial \tilde{b}}(\tilde{b}) &= E\left[2\int_0^\infty \nu(s)d'(s)d'(s)^T(b-\tilde{b})ds\ |P\right] \\
&= 2\int_0^\infty \nu(s)d'(s)d'(s)^T ds\, E[b-\tilde{b}\ |P].
\end{aligned}$$

Thus $e(\tilde{b})$ is minimized at $\hat{b} = E[b|P]$. This estimator is well defined relative to the problem of maximizing a function with more than one maximum. The estimator will also evolve without the risk of catastrophic jumps, since the form of the posterior distribution evolves in a way which is consistent with price changes. Interestingly, this estimate does not depend upon the form of $\nu(s)$.

If, on the other hand, we are considering one of the models in the unrestricted exponential-polynomial class we have the same problems of non-linearity in the exponential parameters. This arises from the fact that the minimization problem here is essentially the same as the maximization problem described in Sections 1.4 and 1.5 for a zero-coupon bond market.

Clearly any loss function which is quadratic in b and $\tilde{b}$ will also have the same properties: for example, if we replace the forward-rate curve by the spot-rate curve.

3 Further remarks

3.1 Choice of m

To get a consistently good fit in the UK gilts market we require $m = 4$: that is, 4 exponential terms. Inevitably we require more terms than if we estimate both the linear and non-linear parameters. The new approach with 4 exponential terms is roughly equivalent to estimating b and c in a model with only 2 exponential terms (but in each case we are still only estimating 5 parameter values). However, the restricted-exponential model has the advantage that it will fit much better on dates where more than one turning point in the forward-rate curve is apparent.

With $m = 4$ we can have a very rich or wide range of yield curves with up to 3 turning points.

3.2 A more general family of curves

Bjørk and Christensen (1997) consider what families of curve are consistent with the evolution of certain models for the term structure. They describe a more general class of model: the restricted-exponential-polynomial family. Any curve in this family in which the polynomials are all of degree 0 is in the restricted-exponential family. A simple example is the Vasicek (1977) model. This is a one-factor model under which the forward-rate curve evolves within the family of curves $\{f(t,t+s) = b_0 + b_1\exp(-c_1 s) + b_2\exp(-c_2 s)\}_{b_0,b_1,b_2}$ where c_1 and $c_2 = 2c_1$ are fixed. There are further restrictions on the parameters b_0, b_1 and b_2 (since we are working with a one-factor, Markov model) but the curve does evolve within this higher-dimensional family.

4 Analysis of German bond-price data

In this section we will consider the application of the restricted-exponential model to the German bond market. The data relate to the period 4 January 1996 to 12 April 1997 with daily prices and between 85 and 90 stocks quoted on each date. An analysis covering a longer period would be desirable. However, the present analysis was severely hampered by the lack of high quality, raw price data covering longer periods. Despite this, the present analysis comes to a number of qualitative statements which will hold for other time periods. However, the more general use of the model (for example, in the construction of yield indices) would benefit from analysis of a longer series of data.

We fit the following model for the forward rate curve:

$$f(t,t+s) = b_0 + b_1 e^{-c_1 s} + b_2 e^{-c_2 s} + b_3 e^{-c_3 s} + b_4 e^{-c_4 s}$$

as in Cairns (1998). For the vector $c = (c_1, c_2, c_3, c_4)$ we assume the central set of values $(0.1, 0.2, 0.4, 0.8)$ (one of those considered in the previous paper). Later in this section we will consider other choices for c.

Recall also that we will use the following statistical model for coupon bond prices:

$$\log P_i \sim N\left(\log \hat{P}_i(\phi), \sigma^2(P_i, d_i)\right)$$

where d_i is the duration of stock i and price errors given ϕ are assumed to be independent. $\hat{P}_i(\phi)$ is the discounted value of the coupon and principal payments calculated with reference to the forward-rate curve when the parameter set is equal to ϕ. The error structure is the same as that in Cairns (1998): that is:

$$\begin{aligned}
\sigma^2(p,d) &= \frac{\sigma_0^2(p)\left(\sigma_\infty^2 d^2 b(p) + 1\right)}{\sigma_0^2(p) d^2 b(p) + 1} \qquad (1)\\
\text{where } \sigma_0^2(p) &= (100p)^{-2}\\
b(p) &= \frac{\sigma_d^2}{\sigma_0^2(p)\left(\sigma_\infty^2 - \sigma_0^2(p)\right)}\\
\sigma_d &= 0.0004\\
\sigma_\infty &= 0.001
\end{aligned}$$

This means that:

$$\lim_{d\to 0}\sqrt{VarP_i} \approx P_i/100P_i = 0.01$$

(which arises from rounding errors in prices)

$$\lim_{d\to 0}\frac{\partial\sigma^2(p,d)}{\partial(d^2)} = \sigma_d^2$$

(that is, for small d, the standard deviation of gross redemption yields is approximately 4 basis points)

$$\lim_{d\to\infty}\sigma^2(p,d) = \sigma_\infty^2$$

(that is, the standard deviation of long-dated stock prices is limited to about 0.1 per 100*DM* nominal).

4.1 Removal of unusual stocks

Group 1: (outliers)

The first stage of the analysis involved the removal of those stocks which had prices significantly out of line with neighbouring bonds. Such stocks were identified by features such as:

- relatively high volatility;
- long periods of significant cheapness or dearness;
- significant step jumps in prices out of line with the rest of the market.

Such features can arise because of the relatively small size of an issue, special tax status, or implicit option characteristics amongst other reasons.

Some of these stocks are illustrated in Figure 4. For example, take the bond $6\frac{1}{2}\%$ 2/1997:

- initially this stock is well in line with the market as a whole with y-values or price errors (actual minus fitted price per 100*DM* nominal) close to zero;
- later on though, the stock becomes, relatively, very expensive.

Figure 4 also shows up an interesting comparison between two very similar stocks $6\frac{3}{8}\%$ 8/1998 and $6\frac{3}{4}\%$ 8/1998. The former is in line (price errors close to zero) while the former is clearly very cheap. We therefore chose to exclude $6\frac{3}{4}\%$ 8/1998. From the graph we can also see that $4\frac{3}{4}\%$ 11/2001 became very expensive shortly after its introduction.

Two further groups of stocks required more careful thought.

Group 2

First we considered a group of stocks all maturing during 1999. ($6\frac{1}{2}\%$ 1/1999, $6\frac{3}{4}\%$ 1/1999, 7% 2/1999, 7% 4/1999, $6\frac{3}{4}\%$ 6/1999, 7% 9/1999, 7% 10/1999 and $7\frac{1}{8}\%$ 12/1999). During the first half of 1996 these stocks exhibited significantly higher volatility and were

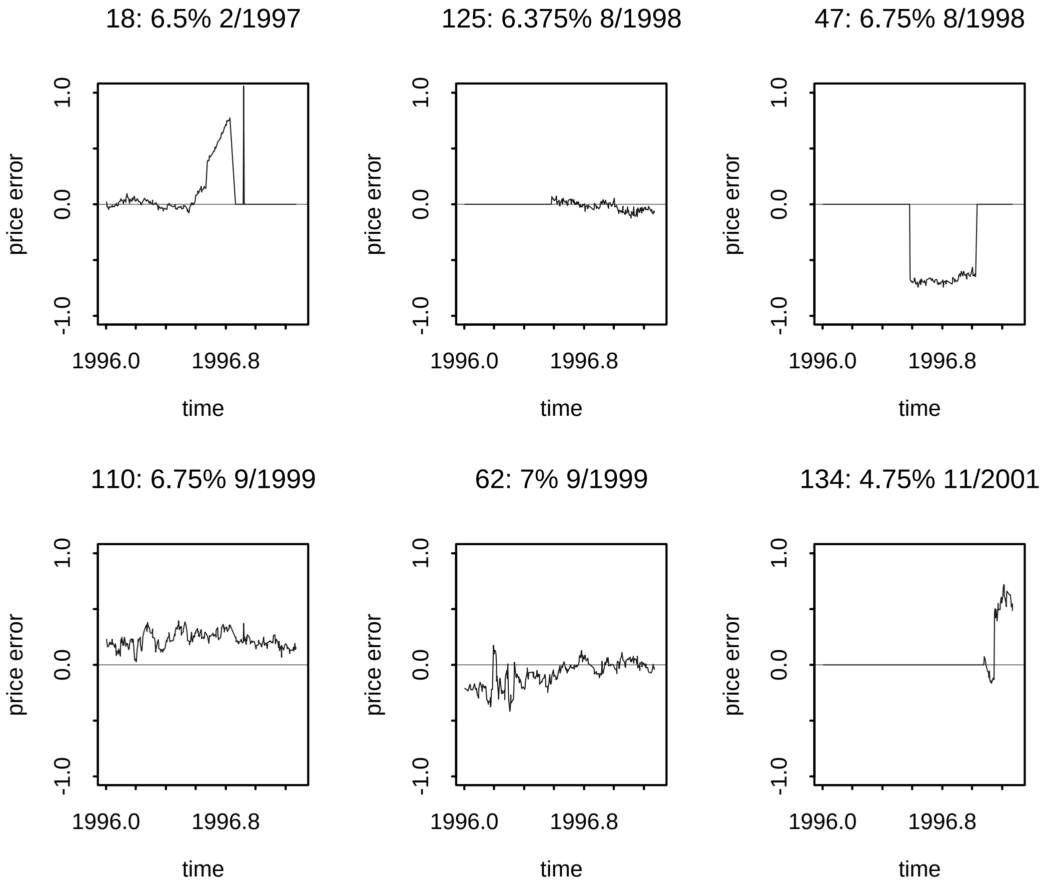


Figure 4: Price errors (actual price minus fitted price per 100*DM* nominal) for selected stocks.

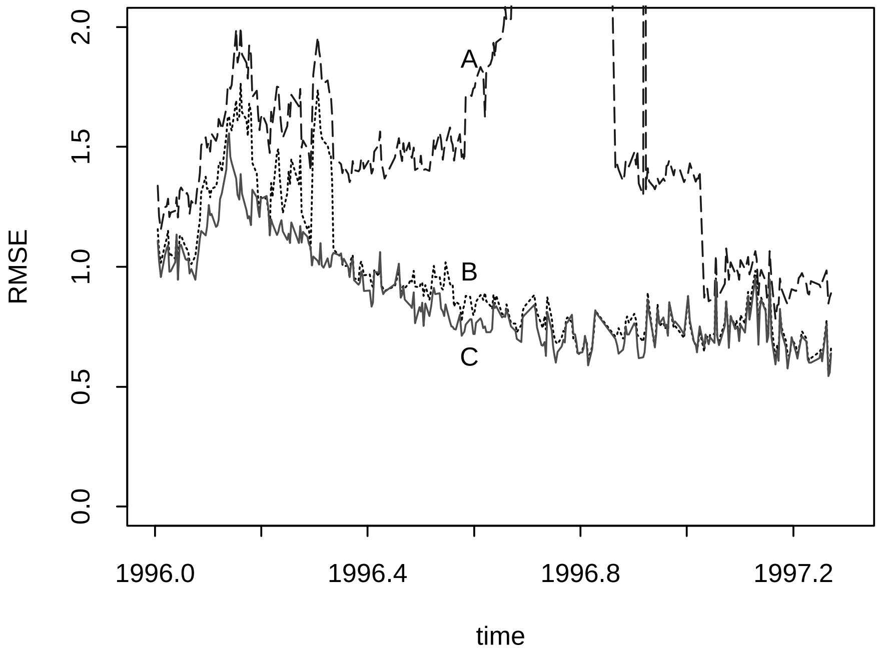


Figure 5: Development of the Root Mean Squared Error (RMSE) over time. A (dashed line): all stocks. B (dotted line): excluding outliers and Group 3. C (solid line): excluding Groups 1, 2 and 3.

significantly cheaper than the remaining (eleven) 1999 stocks (with no obvious coupon effect). Figure 4 illustrates this point (7% 9/1999 versus $6\frac{3}{4}$% 9/1999).

Group 3

Second, five stocks ($5\frac{1}{8}$% 11/2000, $5\frac{1}{4}$% 2/2001, 5% 5/2001, 5% 8/2001 and 6% 9/2003) exhibited stable but very expensive prices relative to the rest of the market. For the bulk of this analysis these stocks were excluded. However, they do lend support to the suggestion of a small coupon effect similar to that which existed in the UK for many years. In particular, other stocks with similar maturities and with gross redemption yields more in line with other market yields had higher coupon rates.

Figure 5 shows the effect of removing various stocks on the quality of fit. If the Root Mean Squared Error (RMSE) is close to 1, the error structure, as a whole, in equation (1) is reasonable (though the dependence upon duration may or may not be quite right). If the RMSE is too high then the quality of fit is less good. In Figure 5 we show three sets of results:

A: all stocks are included;

B: Groups 1 and 3 are excluded;

C: Groups 1, 2 and 3 are excluded.

We can see from Figure 5 that removal of the outliers makes a very significant improvement. Further removal of the volatile and cheap stocks in Group 2 also makes a significant difference in the first 4 months of 1996 but not after.

In Figure 6 we compare results where:

A: Groups 1, 2 and 3 are excluded;

B: Groups 1 and 2 are excluded.

We can see from the top graph that the quality of fit improves significantly. In Figure 6 (bottom) we plot observed gross redemption yields for the stocks against term to maturity (black dots). The four crosses represent those stocks in Group 3. Superimposed on these points are the two par yield curves implied by the parametric models fitted to sets A and B respectively. We can see that the removal of Group 3 stocks makes very little difference to the fitted curve (in fact the two curves, which differ by at most 2 basis points, lie almost on top of each other). This, of course, is a consequence of these stocks lying in the middle of the range of terms to maturity.

The remainder of this analysis assumes that Groups 1, 2 and 3 of stocks are excluded. In Figures 7 and 8 we show the development of price errors for selected stocks. In Figure 7 the first two rows give results for typical short-dated stocks. These all have errors of small magnitude and low volatility reflecting the short duration of these stocks. The third and fourth rows allow comparison of pairs of stocks with similar maturity dates but different coupons. This again suggests a small coupon effect.

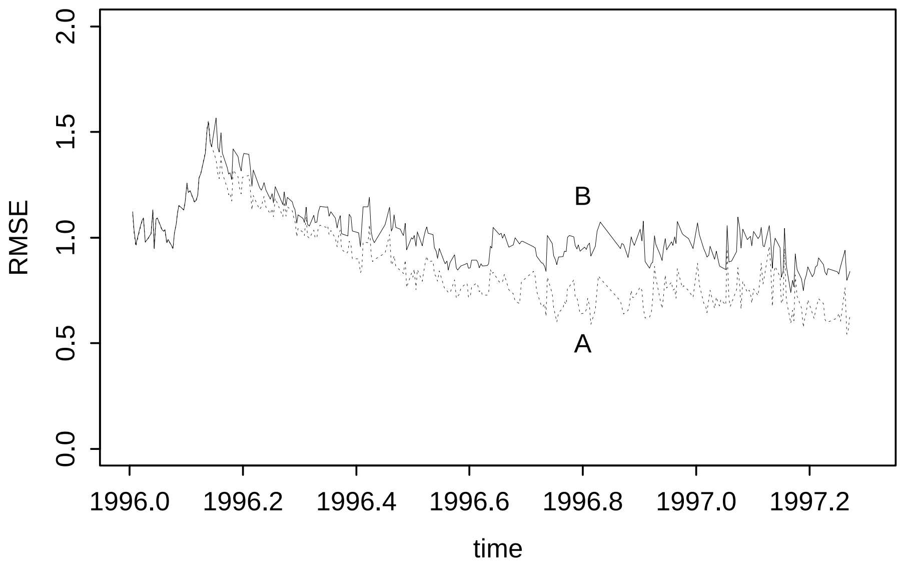
In Figure 8 we give price errors for longer-dated stocks. In each graph the four vertical lines correspond to the introduction of four new stocks (6% 1/2006, 6% 2/2006, $6\frac{1}{4}$% 4/2006 and 6% 1/2007).

The German bond market is really divided up into two parts:

- 6% 2016 and $6\frac{1}{4}$% 2024 (now supplemented by $4\frac{3}{4}$% 2028);
- stocks with a term to maturity of strictly less than 10 years.

Thus we have a very sparse market in long-dated bonds (greater than 10 years) and a relatively dense market in bonds with a term of less than 10 years.

In the UK the range of maturity dates is more uniformly spread out with a premium attached to the longest-dated stock (6% 2028). In the German market the sparseness of the long-dated market makes this difficult to identify. However, Figure 8 demonstrates that there is a clear premium attached to the two or three longest-dated bonds in the under-ten-year class. For example, we can observe in Figure 8 that 6% 1/2006 is as much as 1*DM* overpriced relative to the market early in 1996 when it is the longest-dated stock under 10 years. This price premium gradually declines as longer-dated stocks (6% 2/2006, $6\frac{1}{4}$% 4/2006 and 6% 1/2007) are first anticipated and then introduced. The same pattern can be seen with other stocks.



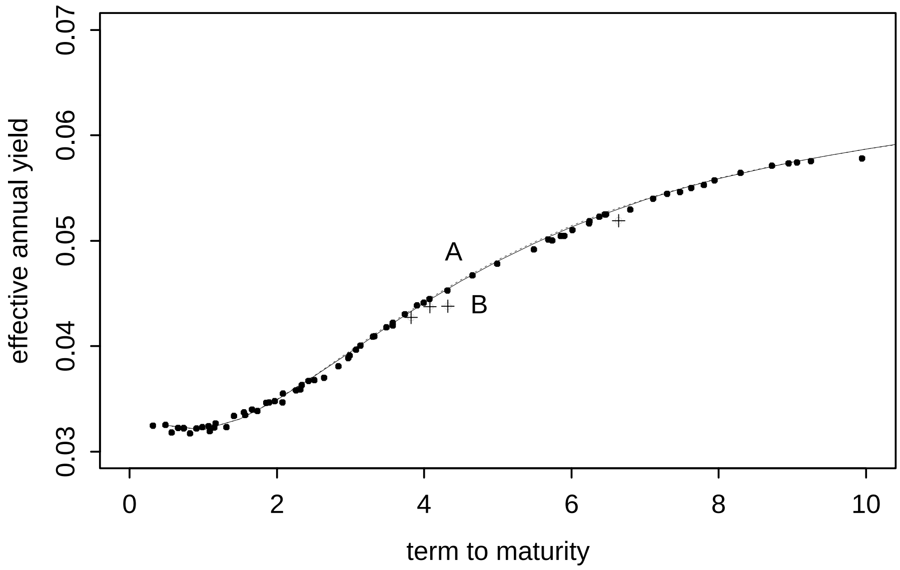


Figure 6: The effect of removing expensive low-coupon stocks. A (dotted line): Groups 1, 2 and 3 excluded. B (solid line): Groups 1 and 2 excluded. Lower graph: observed gross redemption yields on 24 January 1997 (dots) (crosses for Group 3 stocks) and fitted par-yield curves A and B.

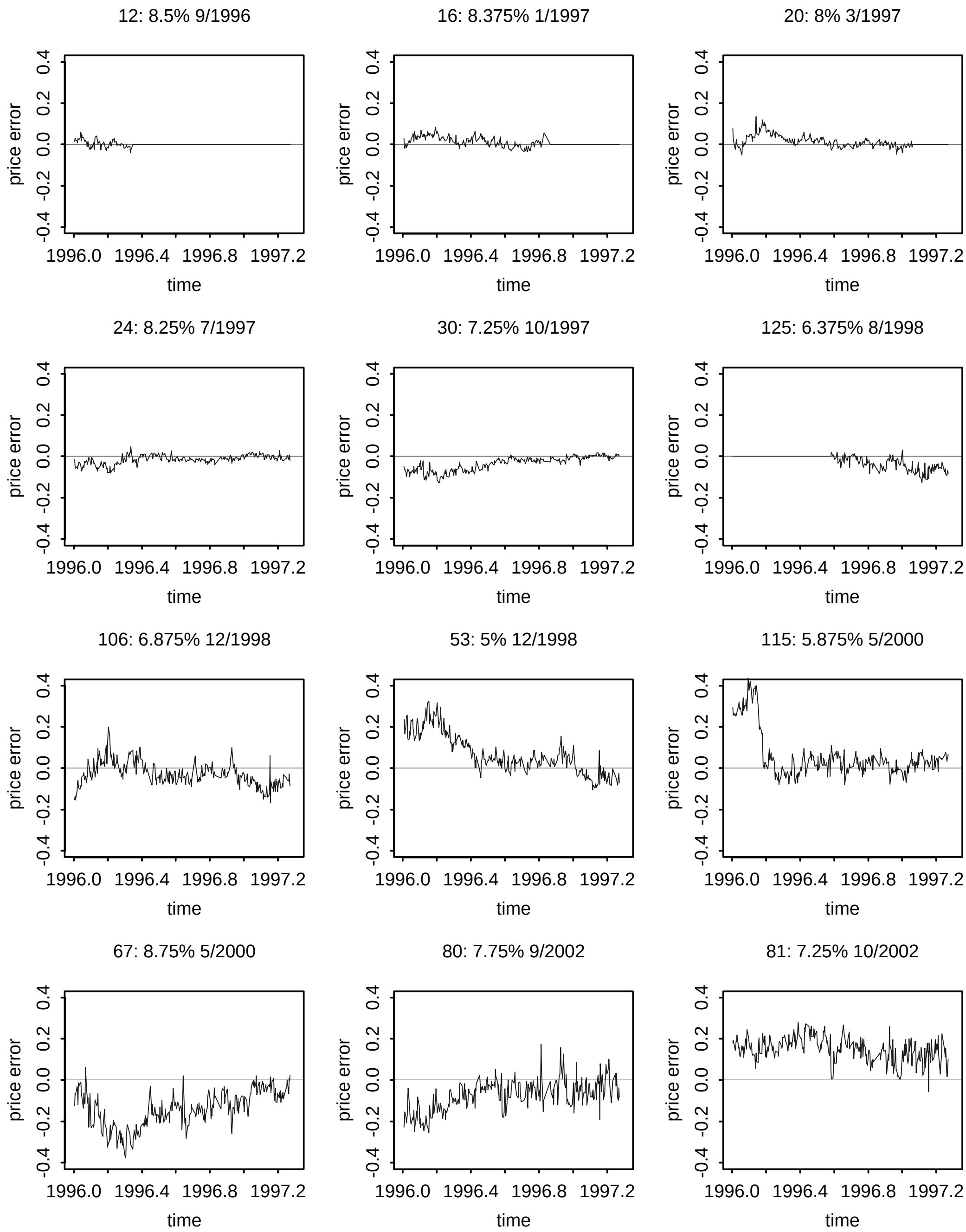


Figure 7: Price errors for selected stocks.

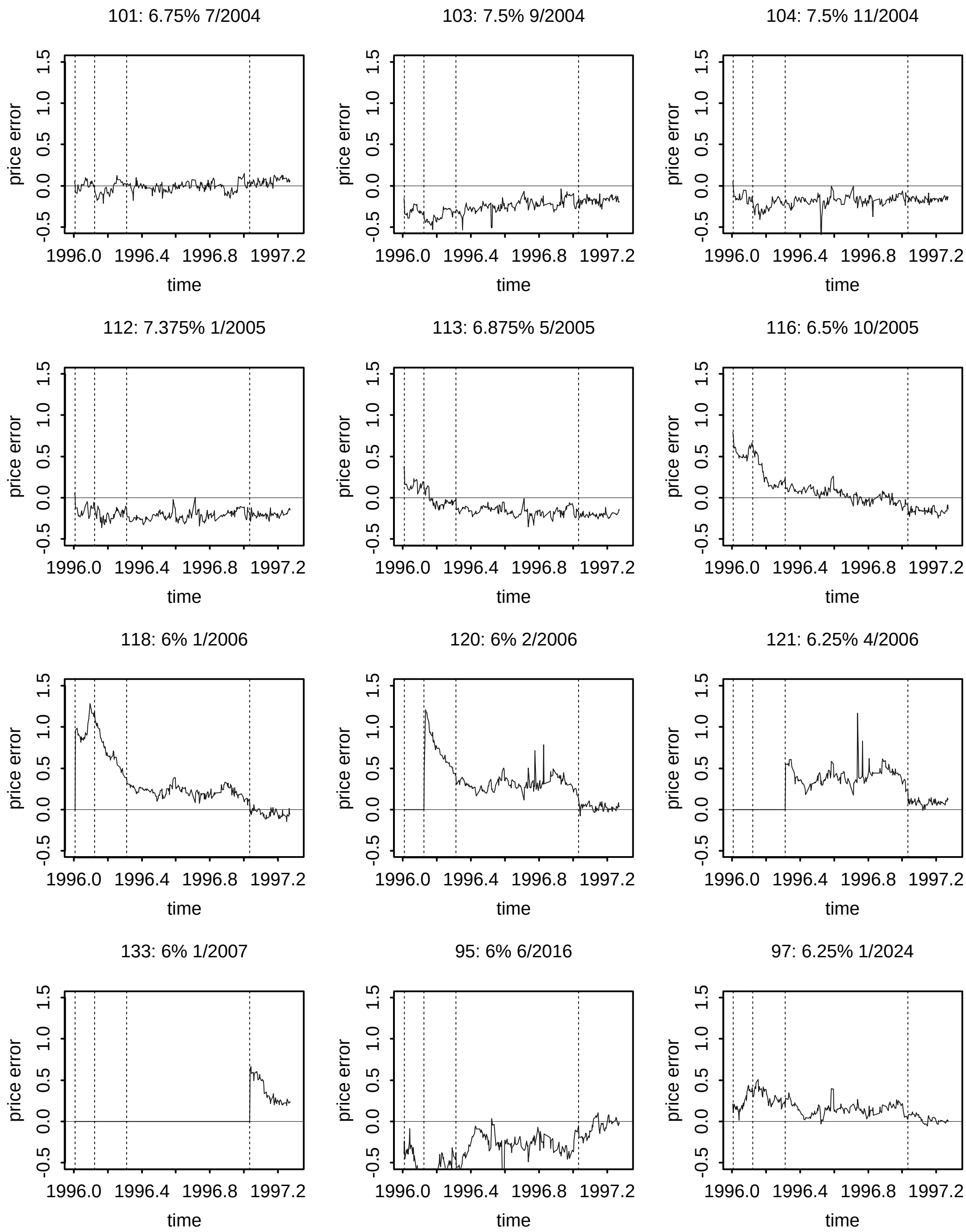


Figure 8: Price errors for selected stocks.

This view is backed up by Figure 10 (top) where the gross redemption yields for the longest stocks just under 10 years appear to be low when we consider yields on the 6% 2016 and $6\frac{1}{4}$% 2024 stocks. Without these two stocks we might postulate that the par yield curve would level off at about 6% at 10 years or even drop (for example, see Figure 4, top, curve E).

It can be seen that there is some problem in fitting well both of the 2016 and 2024 stocks. If we exclude, for example, 6% 2016 then we find that the model is able to get price errors on the 2024 stock down almost to zero. Alternatively, we could increase the weight attached to the 2016 and 2024 to force a closer fit. This additional weight would have to be very substantial to counterbalance the large numbers of stocks with less than 10 years to maturity.

In Figure 9 we consider how different values assigned to the vector c affect the fit. In the upper graph we have:

A: $c = (0.2, 0.4, 0.8, 1.6)$; and

B: $c = (0.05, 0.1, 0.2, 0.4)$;

preserving the power sequence in the c_i proposed by Cairns (1998). The corresponding lines A and B have plotted the ratio of the RMSE to the RMSE estimated using $c = (0.1, 0.2, 0.4, 0.8)$. (Plotting the ratio of RMSE's makes comparison much easier.) We can see that, of the two, $c = (0.2, 0.4, 0.8, 1.6)$ is better, but this is not as good as the central $c = (0.1, 0.2, 0.4, 0.8)$.

In the lower graph in Figure 9 we have:

A: $c = (0.2, 0.4, 0.6, 0.8)$; and

B: $c = (0.1, 0.2, 0.3, 0.4)$;

using an arithmetic sequence as suggested by Chaplin (1998). We can observe that $c = (0.1, 0.2, 0.3, 0.4)$ gives relatively poor results while $c = (0.2, 0.4, 0.6, 0.8)$ does perform well and only marginally worse than $c = (0.1, 0.2, 0.4, 0.8)$. We can also note here that over the period being analysed $c = (0.2, 0.4, 0.6, 0.8)$ was able to reduce significantly the price errors on the 2016 and 2024 stocks observed in Figure 8. As a consequence it was felt that no firm conclusion could be made preferring $c = (0.1, 0.2, 0.4, 0.8)$ over $c = (0.2, 0.4, 0.6, 0.8)$ or vice versa.

In Figure 10 we investigate the effects of changing the way in which we estimate the long end of the forward-rate curve.

A: As before, with Groups 1, 2 and 3 excluded and $c = (0.1, 0.2, 0.4, 0.8)$.

B: As A but with 6% 2016 excluded in addition.

C: As A but 6% 2016 and $6\frac{1}{4}$% 2024 excluded; estimating b_1 to b_4; $b_0 = 0.06$ fixed.

D: As C but $b_0 = 0.03$ fixed.

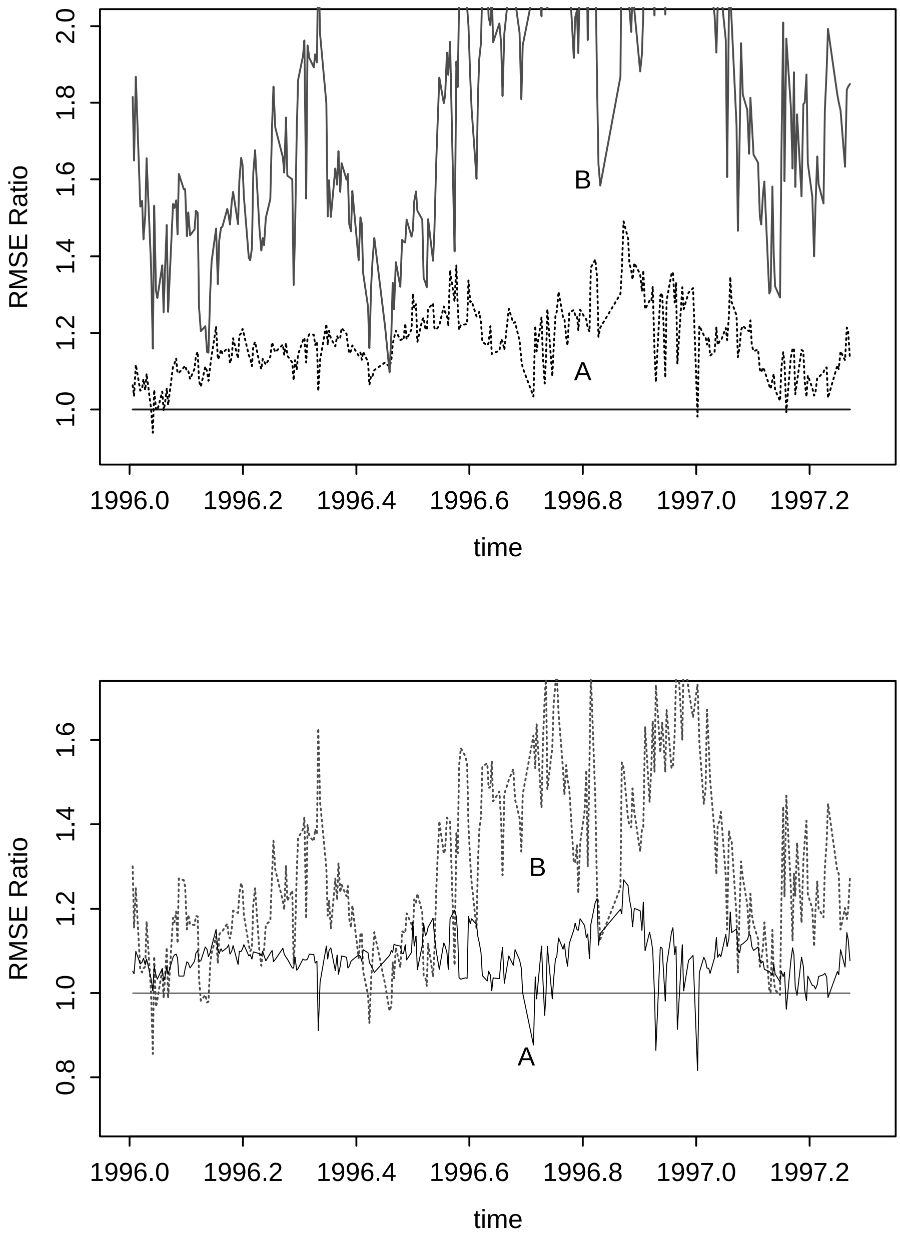


Figure 9: Variation of c from the central case of $0.1, 0.2, 0.4, 0.8$. The lines plotted give the ratio of the RMSE under the alternative proposal for c to the RMSE under the central case. *Upper graph*: A (dotted line): $c = (0.2, 0.4, 0.8, 1.6)$. B (solid line): $c = (0.05, 0.1, 0.2, 0.4)$. *Lower graph*: A (solid line): $c = (0.2, 0.4, 0.6, 0.8)$. B (dotted line): $c = (0.1, 0.2, 0.3, 0.4)$.

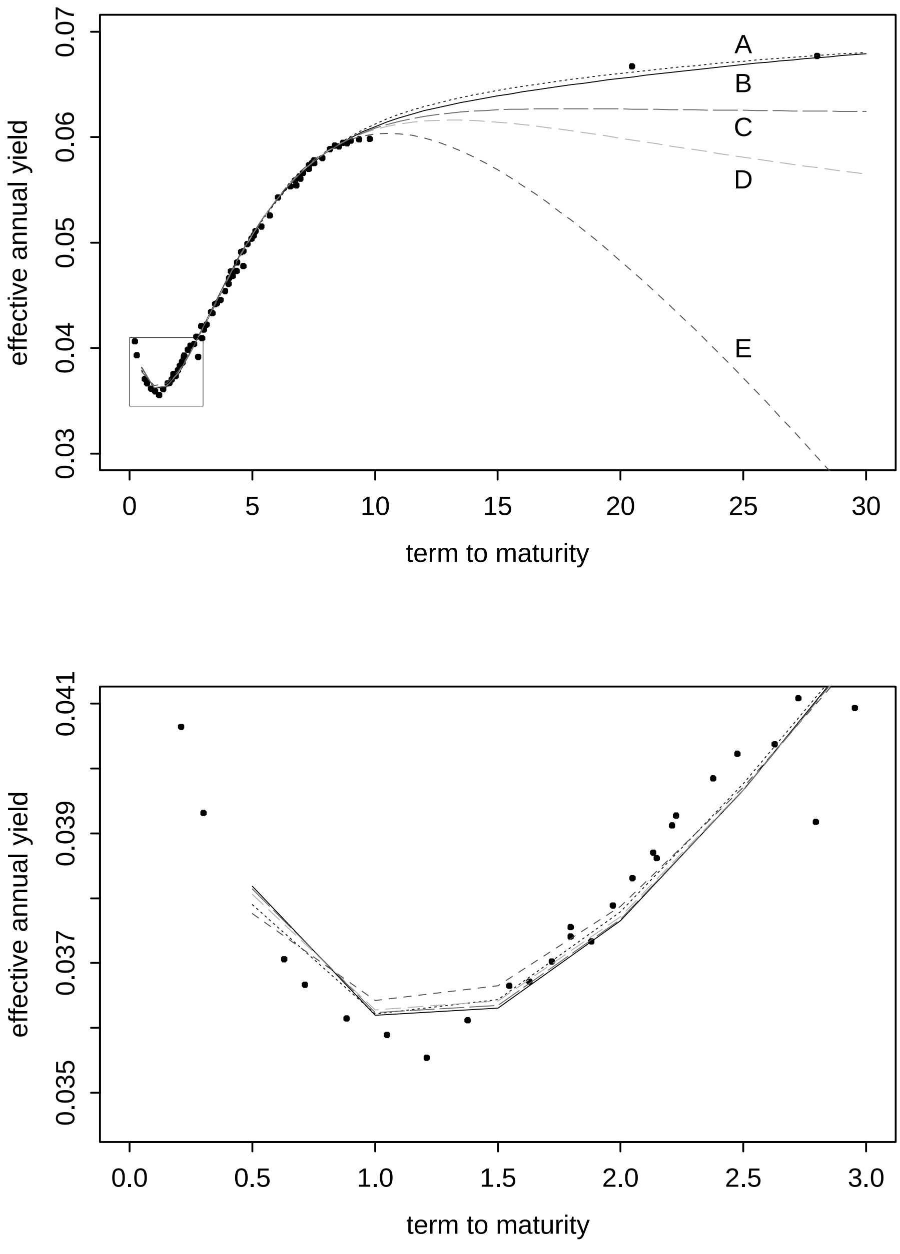


Figure 10: Observed gross redemption yields on 4 January 1996 (dots) and fitted par-yield curves A to E. A, B (dotted and solid line) Groups 1, 2 and 3 excluded and (B only) 6% 2016 excluded. C, D, E (long, medium and short dashed curves) $6\frac{1}{4}\%$ $1/2024$ also excluded. C: $b_0 = 0.06$. D: $b_0 = 0.03$. E: b_0 estimated. Lower graph is a close up of the upper graph.

E: As A but 6% 2016 and $6\frac{1}{4}$% 2024 excluded and estimating b_0 to b_4.

Both graphs in Figure 10 give observed gross-redemption yields and fitted par-yield curves on 4 January 1996 with the lower graph being a close up on terms up to three years.

Curve E clearly fits the data best, but the lack of a long-dated stock or a constraint on b_0 meant that the curve fitting process was 'fooled' by the price premium on the longest stocks under 10 years. (Indeed the estimate for b_0 is much more volatile than that under cases A abd B and it is very often negative.) On the other hand we see that the quality of fit for curves A to D is very similar with the only differences occurring beyond the range of the data at 10 years (curves C and D). In the lower graph the differences between the 5 curves is at most about 3 basis points.

In Figure 11 we look at the effect of the removal of 6% 2016 and of changing c on the long and the short end of the par-yield curve.

A: As before, with Groups 1, 2 and 3 excluded and $c = (0.1, 0.2, 0.4, 0.8)$.

B: As A but with 6% 2016 excluded in addition.

C: As A but $c = (0.2, 0.4, 0.6, 0.8)$.

At the long end (upper graph) we can see that the removal of 6% 2016 has the effect of lowering the par yield curve slightly between terms 10 and 28 years (although the $6\frac{1}{4}$% 2024 price error is reduced). Curve C indicates that $c = (0.2, 0.4, 0.6, 0.8)$ allows us to get a better fit at the long end. This is confirmed by a plot of the price errors over time for the 2016 and 2024 stocks.

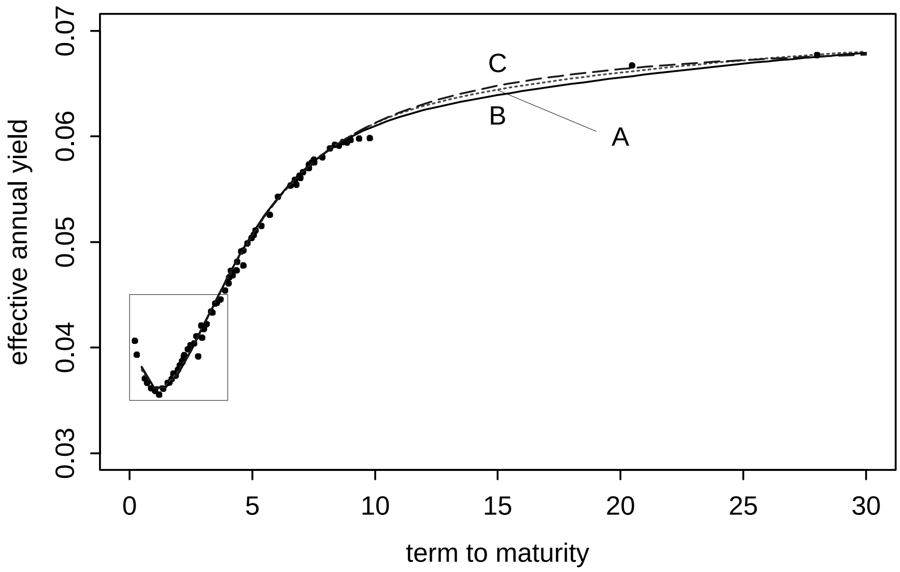
At the short end (lower graph) the differences between A, B and C are very small. These differences can be larger as illustrated in Figure 12.

In Figure 12 we plot differences in par yields over time for terms 1, 5, 10 and 28 years to maturity. In each graph two lines are plotted (with A, B and C as defined above):

1. par yield for A minus par yield for B;
2. par yield for A minus par yield for C.

We can note the following points:

- Differences are smallest at term 5 where we are in the middle of the bulk of the data. Differences become larger towards the extremities of the data (terms 1 and 10).
- The jump in the term 1 curves in the first half of 1996 appears to correspond to the removal of the shortest-dated stock at that time: 8.5% 9/1996. Clearly A reacted in a different way to this removal than B and C.
- At term 10 we see that the removal of 6% 2016 has a relatively significant effect, while a change from $c = (0.1, 0.2, 0.4, 0.8)$ to $c = (0.2, 0.4, 0.6, 0.8)$ has very little effect.



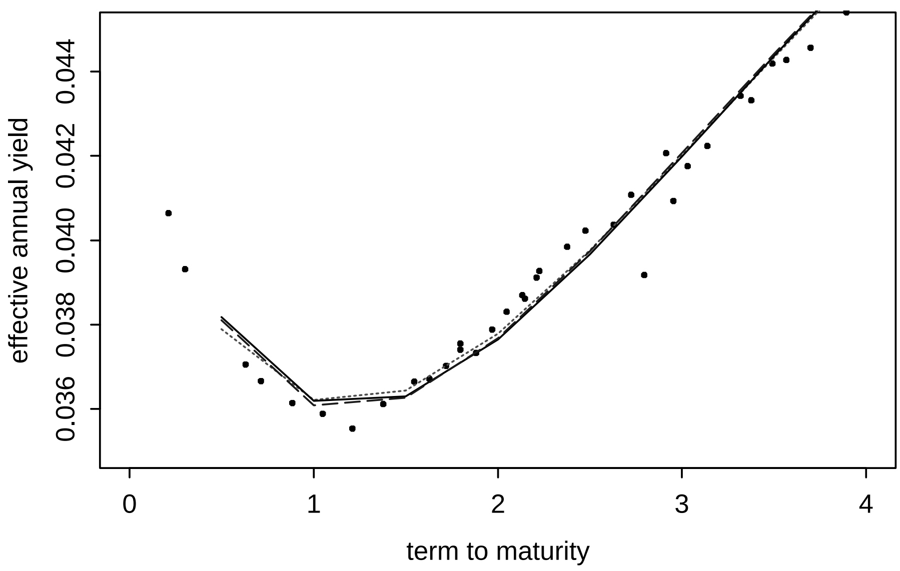


Figure 11: Observed gross redemption yields on 4 January 1996 (dots) and fitted par-yield curves A to C. A (dotted curve): central case. B (solid curve): with 2016 removed. C (dashed curve): with $c = (0.2, 0.4, 0.6, 0.8)$. Lower graph is a close up of the upper graph.

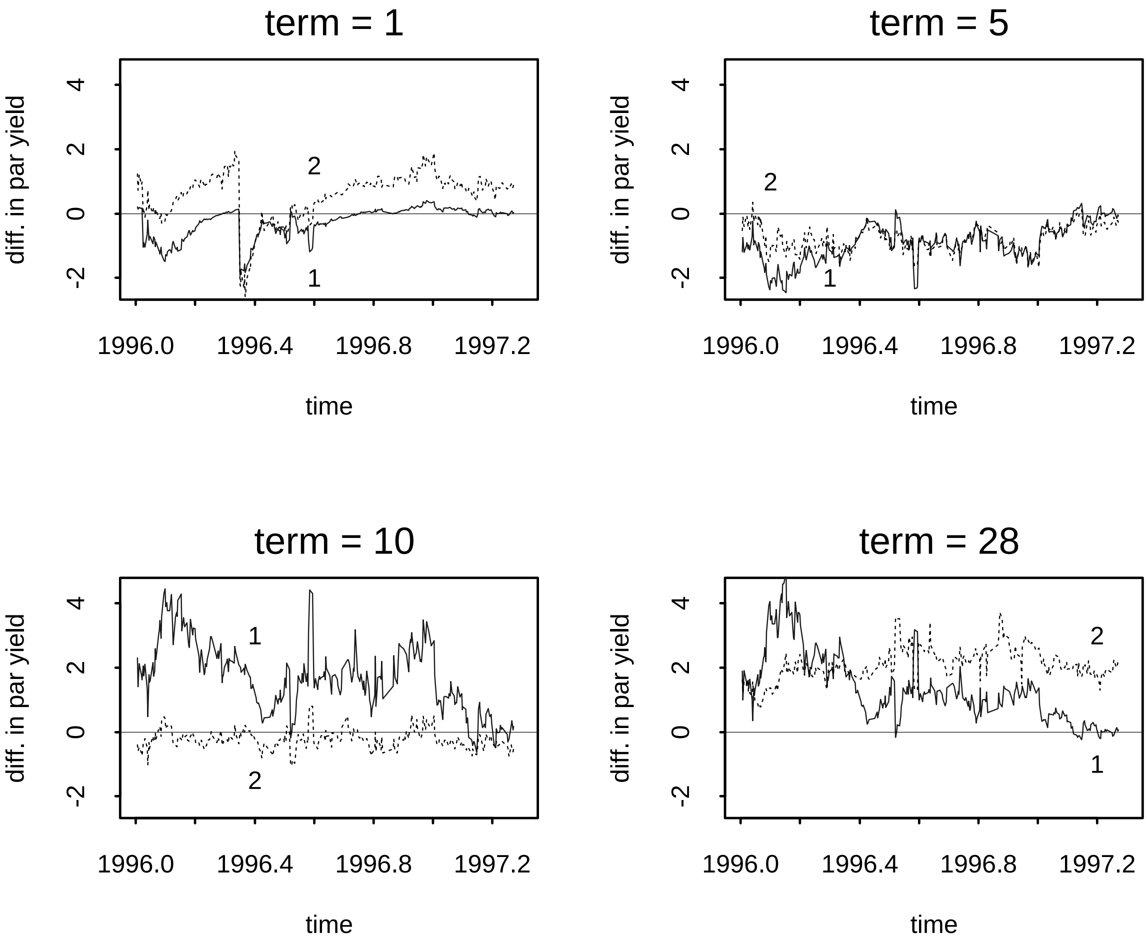


Figure 12: Differences in par yields (in basis points) for: 1 (solid curve): case A minus case B and 2 (dotted curve): case A minus case C.

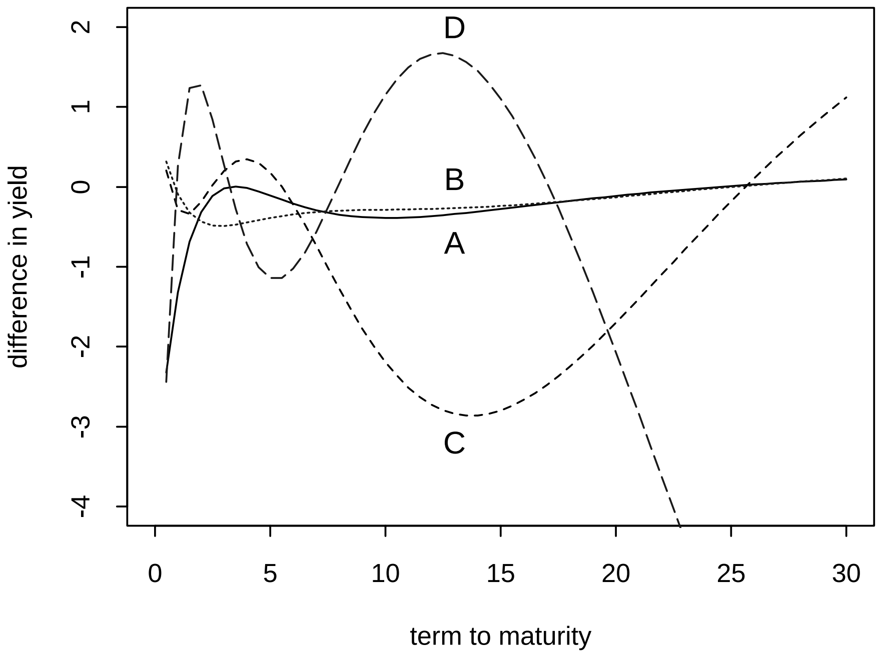


Figure 13: Sensitivity of fitted par yield curves to changes in prices of single stocks. Difference in yield before and after the price change is plotted in basis points. A: change in $8\frac{3}{8}$% 1/1997. B: change in 5% 1/1999. C: change in $6\frac{1}{2}$% 10/2005. D: $6\frac{1}{4}$% 1/2024.

- At term 28 we can match the shape of the curve number 1 with the shape of the price errors curve for $6\frac{1}{4}$% 2024 in Figure 8. This is because after the removal of 6% 2016 price errors for the 2024 stock a reduced almost to zero.
- For terms between 10 and 20 differences in fitted par yield curves are generally larger because of the sparsity of data in this region.
- For terms less than 1 year or above 28 years differences also become more marked as we begin to extrapolate beyond the range of the data.

Finally we considered the effects on the fitted par yield curve of changes in the price of individual stocks. Results of this analysis are plotted in Figure 13. We took data on 4 January 1996 and for each stock price varied we plot the difference (in basis points) between the fitted par yield curves before and after the change. Here we give the results for 4 different stocks with different terms to maturity:

A: Add 0.20 to the price of $8\frac{3}{8}$% 1/1997 (1 year to maturity);

B: Add 0.50 to the price of 5% 1/1999 (3 years);

C: Add 1.00 to the price of $6\frac{1}{2}$% 10/2005 (10 years);

D: Add 1.20 to the price of $6\frac{1}{4}$% $1/2024$ (28 years).

This represents a reduction in the gross redemption yield of each stock of around 15 to 20 basis points.

We note the following points:

- The price change which has the smallest effect overall is for the 3-year stock. This is because it falls in the middle of the data and the effect of any individual price change is dampened by the existence of many surrounding stocks.

- Case D is the only one where the par yield at the term of the altered stock changes by a similar amount to the change in the gross redemption yield of the stock itself. For the other stocks, the effect of price changes are dampened by neighbouring stocks to a greater or lesser degree.

- The biggest effect in each of the four cases is close to the term of maturity of the bond being altered: that is, 0.5 years for A, 3 years for B, 14 years for C and 28 years for D.

- However, the effect of a price change does not diminish uniformly the further we get from the term of the altered stock. Instead there is something of a wave effect. For example, in D the par yield curve actually *rises* by nearly 2 basis points at term 12. In addition, the par yield curve also falls by over two basis points at term 0.5 with smaller effects in between.

  Anderson & Sleath (1999) report that such wave effects arise in other models (Nelson & Siegel, 1987, and Svensson, 1994). On they other hand, Anderson & Sleath (1999) argue that spline-based curve estimation methods are more stable: restricting the effect of a single price change to the locality of that change. However, it is clear that no method gives reliable estimates outside the range of the data or, indeed close to its edges. If we restrict our attention to terms of 1 up to 10 years (for this German data) then the longer-range effects of individual price changes is very limited.

  We conjecture here that the magnitude of the wave effect under the restricted exponential model should be less than that under the Svensson model. This is because we have the fixed parameters $c_1, \ldots, c_4$ should help to localise sensitivity. In contrast estimation of c_1 and c_2 under the Svensson model may lead to greater sensitivity overall in addition to the instabilities discussed earlier in this paper.

4.2 Conclusions

The original intention in 1997 of this part of the research was to investigate the suitability of the restricted-exponential model to European bond markets. The need to analyse data from many countries has now been removed by the introduction in 1999 of the single European currency. As a consequence all countries participating in the Euro must have similar yield curves to avoid arbitrage opportunities (subject to local variations in the structure of bonds and in taxation). Thus the need to analyse each country is no longer there.

This analysis suggests that the restricted exponential model should apply just as well to German and other European bond markets as it does in the UK (Cairns, 1998, Feldman *et al.*, 1998). However, the existence of different structures in each country is likely to cause difficulty with the construction of a single bond-yield index for the Euro zone.

The best choice for the vector c still remains an open question requiring longer runs of data to provide an adequate solution.

Acknowledgements

AC has benefitted from conversations with many individuals: in particular, David Wilkie, Geoff Chaplin, Andrew Smith, Gerry Kennedy, Hans Bühlmann, Freddy Delbaen and Julian Wiseman. Thanks go also to participants at two conferences where part of this work was presented: Quantitative Methods in Finance, Cairns, August 1997, and Groupe Consultatif 10th Colloquium: Interest Risk in Insurance and Pensions, Barcelona, September 1997.

This work was carried out while the authors were in receipt of a grant from the Faculty and Institute of Actuaries.

References

1. ANDERSON, N., AND SLEATH, J. (1999) New estimates of the UK real and nominal yield curves. Bank of England Quarterly Bulletin39384-392

2. BJØRK, T. AND CHRISTENSEN, B. J. (1997) Forward rate models and invariant manifolds. Preprint, Stockholm School of Economics

3. CAIRNS, A.J.G. (1995) Uncertainty in the modelling process. Transactions of the 25th International Congress of Actuaries, Brussels **1**, 67-94.

4. CAIRNS, A.J.G. (1998) Descriptive bond-yield and forward-rate models for the British government securities' market. British Actuarial Journal **4**, 265-321

5. Chaplin, G.B.1998A review of term-structure models and their applicationsBritish Actuarial Journal4323-349

6. DALQUIST, M., AND SVENSSON, L.E.O. (1996) Estimating the term structure of interest rates for monetary policy analysis. Scandinavian Journal of Economics **98**, 163-183

7. DEACON, M., AND DERRY, A. (1994) Estimating the term structure of interest rates. Bank of England Working Paper Series Number 24

8. DOBBIE, G.M. AND WILKIE, A.D. (1978) The FT-Actuaries Fixed Interest Indices. Journal of the Institute of Actuaries **105**, 15-27 and *Transactions of the Faculty of Actuaries* **36**, 203-213 (1978)

9. FELDMAN, K.S., BERGMAN, B., CAIRNS, A.J.G., CHAPLIN, G.B., GWILT, G.D., LOCKYER, P.R., AND TURLEY, F.B. (1998) Report of the fixed interest working group. British Actuarial Journal **4**, 213-263

10. FISHER, M., NYCHKA, D., AND ZERVOS, D. (1995) Fitting the term structure of interest rates with smoothing splines. *Finance and Economic Discussion Series* 95-1 Federal Reserve Board

11. FLESAKER, B., AND HUGHSTON, L.P. (1996) Positive interest. Risk **9(1)**, 46-49

12. 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

13. JARROW, R. (1996) *Modelling Fixed Income Securities and Interest Rate Options.* New York: McGraw-Hill

14. LORIMIER, S. (1995) Interest rate term structure estimation based on the optimal degree of smoothness of the forward rate curve. Ph.D. Thesis, Universiteit Antwerpen, Depertement Wiskunde-Informatica

15. MASTRONIKOLA, K. (1991) Yield curves for gilt-edged stocks: a new model. Discussion paper number 49. Bank of England.

16. MCCULLOCH, J.H. (1971) Measuring the term structure of interest rates. Journal of Business **44**, 19-31

17. MCCULLOCH, J.H. (1975) The tax-adjusted yield curve. Journal of Finance **30**, 811-830

18. MILLS, T.C. (1993) *The Econometric Modelling of Financial Time Series.* Cambridge: CUP

19. NELSON, C.R., AND SIEGEL, A.F. (1987) Parsimonious modeling of yield curves. Journal of Business **60**, 473-489

20. REBONATO, R. (1996) *Interest Rate Option Models.* Chichester: Wiley

21. SILVEY, S.D. (1970) *Statistical Inference.* London: Chapman and Hall

22. 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

23. VASICEK, O. (1977) An equilibrium characterisation of the term structure. Journal of Financial Economics **5**, 177-188

24. VASICEK, O.E., AND FONG, H.G. (1982) Term structure modelling using exponential splines. Journal of Finance **37**, 339-348

25. WAGGONER, D. (1997) Spline methods for extracting interest rate curves from coupon bond prices. Working paper, No 97-10, Federal Reserve Bank of Atlanta

26. WEI, W.W.S. (1990) *Time Series Analysis.* Redwood City: Addison-Wesley

27. WISEMAN, J. (1994) The exponential yield curve model. J.P.Morgan, 16 December

Appendix A

A.1 Zero-coupon bond market

Define $Z(b;t) = \exp[-b^T d(t)]$ where $b^T = (b_0, \ldots, b_m)$, $d(t)^T = (d_0(t), \ldots, d_m(t))$, $d_0(t) = t$ and $d_k(t) = (1 - \exp(-c_k t))/c_k$ for $k = 1, \ldots, m$.

For a bond maturing at t_i write $d_i = d(t_i)$ and $d_{ik} = d_k(t_i)$.

Suppose that the observed prices are P_1, P_2, ..., P_N. The likelihood function is

$$
\begin{aligned}
l(b;P) &= -\frac{1}{2}\sum_{i=1}^{N}\left\{\log 2\pi\sigma^2(t_i) + (\log P_i - \log Z(b;t_i))^2/\sigma^2(t_i)\right\} \\
&= -\frac{1}{2}\sum_{i=1}^{N}\lambda_i(\log P_i + d_i^T b)^2 + \text{constant} \\
\text{where } \lambda_i &= 1/\sigma^2(t_i).
\end{aligned}
$$

Maximising the likelihood is thus equivalent to minimising the function

$$
\begin{aligned}
g(P|b) &= \frac{1}{2}\sum_{i=1}^{N}\lambda_i(\log P_i + d_i^T b)^2 \\
\Rightarrow \frac{d^2 g}{db} &= \sum_i \lambda_i d_i d_i^T.
\end{aligned}
$$

The matrix of second derivatives is constant and positive definite.

Proposition 2.1(a)

Hence $g(P|b)$ is convex and has a unique minimum in b.

If we wish instead to use Bayesian methods it is necessary for us to specify also a prior density function. Suppose that the log-prior density function is denoted by $p(b)$. The log-posterior density function is then

$$p(b|P) = p(b) - g(P|b) + \text{constant}.$$

Corollary 2.2(a)

If $p(b)$ is convex then $-\partial^2 p(b)/\partial b^2$ is positive semi-definite. Thus the matrix of second derivatives of minus the log-posterior density function $-\partial^2 p(b|P)/\partial b^2$ is positive definite and there is a unique value of b for which $\partial p(b|P)/\partial b = 0$.

A.2 Coupon-bond market

Suppose that there are N bonds each with a nominal value of 1. Bond i has cashflows $c_{i1}, \ldots, c_{in_i}$ at times $0 < t_{i1} < \ldots < t_{in_i}$ respectively. Given b, the theoretical price of each bond is then:

$$
\begin{aligned}
\hat{P}_i(b) &= \sum_{j=1}^{n_i} c_{ij} Z(b;t_{ij}) \\
&= \sum_{j=1}^{n_i} c_{ij} \exp(-b^T d(t_{ij})) \\
\text{Write } \; d_{ij} &= d(t_{ij}) \\
d_{ijk} &= d_k(t_{ij}) \\
g(b) &= \frac{1}{2} \sum_{i=1}^{N} \lambda_i [\log P_i - \log \sum_{j=1}^{n_i} c_{ij} Z(b;t_{ij})]^2
\end{aligned}
$$

(Recall that the log-likelihood is $l(b;P) = -g(b) + \text{constant}$.)

Now

$$
\begin{aligned}
Z(b;t) &= \exp(-b^T d(t)) \\
\Rightarrow \frac{dZ(b;t)}{db} &= -Z(b;t) d(t) \\
\frac{d^2 Z(b;t)}{db^2} &= Z(b;t) d(t) d(t)^T. \\
\text{Thus } \; g'(b) &= -\sum_i \lambda_i [\log P_i - \log \sum_j c_{ij} Z(b;t_{ij})] \frac{\sum_j c_{ij} \frac{dZ(b;t_{ij})}{db}}{\sum_j c_{ij} Z(b;t_{ij})} \\
&= -\sum_i \lambda_i [\log P_i - \log \sum_j c_{ij} Z(b;t_{ij})] \sum_j f_{ij}(b) d_{ij} \\
&= -\sum_i \lambda_i [\log P_i - \log \sum_j c_{ij} Z(b;t_{ij})] \bar{d_i}(b) \\
\text{where } \; \bar{d_i}(b) &= \sum_j f_{ij}(b) d_{ij} \\
\text{and } \; f_{ij}(b) &= \frac{c_{ij} Z(b;t_{ij})}{\sum_j c_{ij} Z(b;t_{ij})} \\
g''(b) &= \sum_i \lambda_i \left(\frac{\sum_j c_{ij} \frac{dZ(b;t_{ij})}{db}}{\sum_j c_{ij} Z(b;t_{ij})} \right) \left(\frac{\sum_j c_{ij} \frac{dZ(b;t_{ij})}{db}}{\sum_j c_{ij} Z(b;t_{ij})} \right)^T - \sum_i \lambda_i [\log P_i - \log \sum_j c_{ij} Z(b;t_{ij})] V_i(b) \\
&= \sum_i \lambda_i \bar{d_i}(b) \bar{d_i}(b)^T - \sum_i \lambda_i [\log P_i - \log \sum_j c_{ij} Z(b;t_{ij})] V_i(b) \\
\text{where } \; V_i(b) &= \frac{d}{db} \left(\frac{\sum_j c_{ij} \frac{dZ(b;t_{ij})}{db}}{\sum_j c_{ij} Z(b;t_{ij})} \right) \\
&= \frac{(\sum_j c_{ij} Z(b;t_{ij}) d_{ij} d_{ij}^T)(\sum_j c_{ij} Z(b;t_{ij})) - (\sum_j c_{ij} Z(b;t_{ij}) d_{ij})(\sum_j c_{ij} Z(b;t_{ij}) d_{ij}^T)}{(\sum_j c_{ij} Z(b;t_{ij}))^2} \\
&= \sum_j f_{ij}(b) d_{ij} d_{ij}^T - (\sum_j f_{ij}(b) d_{ij})(\sum_j f_{ij}(b) d_{ij})^T
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_j f_{ij}(b)(d_{ij} - \bar{d_i}(b))(d_{ij} - \bar{d_i}(b))^T \\
&\geq 0
\end{aligned}
$$

Now write $X_i(b) = \log P_i - \log \sum_j c_{ij} Z(b; t_{ij}) = \log[P_i/\hat{P}_i(b)]$.

$$
\begin{aligned}
g'(b) &= \sum_i \lambda_i X_i(b) \bar{d_i}(b) \\
g''(b) &= \sum_i \lambda_i \left(\bar{d_i}(b) \bar{d_i}(b)^T - X_i(b) V_i(b) \right)
\end{aligned}
$$

Now in a zero-coupon bond market $V_i(b) \equiv 0$ and the $\bar{d_i}(b)$ are constant and do not depend on b. Thus $g''(b)$ is constant and positive definite confirming the simpler derivation above.

A.3 Assumptions

We make the following assumptions:

The range of acceptable values for b is denoted by $\mathcal{B}$. Each $b \in \mathcal{B}$ must satisfy the following criterion: for all $0 < t < s$, $1 > Z(b;t) > Z(b;s)$. This criterion is equivalent to the assumption that the forward-rate curve $f(t, t+s)$ is non-negative for all t and for all $s > 0$: that is,

$$
\begin{aligned}
\mathcal{B} &= \{ b : \ b^T d'(t) \geq 0 \text{ for all } t \} \\
\text{where } \ d'(t)^T &= (1, \ \exp(-c_1 t), \ \ldots, \ \exp(-c_m t) \ ).
\end{aligned}
$$

Suppose that $u > 0$ and that $b \in \mathcal{B}$. For any $0 < t < s$ we note that $1 > Z(b;t) > Z(b;s)$ and therefore $0 < b^T d(t) < b^T d(s)$. Thus

$$
\begin{aligned}
0 < u.b^T d(t) &< u.b^T d(s) \\
\Rightarrow 1 > \exp(-ub^T d(t)) &> \exp(-ub^T d(s)) \\
\Rightarrow 1 > Z(bu;t) &> Z(bu;s)
\end{aligned}
$$

Thus $b \in \mathcal{B}$ if and only if $bu \in \mathcal{B}$ for all $u > 0$. That is, $\mathcal{B}$ is a cone.

Furthermore, suppose that b_A and b_B are in $\mathcal{B}$. Then for any λ such that $0 < \lambda < 1$, and for any $0 < t < s$:

$$
\begin{aligned}
0 < b_A^T d(t) &< b_A d(s) \\
\text{and } \ 0 < b_B^T d(t) &< b_B d(s) \\
\Rightarrow 0 < (1-\lambda) b_A^T d(t) + \lambda b_B^T d(t) &< (1-\lambda) b_A^T d(s) + \lambda b_B^T d(s) \\
\Rightarrow 0 < [(1-\lambda) b_A + \lambda b_B]^T d(t) &< [(1-\lambda) b_A + \lambda b_B]^T d(s) \\
\Rightarrow [(1-\lambda) b_A + \lambda b_B] &\in \mathcal{B}
\end{aligned}
$$

Thus $\mathcal{B}$ is a convex cone.

A.4 Locality of possible minima

We define the following sets:

$$\begin{aligned} B_0 &= \{b \in B : \ P_i < \hat{P}_i(b) \text{ for all } i\} \\ B_1 &= \{b \in B : \ P_i > \hat{P}_i(b) \text{ for all } i\} \\ B_2 &= B \backslash B_1. \end{aligned}$$

Clearly for all $b \in B_0$, $\hat{P}_i(bs) > P_i$ for all $0 < s \leq 1$. That is, $b \in B_0$ implies that $bs \in B_0$ for all $0 < s \leq 1$.

Similarly, $b \in B_1$ implies that $bs \in B_1$ for all $1 \leq s < \infty$.

Lemma A.1

There does not exist $b \in B_0$ or $b \in B_1$ such that $g'(b) = 0$ is positive definite.

Proof Suppose that there exists such a $b \in B_1$.

Let $s_0 = \inf\{s : bs \in B_1\}$. Thus $\hat{P}_i(bs) < P_i$ for all $s > s_0$ and $\hat{P}_i(bs)$ is decreasing with s for $s > s_0$. Hence $g(bs)$ is an increasing function of s for $s > s_0$ which indicates that there cannot be a minimum at b.

Similarly there does not exist such a $b \in B_0$.

Lemma A.2

B_1 is convex.

Proof

Note that $\hat{P}_i(b)$ is convex in b for all i.

Suppose b_A and b_B are members of B_1.

Then $\hat{P}_i(b_A) < P_i$ and $\hat{P}_i(b_B) < P_i$.

For $0 < \lambda < 1$: since $\hat{P}_i(b)$ is convex,

$$\hat{P}_i((1-\lambda)b_A + \lambda b_B) < (1-\lambda)\hat{P}_i(b_A) + \lambda \hat{P}_i(b_B) < P_i$$

Thus $b_A, b_B \in B_1 \Rightarrow (1-\lambda)b_A + \lambda b_B \in B_1$.

Given $b_A, b_B \in B_1$ this is true for all i. Thus B_1 is convex.

Proposition 2.1(b)

For small coupon rates there is a unique maximum.

Proof

We proceed as follows:

(a) Establish a means of moving continuously and smoothly from zero-coupon bonds, via a new parameter, γ ($\gamma = 0$ giving a zero-coupon bond market and $\gamma = 1$ giving us the true coupon-bond market).

(b) Establish that $B_2 = B \backslash B_1$ is finite, and let B_3 be some finite expansion of B_2 and which contains B_2 for all values of $\gamma : 0 \leq \gamma \leq 1$.

(c) Within B_3, the shape of the log-likelihood function $g(b)$ can only deform slightly as we increase γ from 0. In particular, for small γ there will still only be one maximum.

(a) We start with a coupon-bond market with N stocks. For stock i we have an observed price P_i. The n_i future cashflows under this stock are $c_{i1}, \ldots, c_{in_i}$ at times $t_{i1}, \ldots, t_{in_i}$. The final payment is made up of nominal capital of 1 and a final coupon payment of $c_{in_i} - 1$.

Now specify an arbitrary forward-rate curve $\tilde{f}(t, t+s)$: for example, the curve fitted to yesterday's prices. This has a corresponding set of zero-coupon bond prices $\tilde{Z}(s)$ for maturity in s years.

The actual coupon bond prices are P_i which we can write as

$$P_i = \left(\sum_{j=1}^{n_i} c_{ij} \tilde{Z}(t_{ij}) \right) e^{\varepsilon_i}$$

where $\varepsilon_1, \ldots, \varepsilon_N$ are defined by this identity.

We define the γ-coupon bond market as follows. For each stock we multiply the original coupon payments for each stock by γ but retain the full redemption payment. Thus the cashflows for stock i are $c_{i1}(\gamma), \ldots, c_{in_i}(\gamma)$ where

$$c_{ij}(\gamma) = \begin{cases} \gamma c_{ij}, & 1 \leq j \leq n_i - 1 \\ 1 + \gamma(c_{in_i} - 1), & j = n_i \end{cases}$$

The price of stock i in the hypothetical γ-coupon bond market is defined as

$$Q_i(\gamma) = \left(\sum_{j=1}^{n_i} c_{ij}(\gamma) \tilde{Z}(t_{ij}) \right) e^{\varepsilon_i}$$

Clearly $\gamma = 0$ represents a zero-coupon bond market while $\gamma = 1$ returns us to the original coupon-bond market.

Proof of (b):

Let $R(t, t+s; b) = b^T d(s)/s$ be the spot rate at time t for maturity at time $t+s$.

Let t_0 be the shortest dated time to a coupon or a redemption payment.

Let $r_m = \inf\{R(t, t+t_0; b) : \ |b| = 1, b \in B\}$ and let $B_m = \{b : \ R(t, t+t_0; b) = r_m\}$.

Let $E(\lambda) = \{b : \ b \in B, |b| = \lambda\}$, $\tilde{E}(\lambda) = \{b : \ b \in B, |b| \leq \lambda\}$.

$E(1)$ is a closed set so that r_m is attainable: that is, B_m is non-empty.

$r_m > 0$. Otherwise there exists b such that $R(t, t+t_0; b) = 0$. This implies that $f(t, t+s; b) = 0$ for $0 < s < t_0$. This can only be true if $b \equiv 0$.

Let $\tilde{\lambda}(\gamma) = \sup_i \frac{1}{r_m t_0} \log \frac{F_i(\gamma)}{Q_i(\gamma)}$ where $Q_i(\gamma)$ is the price of stock i in the γ-coupon bond market and $F_i(\gamma) = \sum_{i=1}^{n_i} c_{ij}(\gamma)$ is the total amount of the future payments under that stock.

Consider the γ-coupon bond market. For all $b \in B, \ |b| \geq \tilde{\lambda}$, for all i, the theoretical price of stock i is:

$$
\begin{aligned}
\hat{Q}_i(\gamma)(b) &= \sum_{j=1}^{n_i} c_{ij}(\gamma)\exp(-R(t,t+t_{ij};b)t_{ij}) \\
&= \sum_{j=1}^{n_i} c_{ij}(\gamma)\exp(-R(t,t+t_{ij};b/|b|)t_{ij}|b|) \\
&\leq \sum_{j=1}^{n_i} c_{ij}(\gamma)\exp(-r_m t_0|b|) \\
&\leq F_i(\gamma)\exp(-r_m t_0\tilde{\lambda}) \\
&\leq Q_i(\gamma)
\end{aligned}
$$

Thus $B_2(\gamma) = B\backslash B_1(\gamma) \subset B \cap \tilde{E}\left(\tilde{\lambda}(\gamma)\right)$.

Choose some λ' such that $\sup_{0\leq\gamma\leq 1}\tilde{\lambda}(\gamma) < \lambda < \infty$.
Let $B_3 = B \bigcap \tilde{E}(\lambda')$.

Write $g(b,\gamma)$ for the log-likelihood function for the γ-coupon bond market.
Clearly $g(b,\gamma)$ is infinitely differentiable.
For any γ ($0 \leq \gamma \leq 1$), for any i and for any $b \in B \setminus B_3$, $\hat{Q}_i(\gamma)(b) < Q_i(\gamma)$.
Thus, by Lemma A.1, there can be no $b \in B$ outside $\tilde{E}(\lambda')$ such that $\partial g/\partial b(b,\gamma) = 0$ at b.
Hence no local maxima can come sliding in from infinity as soon as $\gamma > 0$.
Let $\hat{b} \in B_3$ be the unique value of b such that

$$
\frac{\partial g}{\partial b}(\hat{b},0) = 0.
$$

Note that $-\partial^2 g/\partial b^2(b,0)$ is constant and strictly positive definite. (This leads to the unique maximum $\hat{b}$ mentioned above for the zero-coupon bond market.)
Hence there exists $\gamma_1 > 0$ such that $-\partial^2 g/\partial b^2(b,\gamma)$ is strictly positive definite for all $0 < \gamma < \gamma_1$ and for all $b \in B_3$ (since g is C^∞ and B_3 is finite).
Thus for each γ ($0 < \gamma < \gamma_1$) there is a unique $\hat{b}(\gamma) \in B_3$ (and hence B) which maximises $g(b,\gamma)$.

Corollary 2.2(b)

Furthermore, if a prior distribution for b is such that -1 times the log-prior density function is positive semi-definite then -1 times the log-posterior density function is also positive definite within B_3.