ELSEVIER

Journal of Banking & Finance 27 (2003) 1–26

Journal of BANKING & FINANCE

www.elsevier.com/locate/econbase

A comparison of yield curve estimation techniques using UK data

Michalis Ioannides *

Watson Wyatt Partners, Watson House, London Road, Reigate, Surrey RH2 9PQ, UK

Received 31 July 2000; accepted 3 May 2001

Abstract

I compare different methods of estimating the term structure of interest rates on a daily UK treasury bill and gilt data that spans the period from January 1995 to January 1999. In-sample and out-of-sample statistics reveal the superior pricing ability of certain methods characterised by an exponential functional form. In addition to these standard goodness of fit statistics, model performance is judged in terms of two trading strategies based on model residuals. Both strategies reveal that parsimonious representations of the term structure perform better than their spline counterparts characterised by a linear functional form. This is valid even when abnormal returns are adjusted for market movements. Linear splines overfit the data and are likely to give misleading results.

JEL classification: E43; C52; G12; G14

Keywords: Bonds; Trading strategy; Term structure of interest rates; Pricing; Splines

1. Introduction

An important "tool" in the development and testing of financial theory is the term structure of interest or forward rates. This relationship between rates and term to maturity has proved to be critical to policy makers and to market practitioners. In particular, forward rates may serve as indicators of monetary policy and as inputs to a pricing model. Indeed, the examination of the term structure theories empowered

* Tel.: +44-1737-274083; fax: +44-1737-241496.

E-mail address: michalis.ioannides@eu.watsonwyatt.com (M. Ioannides).

0378-4266/02/$ - see front matter
PII: S0378-4266(01)00217-5

by theory of contingent claims has led to the derivation of term structure models characterised by the absence of arbitrage.

There are two approaches [1] to term structure modelling. The first approach is the "term structure consistent" approach pioneered by Ho and Lee (1986) and Heath et al. (1992). Recently, Fisher and Gilles (1998) have constructed term structure models in the spirit of Heath et al. (1992) that are consistent with the unbiased expectation hypothesis. This approach works only if an estimate of the initial forward curve is inputted to the Heath et al. (1992) [2] framework. Under such conditions, the pricing and hedging of all types of interest rate claims is possible and consistent with an initial estimate of the forward or spot yield curve.

The second approach identifies a state variable with every yield of maturity τ of zero coupon bonds. This approach attributed to Duffie and Kan (1996) constructs a multifactor model of the term structure in which the yields estimated using a term structure estimation method are the state variables. Although, the idea [3] of using yields as the state variables captures almost all of the dynamics of the term structure, Steeley (1997) conducts a principal component analysis on the Bank of England's term structure estimates to find that at least two factors [4] are enough to capture almost all of the dynamics of the UK term structure. In a different study, Elton et al. (1990) employ McCulloch's (1990) term structure estimates to identify suitable proxies for the unobserved state variables that drive the US term structure of interest rates. Furthermore, Elton and Green (1998) adopt a term structure estimation method, which they augment with tax and liquidity variables [5] to account for the deviations between the actual and the model bond prices.

Despite the widespread use of the term structure of interest rates, a fundamental issue that other researchers seem to ignore is which term structure estimation method should be used in the first place to imply these estimates. Chambers et al. (1984) point out that the use of linear least squares may result in a singular matrix arising from columns that are perfectly collinear. This problem is more acute with longer maturities. An alternative problem arises when the estimation method employs cubic

[1] An alternative modelling approach to the ones mentioned is the one proposed by Ball and Torous (1983) and Schaefer and Schwartz (1987). Unlike the other single factor models that specify a process for the instantaneous short rate, Ball and Torous and Schaefer and Schwartz model the price process of pure discount bonds. Rady and Sandmann (1994) labelled this approach as the *direct approach* to debt option pricing. Another alternative and equally important modelling approach is the *pricing kernel* approach introduced by Constantinides (1992).

[2] Other term structure models that take as an input an estimate of the spot yield curve are Black et al. (1990) and Hull and White (1993). Central to each model is the use of time dependent variables that force the model estimates of the yield curve to equal that market yield curve. In such a way, the model is calibrated to market data. It is then possible to accurately recover the price of plain vanilla bonds.

[3] The idea of using yields as factors originates from the studies of Pearson and Sun (1994) and Chen and Scott (1995), who estimated two factor CIR SR models by performing a change of basis to identify the state variables with yields.

[4] Litterman and Scheinkman (1991) also recognise that the level, the steepness and curvature factors can explain 98% of the variation in yields of US coupon bonds.

[5] Elton and Green (1998) provide evidence to suggest that the tax and liquidity effects are quite small.

splines to fit the term structure of forward rates. Occasionally, such an approach may drift off to negative values.

Another crucial issue in the estimation of the term structure lies in the use of an over-parameterised term structure estimation method, which raises concerns about *overfitting* rather than *genuinely* fitting the bond dataset. Here, I am interested in a term structure estimation method that not only fits the most salient features of the historical bond dataset but also recognises the most stable relationships between the parameters that are useful for out of sample pricing. I also seek to recognise those over-parameterised term structure estimation methods. These methods fit accidental or random features of the bond data that will not recur and are of no use in out of sample pricing.

I investigate two mainstream approaches of estimating the term structure of interest rates, a parsimonious representation defined by an exponential decay term and a spline representation classified into a parametric and non-parametric splines. The former approach is developed by Nelson and Siegel (1987) and extended by Svensson (1994). The parametric cubic spline is introduced to finance by McCulloch (1971, 1975) and the non-parametric splines are developed by Adams and van Deventer (1994), Fisher et al. (1995), Tanggaard (1995) and Waggoner (1997).

Here, I concentrate on seven popular term structure estimation techniques. [6] The Nelson and Siegel and Svensson functions, the McCulloch cubic spline, three non-parametric splines and a non-parametric spline characterised by a maturity dependent penalty. I refer to the three non-parametric splines as linear, exponential and integrated B-splines. I estimate each method on daily UK government bond data spanning a period of four years, from January 1995 to January 1999. Hence, each day I observe seven different estimates of the term structure of interest rates. Second, I choose the "best" function in terms of its ability to replicate market prices of bonds, by performing in-sample and out-of-sample tests. Specifically, I compare estimated term structures by the differences between their in-sample and out-of-sample mean absolute error statistics.

Earlier work by Buono et al. (1992) examined the ability of three term structure estimation techniques, namely the ordinary least squares method, the exponential polynomial and a recursive technique, to accurately estimate the forward curve. Unlike this paper, which uses real market data, Buono et al. based their comparison on Monte Carlo simulations. This work resembles the work of Bliss (1997), who compared four alternate term structure estimation techniques on the basis of their out of sample performance and on the randomness of their fitted price errors.

This work departures from the work of Buono et al. (1992) and Bliss (1997) by applying the trading and filter rule tests of Sercu and Wu (1997). I construct a trading strategy by buying (selling) undervalued (overvalued) bonds according to a given term structure estimation technique and compute the abnormal returns (ARs) as

[6] Having considered actual equilibrium or no-arbitrage models of the term structure, we would have been involved in answering the question of how many stochastic factors are necessary to describe the dynamics of the term structure. This is another topic on its own right and is examined in a companion paper.

defined by three different benchmarks. These alternative definitions of ARs filter out the general market movements and seek out the mere noise in returns. In this way, I identify potential misspecification or mis-estimation of a given term structure estimation technique and isolate "genuine" mispricing. I seek to identify the best model in terms of its ability to capture information in bond residuals and generate ARs. [7] This study is the first to combine and apply this approach to the UK market.

Overall, the performance of both parsimonious functions is favoured over their spline counterparts. Both the in-sample fit and the out-of-sample pricing ability of these techniques to replicate the existing term structure is considerably better than splines defined by linear functional form and marginally better than splines defined by an exponential functional form. In the trading rule test, I render a function superior to another competing alternative when this function produces ARs across all three benchmarks. I would then expect to see that these ARs do not persist. On the other hand, I characterise a function as misspecified or mis-estimated only when I observe ARs in one or two rather than in all three benchmarks. With this in mind, I identify the parsimonious specifications and the non-parametric spline with maturity dependent penalty as superior approaches to the competing alternatives of McCulloch's cubic spline, the linear, exponential and integrated B-splines. However, the performance of exponential and integrated B-splines is marginally close to the performance of the parsimonious methods. Finally, to confirm the results of the trading rule test, we conduct a filter rule test to reveal that the weighting scheme utilised by the trading rule test is not optimal.

The following section describes the alternative methods of the term structure and deals with estimation method and analysis of in-sample and out-of-sample results. Section 3 presents the model specification tests and results and Section 4 concludes.

2. Estimation methods of the term structure of interest rates

The concept of no-arbitrage is an important theoretical breakthrough in finance. Ross (1976, 1978) demonstrates that the absence of arbitrage implies the existence of a linear pricing functional that can be used to value all assets. This linear pricing rule allows us to define the price of the ith bond by the value additivity principle:

$$P_i = \sum_{j=1}^{n} c_i(\tau_j) v_j. \tag{1}$$

Hence, the term structure of interest rates is given for a set of maturities $\{\tau_1, \ldots, \tau_n\}$ by a set of discount factors v $(v_1, \ldots, v_n)$ such that the observed bond price is equal

[7] Here, our ultimate purpose is to use estimates from the selected technique to estimate multifactor models of the term structure. Typically, such models are used to price and hedge term structure derivatives. If such models are unable to replicate the market prices of underlying securities, their ability to price contingent claims would be questioned. Hence, we need to make sure that estimates of the term structure lie closer to market rates and identify genuine mispricing.

to the present value of the cash flows $\{c_i(\tau_1), \ldots, c_i(\tau_n)\}$. However, a significant body of literature investigates how this relationship breaks down in the presence of heterogeneous taxes, transactions costs, short selling constraints and liquidity effects. [8] The purpose of this study is not to find a method of incorporating these complicated effects into Eq. (1), rather we wish to evaluate the performance of certain approximations to the term structure by assuming the validity of Eq. (1). [9]

Having decided on the pricing function, we are faced with the problem of finding a suitable functional form for the discount factors v. A possible solution would be to use a polynomial [10] to approximate v. We choose to investigate two distinctive classes of functions, one defined by splines and another by exponential polynomials. These are briefly described below.

2.1. A spline representation of the term structure

The spline representation arises from problems associated with modelling the term structure via a simple kth degree polynomial. In particular, McCulloch (1971) was the first to introduce the use of a quadratic polynomial to estimate the discount function. He observed that this approximating function can lead to unstable forward curves and in a companion paper, McCulloch (1975), introduces a cubic spline, to avoid these unreasonable effects. This approximation to the discount function is defined as a family of cubics that are constrained to be continuous and smooth around each knot point. [11] Mathematically, this can be expressed as

McCulloch

$$v(\tau) = 1 + \sum_{j=1}^{k} \alpha_j f_j(\tau). \tag{2}$$

[8] See McCulloch (1975), Litzenberger and Rolfo (1984), Jordan (1984), Ronn (1987), Katz and Prisman (1991) for discussions on tax effects. Also, Dermody and Prisman (1988) and Dermody and Rockafellar (1991) investigate the effects of transaction costs and taxes. See Amihud and Mendelson (1991), for tests of liquidity effects.

[9] Recently, Elton and Green (1998) investigate the tax and liquidity effects in the pricing of US government bonds. They use daily data from the government interdealer market to demonstrate that the role of the liquidity effects is less pronounced than other studies have documented. They also confirm the results of Green and Odegaard (1997), who estimate a zero tax rate to conclude short that tax clienteles do not affect bond prices. However, further tests that use bond triplets (i.e., three bonds with the same maturity but different coupon) document evidence of the tax timing option and liquidity effects which are measured in pennies. Elton and Green (1998) conclude that earlier results reported by previous authors are irrelevant because of "increased inefficiencies in the Treasury market" or because of data problems influencing the calculation of the original estimates. Unfortunately, we do not have interdealer data on UK government bonds to take the tax timing and liquidity effects into consideration.

[10] Weierstrass' approximation theorem (see Williams, 1991) proves that the use of polynomial in approximating a continuous function over an interval on [0,1] results in a small degree of error.

[11] These are polynomial pieces that serve as the building blocks of the spline model. The number and location of the knots points is automatically implied by the data. I insert knot points at certain maturities such that there is an equal amount of data points between knots.

The discount function in Eq. (2) is written as a linear combination of basis functions [12] f_j which are produced by integrating the derivatives of low order polynomials. α_j are the coefficients to be estimated via linear least squares regression and τ denotes maturity. Henceforth, we refer to McCulloch's cubic spline given in Eq. (2) as McCulloch.

Shea (1984) points out that a cubic spline approach equivalent to McCulloch's still suffers from unstable and fluctuating forward rates. According to Shea (1984), a possible cure to irregularities in the forward curve is to impose simple restrictions of fixed proportions on the first derivatives of the discount function at different maturities. Shea (1984) then demonstrates that even these constraints may result in unreasonable shapes of the forward curve in places other than where constraints were imposed. An alternative approach would be to introduce a function which penalises the long end of the curve.

An alternative spline representation that is free from the aforementioned drawback of singularity is another set of basis functions, termed B-splines. As demonstrated by De Boor (1978) and Fisher et al., for a B-spline basis, the set of knot points is augmented around the start and end points of the curve. Then, a cubic spline is constructed from linear combinations of the above basis. Having constructed the cubic B-spline, we then approximate the discount function by adopting three functional forms:

Linear B-Spline

$$v(\tau) = \sum_{k=1}^{n} \beta_k \varphi_k(\tau), \tag{3}$$

Exponential B-Spline

$$v(\tau) = \exp\left(- \sum_{k=1}^{n} \beta_k \varphi_k(\tau) \right), \tag{4}$$

Integrated B-Spline

$$v(\tau) = \exp\left(- \int_0^{\tau} \sum_{k=1}^{n} \beta_k \varphi_k(\tau) \right), \tag{5}$$

where β_k are the parameters to be estimated and $\varphi_k(\tau)$ are the B-spline functions which are defined iteratively. I refer to Eqs. (3)–(5) as the linear B-spline, the exponential B-spline and the integrated B-spline functions respectively. Steeley (1990, 1991) is the first to apply the linear B-spline functions [13] on UK Treasury bond prices whereas Fisher et al. apply all B-spline functions on an equivalent US dataset. Furthermore, following Adams and van Deventer (1994), Fisher et al. (1995) and Tanggaard (1995), I intro-

[12] For definition of these functions refer to the appendix of McCulloch (1975).

[13] Steeley (1991) adopts an alternative representation of B-splines attributed to Powell (1981). He notes that the representation of B-splines defined by Fisher et al. is "non-trivially different" from Powell's representation (see Steeley (1991, footnote 4)). Steeley (1991) also develops confidence intervals to assess the accuracy of his discount functions.

duce a penalty function to produce the smoothest estimated discount function and avoid the problems outlined by Shea's critique. I estimate the linear, exponential and integrated B-splines via linear and non-linear least squares methods by minimising the sum of weighted squared errors plus the penalty: [14]

$$\min_{v(\tau)}\left[\sum_{i=1}^{N} w_i(P_i-\widehat{P}_i)^2+\lambda\int_0^T [v''(\tau)]^2\,\mathrm{d}\tau\right]. \tag{6}$$

The first term of Eq. (6) measures closeness to the data, P_i being the observed bond price and $\widehat{P}_i$ being the model price, weighting the distance by the inverse of the duration of the issue squared, $1/D_i^2$. The second term penalises curvature in the function. The parameter λ controls the trade-off between smoothness and goodness of fit. I choose λ as part of the estimation procedure by minimising [15] a criterion termed generalised cross-validation (GCV). [16] Bliss (1997) terms the Fisher et al. procedure as a "dubious choice for estimating term structures" due to its poor in-sample and out-of-sample performance. Bliss (1997) points out the problematic behaviour of the penalty function prompting the user to implement a maturity dependent λ. Unlike Bliss (1997), I modify the Fisher et al. (1995) procedure in two aspects. First, we introduce a weighting scheme to reflect the different effects that a given price change implies on the short and long end of the yield curve. [17] Second, I implement the GCV procedure using the approximation proposed by Wahba (1990). [18] I choose this approximation rather than the one proposed by Fisher et al. (1995), since the Fisher et. al. simplification to the GCV criterion is true only when the discount function is linear, as in Eq. (3). [19] Thus, I choose to be consistent in estimating the functional forms defined in Eqs. (3)–(5).

An alternative approach is to adopt the recommendations of Bliss (1997) by introducing a step penalty function that varies across maturities. Waggoner (1997) choose a three-tiered step function that transforms the minimisation problem in Eq. (6) to the following: [20]

VRP

$$\min_{v(\tau)}\left[\sum_{i=1}^{N}(P_i-\widehat{P}_i)^2+\int_0^T \lambda(\tau)[v''(\tau)]^2\,\mathrm{d}\tau\right] \tag{7}$$

[14] I thank Mark Fisher for suggesting a simpler method of computing the penalty function. Integration and differentiation are done through direct operations on the knot points.

[15] This minimisation is implemented using golden section line search technique in conjuction with a grid search to avoid local minima.

[16] Refer to Fisher et al. for a detailed application of this concept in this context. See also Wahba (1990) and Hastie and Tibshirani (1990) for a discussion on GCV.

[17] In this way, our method resembles resistant smoothing techniques. Härdle (1990, Chapter 6) provides details on such techniques.

[18] See Wahba (1990, Section 11.2), for further details.

[19] This is not our finding. It is reported in p. 25 of Fisher et al. (1995).

[20] To my knowledge, the properties of the variable roughness penalty have not been investigated. A natural question to ask is, how much is gained or lost in asymptotic accuracy when using such a function?

where the maturity dependent penalty λ is taken to be equal to

$$\begin{aligned} &0.1 \quad \text{if} \quad 0 \leqslant \tau \leqslant 1 \quad \text{or} \\ &100 \quad \text{if} \quad 1 \leqslant \tau \leqslant 10 \quad \text{or} \\ &100\,000 \quad \text{if} \quad 10 \leqslant \tau. \end{aligned} \tag{8}$$

Waggoner (1997) justifies the use of Eq. (7), which he terms variable roughness penalty method (VRP) in terms of the added flexibility introduced by Eq. (7) at the short end and the added stiffness at the long end. Although, Waggoner (1997) choose the constraints in Eq. (8) to correspond to the difference in maturity between bills, notes and bonds, its estimation for the UK market may prove to be inappropriate simply because its divisions are ad hoc. In fact, Anderson and Sleath (1999) introduce an alternative penalty function [21] for the UK market. Although their parameterisation is parsimonious, it is still ad hoc. They do not provide any evidence to support their specification as opposed to Waggoner (1997) step function defined in Eq. (8). Here, we implement Eq. (7) along with Eq. (8) in estimation and testing to uncover its implications for the UK market. Henceforth, we refer to the estimation method defined in Eqs. (7) and (8) as VRP.

2.2. A parsimonious representation of the term structure

An alternative approach to the shortcomings of the classical linear regression and cubic splines techniques is the methodology of Nelson and Siegel (1987). Unlike the spline class that models the discount function, this approach explicitly models the forward curve f, at any maturity τ as follows:

$$f(\tau) = \beta_0 + \beta_1 \exp\left(-\frac{\tau}{\tau_1}\right) + \beta_2\left(\frac{\tau}{\tau_1}\right)\exp\left(-\frac{\tau}{\tau_1}\right) \tag{9}$$

where β_0, β_1, β_2 and τ_1 are the parameters to be estimated. This specification resembles a constant plus Laguerre functions which consist of polynomials each multiplied by an exponential decay term. The yield to maturity on a bond is the average of forward rates and is obtained by integrating Eq. (9) over the term to maturity τ.

Nelson and Siegel

$$r(\tau) = \beta_0 + (\beta_1 + \beta_2)\frac{[1 - \exp(-\tau/\tau_1)]}{\tau/\tau_1} - \beta_2 \exp\left(-\frac{\tau}{\tau_1}\right). \tag{10}$$

The discount function in Eq. (10) becomes

$$v(\tau) = \exp(-mr(\tau)). \tag{11}$$

Nelson and Siegel (1987) illustrate that the assumed functional form of the model is flexible enough to accommodate forward curves with monotonic and humped

[21] Anderson and Sleath (1999) define the penalty function as $\log \lambda(\tau) = L - (L - S)\exp(-\tau/\mu)$, where L, S and μ are estimated by maximising the out-of-sample goodness of fit averaged over the sample period.

shapes. This method implies an intuitive explanation of the parameters: β_0 specifies the long rate to which the forward rate curve horizontally asymptotes, β_1 is the weight attached to the short term component, β_2 is the weight attached to the medium term component and τ_1 is the time constant measuring the point of the beginning of decay.

Nelson and Siegel suggest that a specification such as Eq. (10) provides a framework for generalisation to higher-order models. Svensson (1994) adopts their suggestion and extends Eqs. (9) and (10) to increase the flexibility of the model by adding two extra parameters. Hence, Eq. (10) becomes:

Svensson

$$r(\tau) = \beta_0 + \beta_1 \frac{[1 - \exp(-\tau/\tau_1)]}{\frac{\tau}{\tau_1}} + \beta_2 \left(\frac{[1 - \exp(-\tau/\tau_1)]}{\frac{\tau}{\tau_1}} - \exp\left(-\frac{\tau}{\tau_1}\right) \right) + \beta_3 \left(\frac{[1 - \exp(-\tau/\tau_2)]}{\tau/\tau_2} - \exp\left(-\frac{\tau}{\tau_2}\right) \right). \tag{12}$$

The discount function in Eq. (12) is still Eq. (11). Henceforth, we refer to the parsimonious specifications in Eqs. (10) and (12) as *Nelson and Siegel* and *Svensson* respectively.

2.3. Data and estimation of functions

We estimate the functions specified in Eqs. (2)–(5), (10) and (12) and the function (7) on daily UK treasury bill and bond (gilts) data obtained from Datastream. These are closing mid-prices that cover a period from beginning of January 1995 to beginning of January 1999, a total of 1046 observations. Our database contains market data on clean price, accrued interest, coupon, yield to maturity, Macaulay's duration, convexity and amount outstanding per bond in the sample. There are in total 65 coupon bonds of various maturities and redemption yields. To enable a consistent application of Eq. (1), I eliminate bonds that have embedded optionalities, for example callable and flower bonds. I also eliminate bonds with special liquidity problems, namely those bonds with less than one year to maturity and those with an amount outstanding of less than 1 billion pounds. I ensure that these liquidity constraints are binding for each day in the sample. Table 1 provides a summary of the database.

To represent the short end of the maturity spectrum, we preferred Treasury bills over the London interbank market. Despite the fact that the London interbank market may be more liquid than the Treasury bill market, LIBOR [22] rates are subject to an element of credit risk that we cannot ignore. Having modelled the short end with LIBOR rates and long end with gilt data, we would have ended up with term structures that have different default premiums at different maturities. The paucity of the data at the short end is problematic mainly due to the unavailability of Treasury bill

[22] LIBOR stands for London interbank bid and offer rate.

Table 1
Summary of UK Treasury bill and gilt data

	0.0833	0.25	0.25–2	2–5	5–6	6–7	7–9	9–11	11–15	15–20	20–25
Market Pr.	0.9947	0.9842	1.054	1.0734	1.0752	1.0604	1.1209	1.062	1.092	1.069	1.147
Coupon	0	0	0.1284	0.1053	0.086	0.085	0.0937	0.077	0.081	0.077	0.083
Yield	0.0646	0.0648	0.0631	0.0689	0.0669	0.071	0.071	0.07	0.071	0.072	0.072
Duration	0.0833	0.25	0.4777	1.5845	3.4802	3.757	4.730	6.323	7.350	9.200	10.297
Convexity	0.0797	0.2756	0.7059	4.2956	14.5755	16.180	24.039	40.801	53.749	81.704	101.422
N	1046	1046	256	864	791	1046	1046	1046	1007	1046	1037
K	1	1	5	13	5	2	3	4	5	3	2

The data covers a period from 1st of January of 1995 to 1st of January 1999, a total of 1046 observations. Reported figures are in mean values. Market prices correspond to dirty prices, which are clean prices plus accrued interest. Duration refers to Macaulay's duration. Figures under the 'N' column refer to number of trading days. These may depart from the total number of observations of 1046 due to redemption of bond or an introduction of a new issue. Figures under the 'K' column refer to number of actual bonds under each maturity. In total, there are 44 data points used to estimate the term structure functions.

data beyond three months to maturity. This is also recognised by the Bank of England and reflected in the Bank's estimates of the UK term structure. [23]

I estimate the Nelson and Siegel, Svensson, the exponential B-spline, the integrated B-spline functions and VRP function using BFGS algorithm. [24] I apply McCulloch and linear B-spline functions using ordinary least squares. I choose to minimise the distance between squared price errors [25] weighted by the inverse of the duration of the issue squared. This weight function best adjusts for the differential importance of small price changes at different maturities on estimates of the yield curve. An error in price of three-month treasury bill would not have the same impact as the error in the price of a fifteen year bond. If we assume an equal weight, the pricing of long term issues would be less accurate than the pricing of short term maturity issues due to increasing duration. Similar weighting schemes are adopted by Vasicek and Fong (1982) and Bliss (1997). My main purpose is to use yield or forward curves as an input to a no-arbitrage model of the term structure to price and hedge interest rate options. For purposes other than asset pricing, such as monetary policy, the different yield curve models would be estimated by minimising the distance of weighted squared yield errors.

Fig. 1 depicts the UK term structure of interest rates for representative dates of the sample. Each panel plots the yield curve estimates implied by Nelson and Siegel (NS), Svensson (SV), McCulloch (McC), linear B-Spline (LinB), exponential B-Spline (ExpB), integrated B-spline (IntEx) and VRP functions. Panels A to B display an upward sloping UK yield curve that shifted to downward sloping by the begin-

[23] The Bank's estimates are publicly available and do not incorporate maturities below two years. See Anderson et al. (1996) for further details on the Bank's model.

[24] The name BFGS refers to Broyden–Fletcher–Goldfarb–Shanno minimisation algorithm. This is described in Press et al. (1992).

[25] Market prices are dirty prices defined as the sum of clean price and accrued interest.

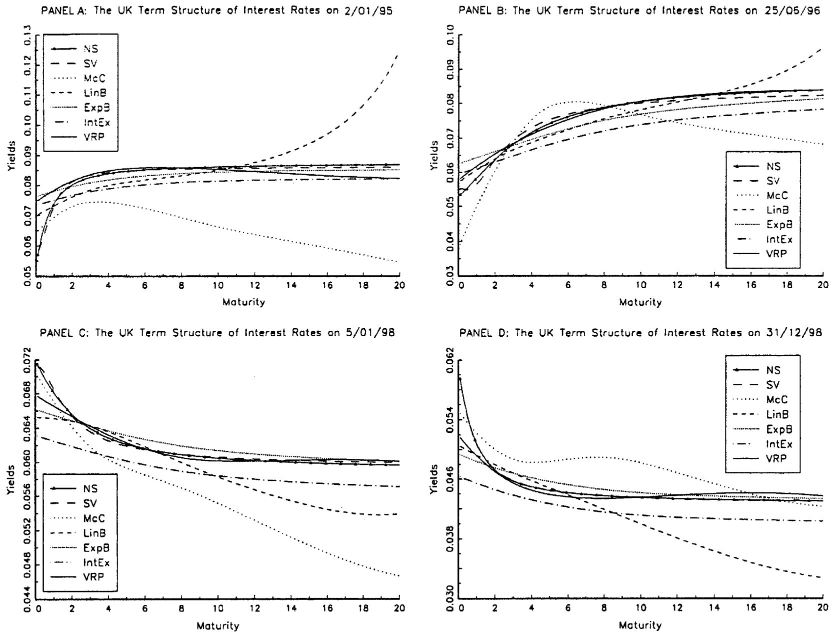

Fig. 1. The fitted curves implied by the seven estimation methods.

ning of 1998 (see Panel C). Panel D shows that the UK yield curve remained downward-sloping for the rest of the sample period.

A striking conclusion from a visual inspection of these panels is the behaviour of McCulloch and of the linear B-spline. Although all other methods produce comparable to indistinct shapes, McCulloch and the linear B-spline experience great difficulty in fitting the long end of the curve. This discrepancy of the two methods characterised by a linear functional form may be due to misestimation or mispecification of the functional form. This preliminary result reinforces the suspicions of Bliss (1997) and strengthens the results of Chambers et al. (1984).

2.4. In-sample and out-of-sample results

I now turn to the discussion of results from the estimation of all competing methods. We concentrate on the analysis of in-sample fitted price errors. I also report results from the pricing of bonds that have been kept out of the estimation sample. These also incorporate bonds that did not fulfil the daily liquidity constraints. On average, the out-of-sample data amounts to 21 coupon bonds.

Table 2 presents the weighted root mean square error for all seven estimation methods. I report mean, maximum, minimum and standard deviation values. A first look at Table 2, reveals the best function in terms of goodness of fit. The weighted mean square error of the Svensson function is the smallest, and amounts to 173.21 basis points. This is closely followed by the Nelson and Siegel function and the

Table 2
Weighted root mean square error across functions

	Nelson and Siegel	Svensson	McCulloch	Linear B-spline	Exponential B-spline	Integrated B-spline	VRP
Mean	173.21	100.00	529.15	387.30	479.58	678.23	264.58
s.d.	300	264.58	458.26	264.58	556.78	728.01	346.41
Min	0	0	200	244.95	0	469.04	100
Max	994.99	969.54	1135.78	748.33	938.08	714.14	877.50

All reported values are expressed in terms of basis points. For example, under the Nelson and Siegel the mean value of 173.21 basis points refers to a root mean square value of 0.017321. The *mean square value* of this mean value is equivalent to 0.0003. We report the mean, the standard deviation, the minimum and maximum root mean square values across all seven functions.

unweighed VRP function. Based on this metric alone, we observe that all competing functions perform well, with the integrated B-spline trailing last with a weighted root mean square error amounting to 678.23 basis points. [26]

Table 3 gives the pricing performance of competing yield curve estimation techniques in terms of explaining variation in market prices. In this table, we classify each bond in the sample into categories depending upon their maturity. The minimum, maximum, and mean values are tabulated along with the unweighed mean absolute pricing error for each category. [27] The magnitude of errors is generally small at the short end and increase at the long end. McCulloch is the exception. It produces the worst and best mean absolute error at the short and long maturities respectively. Although, the smoothing functions of Fisher et al. are unable to fit the long end, these produce a better fit for maturities up to 15 years. [28] This is also true for the VRP. For maturities up to 5 years, we observe that the parsimonious specifications are superior to their spline counterparts. For maturities above 5 years, McCulloch produces the minimum mean absolute error MAE. However, the results of long maturity issues should be interpreted with caution. There is a danger of overfitting the long end of the term structure due to the lack of data at the long end.

This picture changes when I turn to the out-of-sample results. Table 4 presents results from pricing coupon bonds that are excluded from the estimation procedure. Here, I report the mean and MAE values for all competing functions. Once again, I arrange bonds by maturity. Overall, Svensson function performs better on the basis of lower total MAE, with exponential B-spline, Nelson and Siegel, and VRP functions trailing as second best. Comparing MAE values across maturities reveals the marginal superiority of exponential B-spline at 4–5 year and of the VRP function at the 5–10 year sub-intervals, while the Svensson function produce better MAE

[26] Unfortunately, we do not have bid-ask prices to get a feeling of how close the spread is and whether the above errors are within the bid-ask spread.

[27] All values are in basis points.

[28] Bliss (1997) suggested "lowering the tuning parameter" would improve the sample fit. I minimise the GCV criterion over a range of starting values. I experiment with large values (approaching 10^5) to small (approaching 10) in association with a grid search. I choose the smoothness (or tuning) parameter that satisfies the GCV criterion and leads to a better in-sample and out-of-sample fit.

Table 3
Cross-sectional analysis of in-sample residuals

	1m	3m	3m–2y	2y–5y	5y–7y	7y–8y	8y–9y	9y–10y	10y–15y	15y–20y	20y–25y
(A) Nelson and Siegel residuals											
MAE	1	5	25	31	41	43	48	54	62	102	123
Mean	0	1	14	3	16	−7	7	10	−7	−40	−16
Total MAE	49										
(B) Svensson residuals											
MAE	1	3	16	77	40	39	33	33	37	59	76
Mean	0	0	2	6	37	3	17	22	1	−25	−8
Total MAE	34										
(C) McCulloch residuals											
MAE	139	138	144	77	31	26	23	17	19.5	14	7
Mean	−137	−136	64	22	−20	5	−22	11	−1	2	1
Total MAE	54										
(D) Linear B-spline residuals											
MAE	3	10	31	63	88	134	118	111	135	440	1861
Mean	0	−1	6	−60	−24	−107	−90	−98	−101.5	286	1861
Total MAE	216										
(E) Exponential B-spline residuals											
MAE	6	17	58	37	65	68	82	104	120	160	1203
Mean	2	7	54	−7	24	−19	5	14	12	19	−1104
Total MAE	145										
(F) Integrated B-spline residuals											
MAE	52	15	57	80	71	62	65	801	118	265	3508
Mean	−52	−15	−57	−80	−64	−18	35	67	116	265	1181
Total MAE	490										
(G) VRP residuals											
MAE	4	13	57	36	41	49	67	94	109.5	294	1036
Mean	2	5	55	16	24	29	60	85	89.5	267	−1016
Total MAE	142										

The reported values depict the in-sample pricing performance of the seven competing estimation methods for different maturity ranges. For each maturity category, I report the unweighed mean absolute pricing error and the mean pricing error. All values are expressed in terms of basis points.

at the remaining sub-intervals. An exception is the 3–6 month sub-interval, over which Nelson and Siegel performs marginally better than Svensson function. McCulloch is the worst performer for all subintervals.

Fig. 2 exhibits the in-sample mean pricing errors at Panel A and the out-of-sample mean pricing errors at Panel B for each of the competing functions. Panel A clearly shows the ability of the linear, exponential and integrated B-splines to price

Table 4
Cross-sectional analysis of out-of-sample residuals

	3m–6m	6m–1y	1y–2y	2y–3y	3y–4y	4y–5y	5y–10y	10y–15y
(A) Nelson and Siegel residuals								
MAE	28	54	145	51	100	112	161	171
Mean	8	43	101	20	40	83	39	25
Total MAE	71							
(B) Svensson residuals								
MAE	36	10	16	14	37	48	51	124
Mean	19	10	5	−14	−12	1	11	−41
Total MAE	11							
(C) McCulloch residuals								
MAE	103	43	132	1044	906	352	535	1153
Mean	−99	−31	−103	−2	−903	−352	−535	−842
Total MAE	339							
(D) Linear B-spline residuals								
MAE	83	36	54	172	203	106	108	223
Mean	−40	31	−39	−132	−197	−106	−108	−187
Total MAE	82							
(E) Exponential B-spline residuals								
MAE	88	65	56	99	102	29	29	187
Mean	−15	60	−19	8	−34	−29	−29	−82
Total MAE	32							
(F) Integrated B-spline residuals								
MAE	86	64	47	114	184	135	137	761
Mean	−32	64	−25	61	48	135	14	50
Total MAE	54							
(G) VRP residuals								
MAE	85	63	51	123	102	53	17	166
Mean	−16	58	−18	33	−64	−53	6	−38
Total MAE	39							

The reported values depict the out-of-sample pricing performance of the seven competing estimation methods for different maturity ranges. For each maturity category, I report the unweighed mean absolute pricing error and the mean pricing error. All values are expressed in terms of basis points.

accurately Treasury bonds up to maturities ranging between 10 and 15 years. For Treasury bonds with maturities beyond 15 years, these functions produce the highest mean pricing errors. Panel B clearly highlights the inaccuracies of McCulloch.

The implication of these results is twofold. First, the functional form of the models is important. Those functions with an exponential form, namely Nelson and Sie-

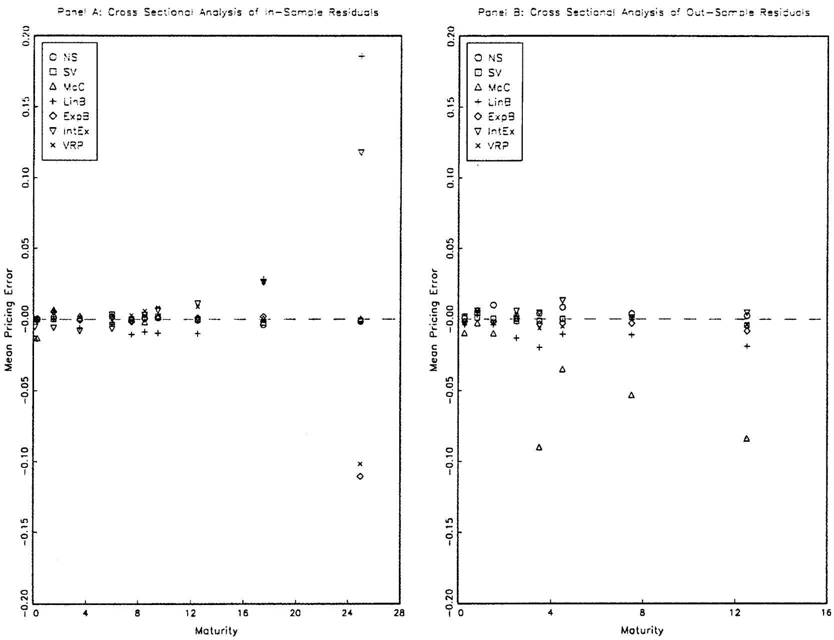

Fig. 2. Cross-sectional analysis of residuals.

gel, Svensson, exponential B-spline and VRP are better performers than their linear counterparts, namely McCulloch and linear B-spline. The B-spline specifications demonstrate enough flexibility at the short end both in-sample and out-of-sample. This performance may be attributed to the weighting of the pricing errors, which serves as an alternative to a maturity dependent penalty. Although the VRP performs better in-sample against all the B-spline specifications, it underperforms slightly the exponential B-spline at the out-of-sample. I suspect that a different functional form other than the step function in Eq. (8) would improve the out-of-sample and in-sample performance of the VRP.

Second, some of this mispricing may be due to misspecification and/or mis-estimation of the parsimonious and spline functions. I formalise these implications by using two tests. This set of tests is borrowed from the work of Wu (1995) and Sercu and Wu (1997) on the information content of bond residuals. I formalise these tests in the following section.

3. Model specification tests

These tests investigate the degree of mispricing by taking positions in bonds that are mispriced according to a given yield curve estimation method, buying bonds that are underpriced and short selling bonds that are overpriced. Although, both of these tests are developed by Wu (1995) and Sercu and Wu (1997), this study is the first to apply this methodology only to term structure estimation methods using daily data from the UK bond market. The procedure described in the following subsection presents these tests. [29]

3.1. Trading and filter rule tests

The functions underlying the parsimonious and spline class of yield curve estimation functions are simply approximation methods. There are no economic assumptions underpinning the use of such functions. I could discriminate between these methods only by their shortcomings and mathematical implications. Nonetheless, such an analysis would prevent the end user from learning more about the implications of residuals arising from the estimation of such functions. Ideally, we want to identify the degree of mispricing from the use of these alternative representations. Some of this mispricing may be due to model misspecification. The last section reveals that all functions perform fairly well, but favours the parsimonious functions of Nelson and Siegel and Svensson, the integrated B-spline and the VRP. In this section, my main purpose is to recognise the best function in terms of its ability to recognise genuine mispricing and thus form abnormal profits, even when I correct for wide market movements. [30]

I apply a contrarian trading strategy by buying (selling) issues that are considered underpriced (overpriced) and weighting each by a factor proportional to the size of mispricing. In this respect, I adopt the work of Wu (1995) and Sercu and Wu (1997), who propose three alternative benchmarks of ARs to verify whether the competing functions allow any profitable trading strategies.

All three benchmarks measure AR and remove the normal component of the holding period returns by adjusting for market movements. These are designed to adjust for parallel shifts and twists in the term structure. The first benchmark measures ARs as the difference between the market holding period return, $\mathrm{HP}_{i,t}$ and the expected holding period, $E_t[\mathrm{HP}_{i,t}|\Phi_{t-1},\Phi_t]$ for a function underlying a given yield curve estimation technique:

$$\mathrm{AR}_{i,t} = \mathrm{HP}_{i,t} - E_t[\mathrm{HP}_{i,t}|\Phi_{t-1},\Phi_t] \tag{13}$$

[29] An alternative way of testing for mispecification is the non-nested test of Davidson and MacKinnon (1981). This test concentrates on testing the functional specification of one yield curve estimation method in the presence of an alternative which sets to explain the same data. Preliminary results from the application of this test render the test statistics inconclusive. This shows that it is hard to discriminate between the competing yield curve estimation techniques. Detailed results are available from the author upon request.

[30] An alternative study by Steeley (1992) investigates the efficiency of the UK Treasury market before and after the Big Bang. His sample covers the period between October 1985 and October 1997. His findings support that the market improved in efficiency after the Big Bang since the mean absolute errors between the market and model prices are smaller after the Big Bang rather than before.

where

$$E_t[\mathrm{HP}_{i,t}|\Phi_{t-1},\Phi_t] = \frac{\widehat{P}_{i,t} - \widehat{P}_{i,t-1} + \mathrm{Coupon}}{\widehat{P}_{i,t-1}}. \tag{14}$$

Note that the information set Φ_t contains the model parameters obtained at day t. Although, this definition designed to absorb the day to day movements in market prices of bonds, it is problematic, since it assumes that the 'true' function is the function whose performance is being tested. This leads to the definition of two other benchmarks. One measures ARs based on the market model as explicitly described in Elton and Gruber (1991). Following Sercu and Wu (1997), we define the AR as follows:

$$\mathrm{AR}_{i,t} = \mathrm{HP}_{i,t} - [\alpha_{i,t-1} + \beta_{i,t}(\mathrm{HP}m_{i,t} - \alpha m_{i,t})] \tag{15}$$

where $\alpha_{i,t-1}$ is the annualised continuously compounded yield on the ith bond calculated on a daily basis and $\alpha m_{i,t}$ is the similarly derived yield on the market portfolio. $\beta_{i,t}$ is the ratio of durations between the duration of the ith bond and the duration of the market portfolio. $\mathrm{HP}m_{i,t}$ denotes the market portfolio holding period return, which we choose to approximate with the Datastream benchmark index. This index is value weighted and includes all live bonds in the UK market over the sample period in question. The Datastream information also include the gross price level and duration of the index.

The above measure only accounts for parallel shifts in the term structure. I should account for both parallel shifts and twists in the underlying term structure. The last benchmark matches the return of a portfolio with the value, duration and convexity of the underlying bond. This benchmark is measured using three portfolios. The first consists of one month and three month treasury bills, the second portfolio consists of bonds that mature between 1 and 8 years and the third portfolio is made up of bonds with more than 8 years to maturity.

Having built the three benchmarks, I proceed to design the trading and filter rule tests in accordance with Sercu and Wu (1997). The trading rule test consists of purchasing all underpriced bonds and short selling all overpriced bonds on a given trading day. Both "purchase" (denoted by subscript p) and "sell" (denoted by subscript s) portfolios are weighted by the size of the mispricing ($\mathrm{RES}_{i,t-i-L}$, where L is the time delay) and adjusted by all the bonds in the sample at any trading day, such that the mean AR is

$$\overline{\mathrm{AR}}_{y,t} = \sum_{i=1}^{N} \frac{\mathrm{RES}_{i,t-i-L}}{\sum_{i=1}^{N_y} \mathrm{RES}_{i,t-i-L}} I_{i,t-i-L}\mathrm{AR}_{i,t} - \frac{1}{N_{o,t}} \sum_{i=1}^{N_{o,1}} \mathrm{AR}_{i,t}, \quad y = \mathrm{p},\mathrm{s}, \tag{16}$$

where I_i is an indicator taking value of $-1(+1)$ if the ith bond is overpriced (underpriced), $\mathrm{AR}_{i,t}$ is defined as one of the benchmarks and $N_{o,t}$ denotes all outstanding bonds in the sample on day t. Hence, I end up with six mean AR series, reflecting three different benchmarks each for the "buy" and sell portfolios. Eq. (16) implies that there is a time delay between the construction of the term structure and the implementation of the trade rule. This time delay is denoted by L and is varied over a window of 0 to 5 days to eliminate any bid ask bias. To ensure that the remaining

AR is merely due to noise and is not model specific, we deduct the AR from holding an equally weighted portfolio of all bonds. Finally, I compute the average AR of the buy and sell portfolios (subscript c).

To test if the strategy outperforms the naive buy and hold portfolio, I compute the cumulative mean AR for the whole sample period:

$$\mathrm{CAR}_{y,t} = \sum_{i=1}^{1046} \overline{\mathrm{AR}}_{y,t}: \quad y = \mathrm{p, s, c}. \tag{17}$$

The t-statistic is based on the consistent heteroscedasticity and autocorrelation covariance matrix of Andrews (1991). [31]

The contrarian strategy above places equal weight on all mispriced bonds, no matter how large the mispricing. The filter test is designed to vary the weighting scheme according to the size of the bond mispricing. In this way, I can verify the results of the trade rule test. The filter rule test identifies mispricing by placing a filter on the bond residuals. If a bond residual is positive and larger than a number of basis points, i.e., the filter, I short sell the bond. On the other hand, if a bond residual is negative and smaller than a number of basis points, I buy the bond. We then vary the filter size. For each filter rule size, we report results for purchase, sell and combined portfolios. The mean AR $\overline{\mathrm{AR}}_{y,t}$, from the filter rule adjusted for broad market movements according to Eqs. (13) and (15) is as follows:

$$\overline{\mathrm{AR}}_{y,t} = \frac{1}{N_t} \sum_{i=1}^{N} \mathrm{AR}_{i,t} I_{i,t-i}. \tag{18}$$

Note that $\mathrm{AR}_{i,t}$ is the AR realised over two consecutive days and $I_{i,t-1}$ is the buy/sell indicator. [32]

3.2. Analysis of results of the contrarian and filter rule tests

Table 5 presents the results of the trading strategy for the best and worst performing functions. The best performing function is the Nelson and Siegel while McCulloch performs the worst among the seven competing functions. [33] The results are

[31] I calculate standard errors using Andrews (1991) and Andrews and Monahan (1992) heteroscedasticity and autocorrelation consistent covariance matrix estimator. One of the main advantages of this method over other popular methods like Gallant (1987), Newey and West (1987) and White (1980) is that the bandwidth or lag parameter is automatically implied by the data.

[32] The t-test statistic is calculated using Andrews (1991) heteroscedasticity and autocorrelation consistent covariance matrix. I follow Bjerring et al. (1983) and Sercu and Wu (1997). I lose 24 days to calculate the standard deviation. The ARs are then standarised into a Student's variable, $Z_t = (\overline{\mathrm{AR}}_t/\sigma_t)$, under the null hypothesis that the filter rule gains no positive returns. The t-statistic converges to a unit normal when the sample is sufficiently large and is equal to $(1/\sqrt{T-26}) \sum_{i=25}^{1046} Z_t$.

[33] Detailed results of the trading strategy for the Svensson, the linear B-spline, the exponential B-spline, the integrated B-spline and the VRP are available from the author upon request.

Table 5
Profits and losses from contrarian trading strategies

Lag	Benchmark: M. Implied Return			Benchmark: Market Model			Benchmark: DCM Model		
	Buy	Sell	Both	Buy	Sell	Both	Buy	Sell	Both
(A) CARs formed with reference to Nelson and Siegel residuals									
0	**0.5009**	0.278	**0.3895**	0.0095	0.0154	0.0125	0.1616	**0.2341**	**0.1978**
	(5.0051)#	(2.7476)#	(3.9974)#	(1.383)	(1.9849)**	(1.7226)*	(1.8967)*	(2.2725)**	(2.1179)**
1	0.1994	0.0729	0.1361	0.0078	0.014	0.0109	0.0356	0.0458	0.0407
	(2.2406)**	(0.7279)	(1.4731)	(1.1338)	(1.7846)*	(1.4955)	(0.4429)	(0.4423)	(0.4445)
2	0.1548	0.0821	0.1185	0.0141	0.0124	0.0133	0.0763	0.0562	0.0633
	(1.7955)*	(0.8241)	(1.2856)	(1.4285)	(1.5844)	(1.6071)	(0.8967)	(0.5467)	(0.7082)
3	0.0596	0.09	0.0748	0.0138	0.0115	0.0126	0.0522	0.0764	0.0643
	(0.7332)	(0.8986)	(0.8274)	(1.3857)	(1.4739)	(1.5329)	(0.6261)	(0.7409)	(0.6929)
4	0.0681	0.0701	0.0691	0.0154	0.0102	0.0128	0.0731	0.0606	0.0669
	(0.8386)	(0.7047)	(0.767)	(1.4908)	(1.314)	(1.5256)	(0.8774)	(0.5895)	(0.722)
5	0.05	0.0693	0.0597	0.0134	0.0088	0.0111	0.0437	0.0513	0.0475
	(0.6213)	(0.6975)	(0.6654)	(1.2902)	(1.1229)	(1.3116)	(0.5291)	(0.4995)	(0.5153)
(B) CARs formed with reference to McCulloch model residuals									
0	−0.0403	−0.0649	−0.0549	0.0205	0.0181	0.0193	−0.0417	−0.0711	−0.0564
	(−5.7018)#	(−4.8287)#	(−5.2028)#	(4.031)#	(3.2479)#	(3.661)#	(−3.4246)#	(−4.746)#	(−4.1797)#
1	−0.0407	−0.0654	−0.0531	0.0202	0.0179	0.019	−0.042	−0.071	−0.0565
	(−5.7405)#	(−4.7829)#	(−5.1931)#	(3.9641)#	(3.1926)#	(3.5976)#	(−3.4502)#	(−4.7556)#	(−4.1965)#
2	−0.0408	−0.0662	−0.0535	0.0198	0.0177	0.0187	−0.0421	−0.071	−0.0566
	(5.7302)#	(4.9345)#	(5.2851)#	(3.8904)#	(3.1434)#	(3.5357)#	(−3.4553)#	(−4.754)#	(−4.1969)#
3	−0.0409	−0.0654	−0.0532	**0.0198**	0.0176	0.0187	−0.042	−0.0706	−0.0563
	(5.7475)#	(4.8747)#	(5.2514)#	(3.9058)#	(3.1366)#	(3.5391)#	(−3.4463)#	(−4.7223)#	(−4.1756)#
4	−0.0411	−0.0654	−0.0533	**0.0197**	0.0176	0.0187	−0.0424	−0.0707	−0.0566
	(−5.7829)#	(−4.8629)#	(−5.2575)#	(3.8746)	(3.1366)#	(3.5222)#	(−3.8406)#	(−4.7373)#	(−4.1979)#
5	−0.0409	−0.0649	−0.0529	**0.0199**	0.0178	0.0188	−0.0424	−0.0695	−0.0559
	(−5.7598)#	(−4.8828)#	(−5.2643)#	(3.9139)#	(3.1618)#	(3.5553)#	(−3.4798)#	(−4.6949)#	(−4.1714)#

The results above are based on the trading of residuals from the daily cross-sectional estimation of each one of the aforementioned models. The trading takes place between 1 January 1995 and 31 of December 1998, a total of 1046 trading days. To implement the trading rule at day t-1, we buy (sell) a bond with negative (positive) residual and weight each position in proportion to the initial amount of mispricing. We implement the trading rule with a varying lag, from zero to five days. CARs are defined as the difference between the market holding period return and one of the following benchmarks: the method implied return, the return on the market model and the return on a portfolio matched in terms of value, duration and convexity. Moreover, these are adjusted by the AR from holding buy-hold portfolio as defined by the residuals of each model. The figures in parenthesis are t-statistics computed by the Andrews (1991) automatic lag selection procedure. Figures in bold represent significant return or lowest significant losses across benchmark models.

Significance at 1% level.

** Significance at 5% level.

* Significance at 10% level.

classified according to the benchmarks employed, the model implied AR, the market model AR defined in Eq. (15) and the value, duration and convexity matched portfolio (DCM). I report in bold the highest statistically significant profits among all competing functions classified in terms of buy, sell and combined portfolios for each time delay and benchmark. I calculate the standard errors for each computed AR series using the prewhitened quadratic spectral kernel of Andrews (1991) and Andrews and Monahan (1992) to estimate a heteroscedasticity and autocorrelation consistent covariance matrix. We report the t-statistics in brackets.

I deem a term structure estimation technique superior to any other alternative technique, if it produces statistically significant cumulative abnormal returns (CARs) under all three benchmarks. With the method implied return benchmark, we implicitly assume the validity of each of the competing functions without capturing any of the wide market movements. While the market model AR escapes from this criticism, it only allows for parallel shifts in the yield curve. The third benchmark, the DCM portfolio accommodates both parallel shifts and twists in the underlying term structure. Furthermore, I correct each AR by the CAR of a buy and hold portfolio. If the apparent mispricing is due to the misspecification or mis-estimation of a particular function, I do not expect this mispricing to be informative. On the other hand, if this mispricing corresponds to genuine movements, we would expect that this mispricing should not persist over time.

First, I observe ARs in excess of the buy and hold strategy across all method implied portfolio returns. Statistically significant returns vary from −0.066% to 0.50% over a period of 1045 days. If trading is immediate (Lag = 0), the Nelson and Siegel function produces the highest statistically significant AR under the buy and combined portfolios. [34] On the other hand, McCulloch produces the lowest statistically significant AR of −0.066%. Under this function, losses are statistically significant across all portfolios irrespective of the time delay between the estimation of the term structure and implementation of the trade.

Second, under the market model and DCM benchmarks, statistically significant returns are of the same magnitude as the model implied ARs and vary between −0.071% and 0.23%. For the parsimonious representation of Nelson and Siegel, the market model and DCM benchmarks produce significant ARs in excess of the buy and hold strategy, only if trade is immediate (Lag = 0). Under the DCM benchmark, the Nelson and Siegel function performs better than any other competing alternative to produce the highest statistically significant profit for the sell and combined portfolios. However, when I vary the time delay between one to five days, our findings suggest that using the DCM benchmark to account for parallel shifts and twists in the term structure eliminates any significant abnormal profits irrespective of the function employed. Under this benchmark, McCulloch underperforms all other competing functions by producing statistically significant losses.

Third, under the market model McCulloch produces the highest significant profits among all other competing alternatives. [35] If these profits were due to unspecified effects, significant profits would also be formed in the model implied and DCM benchmarks. However, for the McCulloch function we report profits only under the

[34] Under the method implied portfolio method, the integrated B-spline produces the highest AR under the sell portfolio when tradind is immediate. This function remains the best performer across the buy, sell and combined portfolios of this method, even when we delay to implement the trade by five days.

[35] This is also true for the linear B-spline and the integrated B-spline.

Table 6
CARs of filter rule tests

bp	Benchmark: return from fitted prices			Benchmark: market model		
	Buy	Sell	Both	Buy	Sell	Both
(A) Nelson and Siegel function—lag zero						
0	**2.4499**	**2.2503**	**2.3501**	**0.0454**	0.0514	0.0484
	(14.7197)#	(12.6573)#	(13.3204)#	(4.4341)#	(2.5588)#	(3.4474)#
25	**2.8054**	**3.3983**	**3.1019**	0.0862	0.0621	0.1242
	(12.7445)#	(10.3299)#	(13.4812)#	(5.4497)#	(5.4543)#	(6.1535)#
50	**1.7221**	**1.0271**	**1.3746**	0.0565	0.1137	0.0851
	(12.3062)#	(5.0317)#	(13.3744)**	(7.2584)#	(4.4557)#	(5.6228)#
75	1.0575	**0.542**	0.7998	0.0273	0.034	0.0306
	(16.4758)#	(5.5729)#	(19.8703)#	(4.3841)#	(1.4874)	(1.5154)
100	0.4238	**0.2074**	0.3156	0.0017	0.0141	0.0079
	(31.2276)#	(4.8911)#	(18.8549)#	(20.6741)#	(7.2532)#	(11.0474)#
(B) McCulloch function—lag zero						
0	0.038	−0.0433	−0.0027	0.0161	0.0576	0.0368
	(10.1743)#	(−3.9385)#	(−0.02831)	(5.9728)**	(6.9578)**	(6.9532)#
25	0.0632	−0.0964	−0.0166	0.0109	0.0136	0.0123
	(10.2293)#	(−3.1207)#	(0.6715)	(3.8233)**	(0.9462)	(2.9468)#
50	0.0586	−0.181	−0.0612	0.0093	−0.0396	−0.0151
	(6.3342)#	(−1.7578)*	(1.3516)	(0.9397)	(−0.5383)	(0.3858)
75	0.055	−0.2085	−0.0768	−0.002	−0.0767	−0.0394
	(4.2004)#	(−2.0813)**	(−0.3709)	(−0.6615)	(−3.0293)**	(−3.5275)#
100	0.01	−0.2643	−0.1272	−0.0009	−0.0502	−0.0256
	(2.237)**	(−5.6987)**	(0.1521)	(−1.0985)	(−2.5139)**	(−3.2999)#

The sample period of the filter rule tests cover 1021 trading days (24 days are lost to allow for computation of standard errors). Bonds are classified into a buy or sell portfolio, given that their residuals exceed a number of basis points, the filter. Two benchmarks are used to adjust normal return, (i) the return implied from fitted prices, (ii) the return on the market model. Furthermore, returns are adjusted by the return on a buy and hold portfolio. T statistics are in parenthesis. Standard errors are calculated using automatic lag selection procedure of Andrews (1991). The largest significant positive (negative) value are in bold.

Significance at 1% level.

** Significance at 5% level.

* Significance at 10% level.

market model benchmark. I conclude that this function suffers from mis-estimation or misspecification. [36]

I now turn to the filter rule test which places equal weighting on all mispriced bonds as opposed to the trading rule test which weights each bond by the size of mispricing (see Eq. (16)). Table 6 presents the results from the implementation of this

[36] I obtain similar results for the linear B-spline. This lets us to conclude that the linear B-spline also suffers from mis-estimation or misspecification. This effect is less pronounced for the exponential B-spline, which generates profits only at the sell portfolio. For the integrated B-spline, I report significant abnormal profits only for two of the benchmarks, namely the model implied and the market model. Adjusting ARs for parallel shifts and twists in the term structure, I observe no significant abnormal profits. These outcomes highlight the problematic behaviour of the penalty function and confirm the suspicions of Bliss (1997).

test. The ARs are defined by two alternative benchmarks, the model implied return of Eqs. (13) and (14) and the market model return of Eq. (15). I choose to apply only these two benchmarks because almost all functions do not generate any ARs under the DCM portfolio. I implement the filter rule tests with a zero time delay [37] for all seven term structure methods. Here, I report results for the Nelson and Siegel and McCulloch. I also place in bold the highest statistically significant profits reported for each buy, sell and combined portfolios for every benchmark and function.

First, I observe statistically significant ARs on both benchmarks, the method implied return and market model. For the Nelson and Siegel profits initially increase as the filter size becomes larger and then diminish towards zero. [38] This implies that it is more profitable to trade on medium-sized residuals rather than trade on large-sized ones. Looking at Table 6 (Panel (A)), the highest AR for the Nelson and Siegel under the buy, sell and combined portfolios is 2.8054%, 3.3983% and 3.1019% respectively when the benchmark return is the model implied return and the filter size is 25 basis points (bp). Under the market model, I also form the highest ARs with a filter size of 25 bp. Are these results in agreement with the results reported by the trading rule test? The contrarian trading rule test places more weight on those bonds that deviate further from their fitted values rather than placing equal weights for each bond in the portfolio. If the trading rule's weighting scheme was an optimal one, I would expect the filter rule test to generate the highest abnormal profits for the biggest filters with a size of 75 and 100 bp. Results indicate that the weighting scheme adopted by the trading rule test is not optimal. [39] This suggests that the trading rule understates ARs generated by these functions. [40]

Second, statistically significant losses on the sell portfolio of the model implied return and market model persist and do not decline to zero only for McCulloch's cubic spline.

Overall, the filter rule results indicate that weighting by the size of mispricing is not optimal. In other words, the filter rule illustrates that profits produced by all functions [41] using the contrarian trading rule are understated. I find that for certain functions returns initially increase as the filter size gets larger and then decrease

[37] I also implement the filter rule test with a delay of one day. I find that results are not different from the ones presented here.

[38] This is also true for the Svensson, the Integrated B-spline and the VRP.

[39] This result is also reported by Sercu and Wu (1997).

[40] This is not the case for the linear and exponential B-spline functions. Both functions generate the highest ARs for the buy and combined portfolios of the model implied return and for the sell portfolio of the market model return when the filter size is at 100 bp and at 75 bp respectively. This implies that trading in bonds with large-sized residuals is more profitable than trading in bonds with medium-sized residuals. In other words, by investing more on bonds whose price deviates away from their fitted value estimated by either the linear or exponential B-spline functions, one may generate abnormal profits. This observation confirms that the weighting scheme adopted by the trading rule test is indeed optimal. However, this is only true when we form a trading strategy under the buy and combined portfolios of the model implied return and for the sell portfolio of the market model using either the linear or the exponential B-spline functions.

[41] Exceptions to this result are the linear and exponential B-splines.

which implies that large sized residuals lead to lower average returns. Results also reveal that the performance of the two parsimonious functions in terms of generating the highest average profits under the model implied return is better than McCulloch's cubic spline, the linear, exponential, integrated B-spline and VRP functions. [42]

4. Conclusions

I compare different methods of estimating the term structure of interest rates on daily UK Treasury bill and gilt data. I examine the Nelson and Siegel and Svensson functions, McCulloch's cubic spline, the linear, exponential and integrated exponential B-splines and the VRP method, a total of seven methods.

The major conclusions stemming from in-sample and out-of-sample analysis of residuals suggest that the parsimonious specifications and VRP method perform better than the linear spline counterparts. In terms of the out-of-sample performance, the non-linear B-splines and the VRP function produced a lower mean absolute pricing error than the Nelson and Siegel specification, but these functions are second best to Svensson. This suggests that the specification of the functional form of the model is important for pricing.

To verify this conclusion, I conduct two further tests to investigate genuine pricing errors. The application of the contrarian trading rule indicates that the best term structure estimation techniques are the parsimonious ones. Of the spline specifications, the integrated B-spline and the VRP function produce equally good results. However, the implementation of the filter rule shows for all functions with the exception of the linear and exponential B-splines that increasing our stake in a bond on the basis of the initial mispricing understates profits.

Acknowledgements

I would like to thank two anonymous referees, Frank Skinner and Harry Kat for numerous constructive comments and suggestions. I also benefited from discussions with Chris Brooks, Antonio Diaz, Mark Fisher, Dimitrios Gkamas, Olan Henry, Piet Sercu, John Sleath, James M. Steeley, Robert I. Webb, Xueping Wu, the participants of the 2000 FMA Conference at Edinburgh, the participants of the 31st annual Money, Macro and Finance conference at University of Oxford and with the seminar participants at the Bank of England.

References

Adams, K., Deventer, D.V., 1994. Fitting yield curves and forward rate curves with maximum smoothness. Journal of Fixed Income (June), 52–62.

[42] Detailed results of the filter rule tests for the Svensson, linear, exponential and integrated B-splines and the VRP are available from the author upon request.

Amihud, Y., Mendelson, H., 1991. Liquidity maturity and yields on U.S. treasury securities. Journal of Finance 46, 1411–1426.

Anderson, N., Breedon, F., Deacon, M., Derry, A., Murphy, G., 1996. Estimating and interpreting the yield curve. Wiley, New York.

Anderson, N., Sleath, J., 1999. New estimates of the UK real and nominal yield curves. Bank of England Quarterly Bulletin, pp. 384–392.

Andrews, D., 1991. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.

Andrews, D., Monahan, J., 1992. An improved heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 60 (4), 953–966.

Ball, C., Torous, W.N., 1983. Bond price dynamics and options. Journal of Financial and Quantitative Analysis 18, 517–531.

Bjerring, J., Lakonishok, J., Vermaelen, T., 1983. Stock prices and financial analysts recommendations. Journal of Finance 38, 187–204.

Black, F., Derman, E., Toy, W., 1990. A one factor model of interest rates and its application to treasury bond options. Financial Analysts Journal 46, 33–39.

Bliss, R., 1997. Testing term structure estimation models. Advances in Futures and Options Research 9, 197–231.

Buono, M., Gregory-Alien, R., Yaari, U., 1992. The efficacy of term structure estimation techniques: A Monte Carlo study. Journal of Fixed Income 1, 52–63.

Chambers, D., Carleton, W., Waldman, D., 1984. A new approach to estimation of the term structure of interest rates. Journal of Financial and Quantitative Analysis 19, 233–253.

Chen, R., Scott, L., 1995. Multi-factor Cox, Ingersoll and Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model. Rutgers University, unpublished paper.

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

Davidson, R., MacKinnon, J., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 49, 781–793.

De Boor, C., 1978. A Practical Guide to Splines. Springer, Berlin.

Dermody, J., Prisman, E., 1988. Term structure multiplicity and clientele in markets with transaction costs and taxes. Journal of Finance 43, 893–911.

Dermody, J., Rockafellar, R., 1991. Cash stream valuation in the face of transaction costs and taxes. Mathematical Finance 1, 31–54.

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

Elton, E., Green, T., 1998. Tax and liquidity effects in pricing government bonds. Journal of Finance 53, 1533–1562.

Elton, E., Gruber, M., 1991. Modern Portfolio Theory and Investment Analysis. Wiley, New York.

Elton, E., Gruber, M., Michaely, R., 1990. The structure of spot rates and immunization. Journal of Finance 45, 629–643.

Fisher, M., Gilles, C., 1998. Around and around: the expectation hypothesis. Journal of Finance 53, 363–383.

Fisher, M., Nychka, D., Zervos, D., 1995. Fitting the term structure of interest rates with smoothing splines. Federal Reserve Board, Finance and Economics Discussion Series, 1.

Gallant, A., 1987. Nonlinear Statistical Models. Wiley, New York.

Green, R., Odegaard, B.A., 1997. Are there tax effects in the relative pricing of U.S. government bonds? Journal of Finance 52, 609–633.

Hardle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.

Hastie, T., Tibshirani, R., 1990. Generalised Additive Models. Monographs on Statistics and Applied Probability, vol. 43. Chapman and Hall, London.

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: A new methodology. Econometrica 60, 77–105.

Ho, T., Lee, S.-B., 1986. Term structure movements and pricing interest rate contingent claims. Journal of Finance 41 (5), 1011–1029.

Hull, J., White, A., 1993. One-factor interest rate models and the valuation of interest rate derivative securities. Journal of Financial and Quantitative Analysis 28, 235–254.
Jordan, J., 1984. Tax-effects in term structure estimation. Journal of Finance 39 (2), 393–406.
Katz, E., Prisman, E., 1991. Arbitrage, clientele effects, and the term structure of interest rates. Journal of Financial and Quantitative Analysis 26 (4), 435–443.
Litterman, R., Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income 1, 54–61.
Litzenberger, R., Rolfo, J., 1984. An international study of tax effects on government bonds. Journal of Finance 39 (1), 1–22.
McCulloch, J., 1971. Measuring the term structure of interest rates. Journal of Business 44 (1), 19–31.
McCulloch, J., 1975. The tax-adjusted yield curve. Journal of Finance 30 (3), 811–830.
McCulloch, J., 1990, U.S. Term Structure Data, 1946–87, Handbook of Monetary Economics, I, Elsevier, Amsterdam. pp. 672–715.
Nelson, C., Siegel, A., 1987. Parsimonious modelling of yield curves. Journal of Business 60 (4), 473–489.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Pearson, N., Sun, T., 1994. Exploring the conditional density in the term structure: An application to the Cox, Ingersoll and Ross model. Journal of Finance 49, 1279–1304.
Powell, M., 1981. Approximation Theory and Methods. Cambridge University Press, Cambridge, MA.
Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. Numerical Recipes in C. Cambridge University Press, Cambridge.
Rady, S., Sandmann, K., 1994. The direct approach to debt option pricing. Review of Futures Markets 13, 461–515.
Ronn, E., 1987. A new linear programming approach to bond portfolio management. Journal of Financial and Quantitative Analysis 22 (4), 439–466.
Ross, S., 1976. Return, risk and arbitrage. In: Friend, I., Bicksler, J. (Ed.), Risk and Return in Finance. Ballinger, Boston, MA.
Ross, S., 1978. A simple approach to valuation of risky streams. Journal of Business 51 (3), 453–475.
Schaefer, S., Schwartz, E.S., 1987. Time-dependent variance and the pricing of bond options. Journal of Finance 42, 1113–1128.
Sercu, P., Wu, X., 1997. The information content in bond model residuals: An empirical study on the Belgian bond market. Journal of Banking and Finance 21, 685–720.
Shea, G., 1984. Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations. Journal of Financial and Quantitative Analysis 19 (3), 253–269.
Steeley, J., 1990. Modelling the dynamics of the term structure of interest rates. The Economic and Social Review 21 (4), 337–361.
Steeley, J., 1991. Estimating the gilt-edged term structure: Basis splines and confidence intervals. Journal of Business Finance and Accounting 18 (4), 513–529.
Steeley, J., 1992. Deregulation and market efficiency: Evidence from the gilt-edged market. Applied Financial Economics 2, 125–143.
Steeley, J., 1997. A two-factor model of the U.K. yield curve. The Manchester School Supplement, pp. 32–58.
Svensson, L., 1994. Estimating and interpreting forward interest rates: Sweden 1992–1994. International Monentary Fund, Working paper no. 114 (September).
Tanggaard, C., 1995. Nonparametric smoothing of yield curves. Department of Finance, Aarhus School of Business, Working paper, D95-1.
Vasicek, O., Fong, H., 1982. Term structure modelling using exponential splines. Journal of Finance 37, 339–356.
Waggoner, D., 1997. Spline methods for extracting interest rate curves from coupon bond prices. Federal Reserve Board of Atlanta, Working paper 97-10 (November).
Wahba, G., 1990. Spline Models for Observational Data. SIAM, Philadelphia.
White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48, 817–828.

Williams, D., 1991. Probability with Martingales. Cambridge University Press, Cambridge.
Wu, X., 1995. Term structures in the Belgian market: Model estimation and pricing error analysis. Ph.D. Thesis, Katholieke Universiteit Leuven.