Journal of Banking & Finance 36 (2012) 121–135

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Models of the yield curve and the curvature of the implied forward rate function

Peter J. Yallup *

Manchester Business School, University of Manchester, Booth Street West, Manchester, M15 6PB, United Kingdom

ARTICLE INFO

Article history:
Received 2 October 2008
Accepted 18 June 2011
Available online 25 June 2011

JEL classification:
E43
G12
C52

Keywords:
Bonds
Yield curve
Forward rate
Discount rate
Curve fitting

ABSTRACT

We examine several alternative models of the UK gilt yield curve using daily data for the period 12 July 1996–10 February 2010. We select the best models according to two criteria: low out of sample errors in pricing bonds and low curvature of the implied forward rate curve function. We suggest additions to some of the models that significantly improve their performance. Some of the new models out perform those typically used by the central banks. In particular this paper suggests that the model used by the Canadian Central Bank which both outperforms other models and is particularly easy to estimate, is well suited to the UK gilt market.

1. Introduction

The association between interest rates and term to maturity, known as the term structure of interest rates (or alternatively the "zero coupon" yield curve) is a fundamental relationship for central banks and other market participants. The yield curve is used to identify rich and cheap securities and to price new issues to the bond market. In addition, the yield curve is used to assess the impact of economic (particularly monetary) policy on the economy through its effect on current interest rates and also through the implied forward rate curve.

In this paper we examine many alternative models of the yield curve and we try to evaluate these across all practical parameter combinations. We select the best models according to two criteria which are most relevant to the expediency of the resulting yield curves. The first criterion is that a superior yield curve model should have low out of sample errors in pricing bonds; the second criterion is that a superior yield curve model should imply a forward rate curve with low curvature. We apply the selection criteria consistently across the different yield curve models.

We use fixed coupon bond data from the UK Government bond (gilt) market which is a large and liquid market. On 27 May 2005 the UK Treasury issued its first "ultra long" fixed coupon gilt, abruptly extending the yield curve from about 35 years to 50 years in maturity. The data used in this paper spans the period 12 July 1996–10 February 2010, and thus allows a comparison of the situation both pre and post this structural change.

We look at a large number of interesting alternative models including some which have not previously been used with UK data. We achieve far better fit to the data (both in sample and out of sample) and we explicitly calculate a measure of the curvature of the forward rate curve. Some of the new models outperform those typically used by the central banks. In particular, this paper proposes that the Li et al. (2001) model used by the Canadian Central Bank which both outperforms other models and is particularly easy to estimate is suitable for estimating the UK gilt yield curve.

1.1. Yield curve models

There are two groups of yield curve models which are present in the literature. The first group of yield curve model (dynamic term structure model) is motivated by the need to price options on long term interest rates (e.g. options on bonds). To do this requires modelling not only the yield curve but also the volatility of those interest rates. In addition, if a model is to be useful in a market where prices are evolving, accuracy in pricing the whole term structure may be less important than the ability to accurately model volatility and to price options quickly. Examples of this group of dynamic term structure models include Cox et al. (1985) and Heath et al. (1992). Determining which of the many alternative dynamic term structure models best fits a given set of data is examined in Aït-Sahalia and Kimmel (2010) and Joslin et al. (2011).

* Tel.: +44 (0) 161 2756333.
E-mail address: peter.yallup@mbs.ac.uk

0378-4266/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbankfin.2011.06.010

The second group of yield curve model is motivated by the need to model the yield curve alone accurately. Here the users of the yield curve models will be central banks and other participants in government bond markets (see Bank for International Settlements, 2005). Such models are used directly to identify rich and cheap bonds and to price new issues. In addition, yield curve models are used for calculating statistics (e.g. implied zero coupon interest rates) that are published for the benefit of all market participants and provide the key inputs to the essential task of determining an appropriate monetary policy,[1] as they show the markets expectation of the future level of interest rates. Thus, much of the published research on these models is conducted by the central banks themselves. It is on this second group of yield curve models that this research is focused.

Yield curve models can be expressed in terms of either the zero coupon yield curve, $r(t)$, the discount function curve, $d(t)$ or the instantaneous forward rate curve, $f(t)$. These are related as follows:

$$d(t) = \exp(-r(t)\cdot t) = \exp\left[-\int_0^t f(s)ds\right] \tag{1}$$

There are four types of model which we will evaluate. These are Nelson and Siegel, polynomial, discount function and cubic spline type models.

2. Estimating yield curve models, practical considerations

2.1. Objective function

To estimate a yield curve model, we need a measure of how well it fits the data. Assume that $\underline{\mathbf{P}}$ is a column vector of gilt prices which we use to estimate a yield curve model. Let $\underline{\mathbf{P}}^*$ be a column vector of price estimates for the gilts based upon their cash flows, discounted by the * yield curve model. Let $\underline{\mathbf{W}}$ be a square "weighting" matrix of appropriate dimension. $\underline{\mathbf{W}}$ has diagonal element w_{ii} > = 0 and zero off diagonal elements. The objective is to:

$$\text{Minimise}: (\underline{\mathbf{P}} - \underline{\mathbf{P}}^*)^T \underline{\mathbf{W}} (\underline{\mathbf{P}} - \underline{\mathbf{P}}^*) \tag{2}$$

Typically the weights w_{ii} are the reciprocal of either the duration or the modified duration of each individual bond squared (e.g. Anderson and Sleath (2001) use the reciprocal of the square of the modified duration).

The objective of this study is to estimate the yield curve. Errors in long term rates of interest will have a greater impact on the pricing of long bonds than the same changes in short term rates will have on short term bonds. In order to select the appropriate weighting scheme we need to understand how changes in yields (or the yield curve) affect the prices of coupon bonds. The modified duration of a bond gives the sensitivity of the price of a bond to changes in yields, or indeed the sensitivity of the yield of a bond to changes in price.

Consider a 5 and 10 year zero coupon bond yielding 5% annually. The modified duration gives the percentage change in the value if yields rise by one basis point (0.01%), (provided that any change in yield is small). Assume that the annual yield rises by 0.1 basis points; we can calculate the new price for each bond and confirm that the modified duration does indeed give the correct value (see Table 1). If we calculate continuously compounded yields the relationship between the modified duration and the change in prices is also valid. In this case the 'new price' of the zero coupon bonds is slightly lower (as a 0.1 basis point move in continuously compounded yields is a slightly larger move than a 0.1 basis point move in annual yields). Note that the duration and the modified duration are the same if we are using continuously compounded yields.

Hence if we use $[100/(\text{Price} \times \text{Modified Duration})]^2$ as the weight in Eq. (2) this will give the desired result. If all bond prices are near 100 this would be equivalent to using $[1/(\text{Modified Duration})]^2$ as the weight. Using the modified weighting scheme allows us to properly handle both high and low (or even zero) coupon bonds.

2.2. Over-fitting yield curve models

One possible problem with estimating yield curve models is the possibility of over-fitting the data. To illustrate, suppose that the yield curve is linear, but observations of yields are measured with some random error. If we have ten observations of yields we could fit a ninth power polynomial to the data exactly. The close fit would suggest that this elaborate model is the best model. Indeed adding terms to any model cannot reduce the fit to the data, as we add terms the fit will typically improve. We need to consider other ways of determining the 'best' model besides those that are purely related to the statistical fit of a model "in sample".

2.3. Effective range of yield curve models

When a yield curve model is estimated, the range of the data used will determine the effective range of the estimated model. For example, if the model is estimated using 1–10 year maturities, it will probably give unreliable estimates of 15 year yields. Indeed some models are not well behaved outside the range of the estimation data used. This consideration will be relevant when the various models are tested.

3. Models of the yield curve

3.1. Nelson and Siegel models of the yield curve

The Nelson and Siegel group of yield curve models comprise the Nelson and Siegel (1987) model and the related models of Svensson (1994) and of Bliss (1986). Nelson and Siegel (1987) introduced a model of the yield curve that is widely used.[2] The model was later extended by Svensson (1994). The Svensson model is specified in terms of the instantaneous forward rate function:

$$f(t) = \beta_0 + \beta_1 \exp\left(\frac{-t}{m_1}\right) + \beta_2\left(\frac{t}{m_1}\right)\exp\left(\frac{-t}{m_1}\right) + \beta_3\left(\frac{t}{m_2}\right)\exp\left(\frac{-t}{m_2}\right) \tag{3}$$

The Nelson and Siegel (1987) model is the same as the above model with $\beta_3 = 0$. Svensson (1994) introduced two extra parameters to give a greater variety of shapes to the instantaneous forward rate and yield curve curves. We can transform the model to give a closed form expression for the yield curve:

$$r(t) = \beta_0 + \beta_1\left(\frac{1-\exp\left(\frac{-t}{m_1}\right)}{\frac{t}{m_1}}\right) + \sum_{i=1}^{2}\beta_{i+1}\left(\frac{1-\exp\left(\frac{-t}{m_i}\right)}{\frac{t}{m_i}} - \exp\left(\frac{-t}{m_i}\right)\right) \tag{4}$$

As $t \to 0$, $r(t) \to \beta_0 + \beta_1$ (also $f(t) \to \beta_0 + \beta_1$) and as $t \to \infty$, $r(t) \to \beta_0$ (also $f(t) \to \beta_0$). So in this model β_0 can be interpreted as a long

[1] A striking recent example is the Bank of England's implementation of a policy of "quantitative easing ". Between 11 March 2009 and 26 January 2010 almost £ 200 billion was used by the Bank of England to purchase gilts from the market.

[2] Diebold and Li (2006) and Christensen et al. (2011) have derived dynamic versions of the Nelson and Siegel (1987) model.

Table 1
The relationship between modified duration and price changes for 5 year and 10 year zero coupon bonds.

Maturity (years)	Annual yields		Continuously compounded yields	
	5	10	5	10
Yield 1 (%)	5.0000	5.0000	4.8790	4.8790
Price 1	78.35262	61.39133	78.35262	61.39133
Yield 2 (%)	5.0010	5.0010	4.8800	4.8800
Price 2	78.34889	61.38548	78.34870	61.38519
Duration	5	10	5	10
Modified duration	4.76190	9.52381	5	10
Change in value (%)	−0.47618	−0.95233	−0.49999	−0.99995

period rate of interest and $\beta_0 + \beta_1$ as an instantaneous short rate. Another desirable property of this model is that both the term structure and the instantaneous forward rate curve are smooth functions.

Bliss (1986) introduced a related model that is the same as the Svensson model but with $\beta_2 = 0$.

3.2. Polynomial models of the yield curve

The polynomial model was examined in Chambers et al. (1984).

$$r(t) = A_1 + A_2 t + A_3 t^2 + \cdots + A_j t^{j-1} \tag{5}$$

Here it is the yield curve which is being modelled as a polynomial function. More recently, in the context of immunisation, de La Grandville (2001) modelled the yield curve as a McLaurin series for which a polynomial function may be a good approximation. If $r(t)$ t can be written as a McLaurin series, then $r(t)$ can be approximated by the polynomial:

$$r(t) \approx \frac{A_0}{t} + A_1 + A_2 t + A_3 t^2 + \cdots + A_j t^{j-1} \tag{6}$$

This implies the following equation for the discount function:

$$d(t) = \exp(-A_0) \cdot \exp(-A_1 t) \cdot \exp(-A_2 t^2) \cdot \exp(-A_3 t^3) \ldots \tag{7}$$

Here the only difference is that the first term in Eq. (6) is t^{-1}, whereas in Eq. (5) the first term is t. The t^{-1} term has a low correlation with the other terms and the addition of this term improves the estimation of the model. Siegel and Nelson (1988) noted the suitability of t^{-1} as an explanatory variable of the yield curve if the yield curve tends asymptotically to a fixed value. It is also possible to model the discount function as a polynomial function, as in Bolder and Gusba (2002).

$$d(t) = \frac{A_0}{t} + A_1 + A_2 t + A_3 t^2 + \cdots + A_m t^{m-1} = \sum_{j=0}^{m} A_t t^{j-1} \tag{8}$$

Bolder and Gusba (2002) do not find this an easy function to work with as the underlying variables (t, t^2, t^3 etc.) are highly correlated and this can lead to numerical problems estimating parameters, due to colinearity. As with the yield curve polynomial, in this paper, we add a t^{-1} term in (8) to improve the estimation of the model.

3.3. Models of the discount function

Li et al. (2001) and Bolder and Gusba (2002) consider models in which it is the discount function (rather than the yield curve or the instantaneous forward rate) that is modelled. A consequence of modelling with the discount function is that these models are relatively easy to estimate.

3.3.1. Exponential models of the discount function

Li et al. (2001) consider a model in which the discount function is modelled as a sum of exponential functions (hereafter the exponential model). This model is included for two reasons. Firstly, Bolder and Gusba (2002) find that this model performs well against other models on Canadian government bond data. The Bank of Canada uses a 9 parameter exponential model to estimate yield curves (see Bank for International Settlements, 2005). Secondly, the model is an interesting alternative to the polynomial yield curve model in Section 3.2 above. In the polynomial model the discount function is expressed as a product of exponential functions (see Eq. (7)) whilst the exponential model is a sum of exponential functions.

$$d(t) = C_1 \exp(-\alpha t) + C_2 \exp(-2\alpha t) + C_3 \exp(-3\alpha t) + \cdots \tag{9}$$

Here α is a long term rate of interest.[3] This model is particularly easy to evaluate as it is "twice linear". Firstly, if we have the discount function, the value of a set of cash flows is a linear function (with cash flows as weights) of discount function values:

$$P = \sum_{t=1}^{T} c(t) d(t) \tag{10}$$

Secondly, the discount function is itself a linear function of exponentials. This means that the coefficients in Eq. (10) can easily be estimated using ordinary least squares techniques (if we use a weighting scheme, then estimation using generalised least squares techniques is appropriate). The exponential model is different to other models of the yield curve of interest rates as it does not include a constant term. A constant term will fit into the sequence of explanatory variables as:

$$C_0 \exp(-0\alpha t) = C_0 \tag{11}$$

Hence an additional model (which we can call the "extended" exponential) is:

$$d(t) = C_0 + C_1 \exp(-\alpha t) + C_2 \exp(-2\alpha t) + \cdots + C_m \exp(-m\alpha t) \tag{12}$$

Note that if we substitute $t = 0$, and $d(0) = 1$[4] into Eq. (12) we get $d(0) = C_0 + C_1 + C_2 + \cdots + C_m = 1$. So, for any particular model (e.g. a 5 parameter extended exponential model with unknown parameters C_0, C_1, C_2, C_3, and C_4) one of the C parameters can always be calculated from the remaining C parameters (i.e. the 5 parameter model has only 4 independent parameters).

3.3.2. Bolder and Gusba, Fourier model of the discount function

Bolder and Gusba (2002) were motivated by the simplicity of the exponential model of Li et al. (2001), to look at other linear models of the discount function. The first was the polynomial model of the discount function mentioned in Section 3.2. Another was a Fourier series model of the discount function (hereafter known as the Fourier model). Here they consider a function made up of the

[3] α is interpreted a long term rate of interest rather than an unknown parameter. In a practical context can be set equal to the yield on the longest bond in the yield curve.

[4] As $d(t) = \exp(-r(t)t)$, if $t = 0$ then $d(0) = 1$.

terms in the Fourier series $\{\sin(nt/\omega), \cos(nt/\omega); n = 0, 1, 2, 3, 4\}$. As $\sin(0) = 0.0$ and $\cos(0) = 1.0$, we have the following model:

$$d(t) = C_1 + C_2 \sin(t/\varpi) + C_3 \cos(t/\varpi) + C_4 \sin(2t/\varpi) + C_5 \cos(2t/\varpi) + \cdots \quad (13)$$

In their estimation, Bolder and Gusba (2002), set $\omega = 10$.

4. Cubic spline models

4.1. McCulloch models of the discount function

Spline models were first used by McCulloch (1971, 1975). Here the discount function is modelled as a series of connected curves. These curves are joined smoothly to each other at points called knots. In McCulloch (1971) the discount function was modelled as a series of connected quadratic curves. At the knot points adjacent quadratic curves would have the same value and the same first derivative. An unfortunate consequence of this scheme was that the instantaneous forward rate curve was not smooth. Indeed the forward rate curve had a series of kinks corresponding to the knots which McCulloch termed "knuckles".

In McCulloch (1975) the discount function was modelled as a series of connected cubic curves, known as a "cubic spline". At the knot points adjacent cubic curves would have the same value, the same first derivative and also the same second derivative. Hence cubic splines would be smoother curves (using quadratics we can only restrict the values and first derivatives without making adjacent curve segments formulae identical). In these cubic spline models the number of knots is restricted, as too many knots leads to numerical problems. In addition the knots must be positioned by the user. Typically they are placed so that there are approximately equal numbers of bonds between each knot point.[5]

4.2. Estimation of cubic splines: B-spline models

In a B-spline model the curve is formed from a linear combination of simple splines called B-splines. Each B-spline is made up of four cubic curve segments which are joined at knot points. Each B-spline obeys the restriction that first and second derivatives of adjacent segments agree at the knot points. Each B-spline will span five adjacent knot points, one at either end and three which define the joins in the B-splines' four component cubic curve segments. Individual B-splines overlap each other, so that at any point which is not a knot, there will be four B-spline curves. A cubic spline is then fitted to a given set of data by taking a linear combination of the B-splines which best fits according to some criteria. As the component B-splines obey the restrictions that first and second derivatives of adjacent segments agree at the knot points, the resultant cubic spline will also fulfil this restriction.

Cubic spline models estimated conventionally involve numerical operations which lead to inaccurate estimates of coefficients. In the worst case these numerical problems can lead to impossible (negative) estimated values of some coefficients. B-splines are a superior way of estimating the existing cubic spline models. The use of B-splines to estimate spline models was first advocated by Shea (1984) and Steeley (1991).

4.3. Smoothing spline models

McCulloch (1971) noted a disadvantage of the quadratic spline model was that the equivalent forward rate curve was not smooth. It had distinct bumps (which he termed "knuckles") corresponding to the knot points. The cubic spline model suffered less from this problem, but as Shea (1984) noted, the problem was still present. Fisher et al. (1995) suggested an improvement on the cubic spline model could be achieved by adding a penalty function to the objective.

$$\sum_{i=0}^{N} w_i (P_i - \widehat{P}_i)^2 + \lambda_t \int_0^T [d(\tau)]^2 d\tau \quad (14)$$

The penalty is expressed in terms of the curvature (second derivative) of the discount function and has the effect of reducing curvature. The parameter λ_t determines the degree of smoothness and is determined daily by selecting a value that minimises the generalised cross validation (GCV) measure. This is the sum of square price errors adjusted by the number of bonds, N, the effective number of parameters θ (which is equal to the number of effective knot points plus two) and a 'cost' parameter γ as in Eq. (15).

$$GCV = \frac{\sum_{i=1}^{n} (P_i - \widehat{P})^2}{(N - \gamma\theta)^2} \quad (15)$$

Fisher et al. (1995) set $\gamma = 2.0$. However, this choice is an arbitrary one. Jarrow et al. (2004) point out that the usual justification for using the generalised cross validation measure is that (with $\gamma = 1.0$) the measure approximates cross validation. In cross validation (such as the "leave out one cross validation" used in Anderson and Sleath (2001), described in Section 8.3 below) part of the data is used to fit the model and the remainder is used to validate the model (by calculating out of sample error statistics). So $\gamma = 1.0$ should approximately find the smoothing parameter that minimises the out of sample errors. Arbitrarily setting $\gamma = 2.0$ reduces the curvature further than is required under the cross validation measure and this results in a smoother discount function, but with higher out of sample errors. Furthermore, using this restriction may bias the comparison of smoothing spline techniques to the non-spline alternatives. Using smoothing splines the number of knot points can be increased above the maximum number for unsmoothed splines. Increasing the smoothing has the effect of reducing the impact of some of the knot points. In effect, the data and the smoothing function determine which knot points are used.[6]

Perhaps the most significant advance in Fisher et al. (1995), was to show how it was possible to fit splines to the yield curve and the forward rate curve as well as the discount curve. In the first case we determine the coefficients, $\beta_{k,}$ which give the best fit of B-splines $\varphi_k(t)$ to the discount curve.[7] The discount curve B-spline is:

$$d(t) = \sum_{k=1}^{n} \beta_k \varphi_k(t) \quad (16)$$

In the second case we determine the coefficients, $\beta_{k,}$ which give the best fit of B-splines $\varphi_k(t)$ to $r(t)t$. Recall that $d(t) = \exp[-r(t)t]$. So fitting Eq. (17) is a yield curve B-spline:

$$d(t) = \exp\left(-\sum_{k=1}^{n} \beta_k \varphi_k(t)\right) \quad (17)$$

In the third case we determine the coefficients, $\beta_{k,}$ which give the best fit of B-splines $\varphi_k(t)$ to $f(t)$. If we recall that $d(t) = \exp\left[-\int_0^t f(s) ds\right]$, fitting Eq. (18) is a forward rate B-spline:

$$d(t) = \exp\left[-\int_0^t \sum_{k=1}^{n} \beta_k \varphi_k(s) ds\right] \quad (18)$$

[5] Fernandez-Rodriguez (2006) examines the problem of finding the optimal spline knot locations.

[6] Laurini and Moura (2010) augment smoothing B-spline estimation with added constraints to ensure that the resulting model has reasonable characteristics (for example, non-negative implied discount rates).

[7] This is essentially a B-spline estimate of the McCulloch (1975) model.

4.4. Variable roughness penalty spline models

Waggoner (1997) noted that the smoothing of the discount curve in Fisher et al. (1995) applied the same degree of smoothing to all parts of the discount function. Instead he proposed to use a different degree of smoothing in different parts of the curve. He divided the curve into three segments, money market rates (up to one year), notes (one to ten years) and bonds (ten years and over). For each of these segments he chose a different smoothing parameter, which was fixed for all periods of estimation.

$$\sum_{i=0}^{N} w_i(P_i - \widehat{P}_i)^2 + \int_0^T \lambda(s)[v''(s)]^2\,ds \tag{19}$$

where $\lambda(s) = \{0.1$ for $0 <= s <= 1$, 100 for $1 <= s <= 10$, $100{,}000$ for $10 <= s\}$.

This technique was termed by Waggoner (1997) the "variable roughness penalty" (VRP) method. Anderson and Sleath (1999, 2001) introduced three improvements to the VRP method of Waggoner. Firstly, in Waggoner (1997) weighting of the price errors is not discussed and hence we can only assume that no weighting scheme was used. Anderson and Sleath (1999, 2001) used the reciprocal of the square of the modified duration as weights. Secondly, they use the B-spline method to estimate the equations and avoid numerical instability problems (see Section 4.2 above). Lastly a smooth function was used for the smoothing parameter, rather than the somewhat ad hoc, step function, smoothing of the discount curve in Waggoner (1997). The weighting scheme used in Anderson and Sleath (1999, 2001) takes the form $\log\lambda(m) = L - (L - S)\exp(-m/\mu)$. The specific parameters used are not mentioned, but from Chart 2 in Anderson and Sleath (2001) we can infer that $L = \log(10{,}000)$, $S = 0$ and $\mu = 1.44$, approximately.

5. Yield curve models used by central banks

Much of the research on yield curve modelling is carried out by the central banks which use the models in a practical context. It is therefore interesting to see which models the various central banks actually use. Fortunately data relating to the use of yield curve models was recently presented in Bank for International Settlements (2005). Looking at Table 2, the Nelson and Siegel (1987) and Svensson (1994) models were widely used. Smoothing splines were used by Japan, the United States and Sweden. The United Kingdom was the only user of a VRP model (it has separate models for fixed coupon bonds and inflation linked bonds). Canada was the sole user of the Li et al. (2001), exponential model.

6. Ioannides (2003)

Ioannides (2003) looked at yield curve estimation in the UK gilt market. His data covered the period from 1 January 1995 to 1 January 1999, and consisted of 1046 observations (days). The techniques estimated were Nelson and Siegel (1987) and Svensson (1994), the McCulloch (1975) cubic spline model, three FNZ smoothing spline models (Fisher et al., 1995) and the VRP model (Waggoner, 1997). The FNZ smoothing spline models are a smoothing spline model of the discount rate, a yield curve model ($r(t)\cdot t$) and a smoothing spline model of the forward curve. The models were examined using within sample analysis, out of sample analysis and an ex-post performance measurement. It was found that the Svensson (1994) method performed the best overall and the various spline methods performed poorly in comparison. Trading rule testing also suggested that the non-spline models outperformed their spline alternatives. The error statistics reported in Ioannides (2003) were surprisingly large. The size of the reported errors may indicate some systematic error in the estimation of the different models.

7. UK gilt market

Prior to April 1996, gilt investors were not a homogenous group with respect to the taxation on capital gains and income from their gilt holdings. Different types of investors favoured different types of gilt according to their tax position. Hence models of the gilt market invariably had to consider an elaborate model of the yield curve which allowed for the effects of taxation. Schaefer (1982) considered a model in which the gilt market was segmented into tax clienteles favouring high, medium or low coupon gilts. On 6 April 1996 the taxation of gilts was changed in preparation for the introduction of a gilt strip market (which started on 8 December 1997). Under the new tax rules the total return (capital gain plus income) on gilts is taxed as income. This change affected all UK gilt holders except individuals who continue to pay no tax on capital gains for coupon paying gilts. Individuals make up a small percentage of the holders of gilts and hence their tax position is unlikely to affect most liquid gilts (although it may impact the pricing of some old rump stocks which were originally issued with very low coupons). On 27 May 2005 the DMO issued a 50 year "ultra long" fixed coupon gilt. This was in response to demand from institutions for long maturities assets to match (or immunize) longer liabilities (for example pensions investment due to the increased longevity of pension beneficiaries). Subsequently a further four ultra long issues have been made to fill the gap between the old issues (at that time up to 34 year maturity) and the new ultra long 50 year maturity.

7.1. Gilt pricing

In the gilt market there are several reasons why some gilts may be priced slightly cheaper or slightly more expensive than other gilts at any particular time. From time to time the Debt

Table 2
Yield curve models used by central banks (adapted from Bank for International Settlements, 2005, Table 1).

Central bank	Estimation method	Minimised error	Relevant maturity spectrum
Belgium	Svensson or Nelson Siegel	Weighted prices	2 days–16 years
Canada	Exponential	Weighted prices	3 months–30 years
Finland	Nelson Siegel	Weighted prices	1–12 years
France	Svensson or Nelson Siegel	Weighted prices	Up to 10 years
Germany	Svensson	Yields	1–10 years
Italy	Nelson Siegel	Weighted prices	Up to 30 years
Japan	Smoothing splines	Prices	1–10 years
Norway	Svensson	Yields	Up to 10 years
Spain	Svensson	Weighted prices	Up to 10 years
Sweden	Smoothing splines and Svensson	Yields	Up to 10 years
Switzerland	Svensson	Yields	1–30 years
UK	VRP	Yields	Up to 30 years
USA	Smoothing splines 2 Curves	Bills: Weighted prices Bonds: prices	Up to 1 year 1–10 years

Management Office will issue further amounts of an existing issue. In recent times the main method of issue has been by a pre-announced auction. After an auction is announced, a gilt may trade slightly cheaper relative to other gilts than at other times. Some gilts are issued in very large amounts and are popular liquid issues; other gilts exist in smaller amounts. Other things being equal, the larger more liquid issues usually trade more expensively than the smaller issues. In this study we have deliberately excluded the smallest 'rump' issues because the gilt-edged market makers are not obliged to make prices for these issues. However there will still be a difference between say a £ 1 billion issue and a £ 15 billion issue. Gilts that are deliverable into the futures contract, in particular the cheapest to deliver, may trade more expensively than other gilts when the futures contract is near to expiry.

From 8 December 1997 to 21 October 2009, all new fixed coupon gilt issues became strippable once the issued size had passed £ 4.0 billion. In effect, these issues have a free option (to convert into strips) not included in other issues. Between 6 April 1996 and 5 April 1999, the strippable gilts also enjoyed a small advantage over the other gilts (as they paid interest gross). These differences mean that strippable gilts may trade expensively relative to other gilts. Since 21 October 2009, some new gilt issues are denoted, by the DMO at issue, as not strippable even when they exceed £ 4.0 billion in size.

If gilt prices are collected from a single market maker, these prices could be at the bid or offer side of the market depending on the inventory and expectations of the particular dealer. In these circumstances, the bid offer spread could add further noise to prices. In this study we have deliberately taken the DMO closing prices, which are averaged over all dealers (after excluding outliers) and should therefore reduce this potential problem.

8. Estimation

In section 3 we discussed several models which are included in our empirical examination. These models are summarised in Table 3.

BFGS is the well known and widely used Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970) search algorithm to find the solution to multivariate, non-linear optimisation problems. Yield curve estimation using the BFGS methods were programmed in C++, using BFGS and line search routines from Press et al. (2001). The generalised least squares technique is a generalisation of ordinary least squares to the case where errors are weighted. The generalised least squares method was programmed in Matlab using code provided in Bolder and Gusba (2002). The Taylor Series Iteration technique used to calculate forward rate and yield curve splines is described in the Appendix. Several of the yield curve models contain an optional number of parameters. Where this is possible we estimate each possible model to evaluate which is the best number of parameters to use.

8.1. Data

Gilt data were obtained from the UK Debt Management Office (DMO) website.[8] These data comprised daily end of day 'reference prices'[9] and nominal amount issued for fixed coupon gilts for the period from 12 July 1996 to 10 February 2010. This resulted in 3433 individual days of gilt prices. From these data, gilts which had a single defined maturity date,[10] had at least 1 year until maturity and which were not at that time designated as 'rump stocks'[11] were used. See summary statistics for the gilt sample data in Table 4. During the period covered by the research the number of fixed coupon, liquid, non-call gilt bonds initially declined (from 36 in 1996 to a minimum of 21 in 2002). This is because the Debt Management Office (and the Bank of England before them) has aimed to create fewer, larger, more liquid issues. Later the number of issues increased, particularly after 2005 when the ultra long issues (with maturities over 35 years) were introduced to extending the maximum maturity to 50 years. By February 2010 there were 34 gilts in the sample. To reflect the abrupt, significant change in the length of the gilt curve, these data were divided into two sub periods; period 1, from 12 July 1996 to 26 May 2005 and period 2, from 27 May 2005 to 10 February 2010.

Table 3
Summary of yield curve models.

Model	Estimated on	Estimation technique
Nelson and Siegel	Yield curve	BFGS
Svensson	Yield curve	BFGS
YC polynomial	Yield curve	BFGS
DR polynomial	Discount rates	Generalised least squares
Exponential model	Discount rates	Generalised least squares
Extended exponential model	Discount rates	Generalised least squares
Fourier model	Discount rates	Generalised least squares
DR spline	Discount rates	Generalised least squares
DR smoothing spline	Discount rates	Generalised least squares
YC spline	Yield curve	Taylor series iteration
YC smoothing spline	Yield curve	Taylor series iteration
FR spline	Forward rates	Taylor series iteration
FR smoothing spline	Forward rates	Taylor series iteration
Anderson Sleath FR spline	Forward rates	Taylor series iteration

8.2. Within sample estimation

For each curve three error statistics are calculated from the actual and theoretical bond prices. These are: a root mean square price error (RMSE) defined in Eq. (20), a mean absolute price error (MAE) defined in Eq. (21) and a root mean square weighted price error (RMS(WE)) defined in Eq. (22). The weighting scheme used is described in Section 2.1 and has the effect of transforming a price error into a yield error. Hence RMS(WE) is effectively a root mean square yield error.

$$RMSE = \sum_{i=1}^{n} \sqrt{\frac{(\widehat{P}_i - P_i)^2}{n}} \tag{20}$$

$$MAE = \sum_{i=1}^{n} \frac{|\widehat{P}_i - P_i|}{n} \tag{21}$$

$$RMS(WE) = \sum_{i=1}^{n} \sqrt{\frac{w_i(\widehat{P}_i - P_i)^2}{n}} \quad \text{where} \quad wi = \left[\frac{100}{P_i}\frac{1}{D}\right]^2 \tag{22}$$

Minimising RMS(WE) was the objective in the estimation of the various yield curves. The within sample error statistics for the various estimated curves are shown in Table 5.

The prices of gilts are expressed as a percentage of par value, i.e. a price of 99.00 is 99% of par (the percentage sign is usually dropped). Therefore the price errors are also expressed in the same units. An error of 0.25 is 0.25% of par. In the case of the weighted

[8] http://www.dmo.gov.uk.

[9] Gilt reference prices are the arithmetic mean of closing mid prices obtained from members of the Gilt-edged Market Makers Association after excluding outliers. For fixed coupon gilts the outliers are any prices more than 0.15% (of par) away from the median price.

[10] Thus excluding any callable or perpetual issues.

[11] A rump stock is an issue for which the gilt market makers are not obliged to quote prices due to the stock being illiquid. Rump stocks are designated by the DMO (Bank of England prior to 1st April 1998).

Table 4
Summary statistics for gilt data.

		Period 1	Period 2	All
Sample	From	12-July-1996	27-May-2005	12-July-1996
Period	To	26-May-2005	10-February-2010	10-February-2010
Days		2241	1192	3433
Observations		58,778	32,645	91,416
Coupon%	Average	7.77	5.47	6.95
	Minimum	4.00	2.25	2.25
	Maximum	15.50	9.75	15.50
Daily number of gilts	Average	26.22	27.39	26.63
	Minimum	21.00	24.00	21.00
	Maximum	36.00	35.00	36.00
Maturity of gilts	Average	9.28	14.08	10.98
	Minimum	1.00	1.00	1.00
	Maximum	34.65	50.55	50.55
Daily value of gilt sample £ billion	Average	226.53	397.13	285.77
	Minimum	289.36	717.16	717.16
	Maximum	190.44	286.57	190.44

errors RMS(WE), we know that the error translates into a yield error. An RMS(WE) of 0.05 represents a yield error of 0.05% (or 5 basis points).

For the non-spline methods we can see that as we increase the number of estimated parameters, the measurement errors all reduce. We know that this will be the case in general as each extra parameter cannot increase the error. In the smoothing spline models it is clear that as the smoothing parameter is increased the errors also increase. Since we know that the effect of the smoothing parameter is to reduce the number of knots which are used in the yield curve model, increasing smoothing is equivalent to reducing the number of parameters in the model.

8.3. Out of sample testing

It is probable that as the number of parameters in the yield curve model is increased, or the smoothing parameter in the spline models is reduced, we will "over fit" the data. Hence, we need to see how the various techniques will perform out of sample. Ideally we could test out of sample by dividing the data, randomly, into two groups. One group, of gilt data, would be used to estimate the various yield curve models and the remaining bonds would be used to test the accuracy of the model (by comparing the prices of those gilts implied by the model to the market prices of those gilts). However, unfortunately we do not have enough gilts available to us to use the data in this way.

In order to test our methods out of sample we adopt the technique used in Anderson and Sleath (1999, 2001) described as "leave one out cross-validation" (Davison and Hinkley, 1997). Here the yield curve is estimated using all but one gilt. The pricing of this gilt implied by the estimated yield curve is then compared to its market price to measure the pricing error. For each gilt used in turn we get a series of yield curves and a series of error estimates. In this research the procedure is slightly modified as we omit the longest or the shortest gilts from this exclusion procedure. If there were 35 gilts, we would estimate the yield curve 33 times, each time excluding one gilt (all except for the two "end" gilts) and we would generate 33 error statistics. If we did not exclude the longest and shortest gilts from this testing procedure we would be using the estimates outside their effective range (e.g. using a curve estimated using 1 to 25 year maturities to price a 30 year bond, see Section 2.3). The out of sample testing was carried out on 528 randomly selected days from the entire sample period. We report the same error statistics as for the within sample testing above [RMSE, MAE and RMS(WE)] in Table 5.

In Table 5 we also compare the rank order of the in sample and the out of sample goodness of fit statistics (RMS(WE)) for the key term best structure model of each type. It is apparent that the ordering of the various models is changed substantially using the out of sample statistics. This suggests that some models may be over-fitting the data. For example in period 1, the Fourier model performs well within sample (ranking 8th overall) but out of sample performs very inadequately (12th).

9. Curvature of the forward rate

In addition to the fit of the various yield curve models to the data, the shape of the implied forward rate curve is also important. The apparent problem with the cubic spline model of McCulloch (1975) was that the implied forward rate curve had distinct distortions in it corresponding to the knot points. The various smoothing spline models were introduced to reduce this problem. However, the widely used method of choosing the appropriate degree of smoothing (the generalised cross validation measure, Fisher et al., 1995) is ad hoc and not well suited to its intended purpose. In addition, it needs to be acknowledged that yield curve models other than the spline models can also suffer from badly behaved forward rate curves.

9.1. Measuring the curvature of the forward rate

Once we have estimated a particular yield curve model it is possible to calculate the curvature of the implied instantaneous forward rate curve explicitly. To calculate the curvature of the forward rate, we first need to calculate the forward rate curve. The yield curve is calculated at discrete points, 0.01 years apart ($t = \{0.00, 0.01, 0.02, 0.03\ldots\}$) and the forward rate is calculated as:

$$f(t) \approx \frac{r(t+.01)(t+.01) - r(t-.01)(t-.01)}{(t+.01)-(t-.01)} \tag{23}$$

The slope of the forward rate is then calculated as:

$$\frac{df(t)}{dt} \approx \frac{f(t+.01) - f(t-.01)}{(t+.01)-(t-.01)} \tag{24}$$

And the curvature is then:

$$\frac{d^2f(t)}{dt^2} \approx \frac{\frac{df(t+.01)}{dt} - \frac{df(t-.01)}{dt}}{(t+.01)-(t-.01)} \tag{25}$$

Table 5

These tables show the goodness of fit statistics for the various yield curve models, both in sample and out of sample. These were: a root mean square price error (RMSE) defined in Eq. (24), a mean absolute price error (MAE) defined in Eq. (25) and a root mean square weighted price error (RMS(WE)) defined in Eq. (26). The units of these measures are the same as the bond prices; that is the percentage of par value. An RMS(WE) of 0.05 is equivalent to a yield error of 5 basis points or 0.05%. The curvature statistic is a measure of the curvature of the Forward Rate defined in Eq. (30). Lambda(t) is the smoothing parameter used in smoothing spline models of the yield curve or discount rate or forward rate. Error statistics in bold denote the model with lowest error of all models of that type estimated. Rank is the rank order of the RMS(WE) error measure. Ticks under Fig. 11 or Fig. 12 denote models appearing in that graph.

Panel 1: Goodness of fit and forward rate curvature statistics for yield curve models - period 1 (12 July 1996 – 26 May 2005)

Fig. 11	Model	Lambda(t)	Parameters	Within Sample Average (2241 days)					Out of Sample Average (341 days)			
				RMSE	MAE	RMS(WE)	Rank	Curve	RMSE	MAE	RMS(WE)	Rank
	DR polynomial		5	0.246659	0.234726	0.042560		4.55	0.407487	0.278472	0.054053	
√	DR polynomial		6	0.209463	0.170508	0.033542		4.69	0.335250	0.230376	0.044931	
√	DR polynomial		7	0.194038	0.149782	0.030757	6	6.03	0.339266	0.230869	0.044262	11
	YC polynomial		3	0.547990	0.370070	0.060578		0.00	0.487923	0.356423	0.059397	
√	YC polynomial		4	0.299353	0.216545	0.039711		0.91	0.334313	0.244240	0.044599	
√	YC polynomial		5	0.225706	0.166573	0.033701		1.77	0.289080	0.210791	0.040331	
√	YC polynomial		6	0.215215	0.159386	0.032352	9	2.31	0.289605	0.212722	0.039489	9
	Fourier		6	0.301945	0.315937	0.048799		14.32	0.735469	0.427898	0.065666	
	Fourier		7	0.202503	0.159792	0.031642	8	9.71	0.550640	0.301462	0.050165	12
√	Exponential		4	0.226222	0.200214	0.036038		0.99	0.299237	0.220139	0.041220	
√	Exponential		5	0.206034	0.171034	0.032731		1.45	0.274804	0.206197	0.039744	
√	Exponential		6	0.196612	0.156313	0.029973		2.71	0.272048	0.203525	0.038252	
√	Exponential		7	0.187630	0.144001	0.028564		4.24	0.270900	0.203076	0.038103	6
√	Exponential		8	0.185974	0.139533	0.027554	3	6.31	0.332953	0.223824	0.040222	
	Extended exponential		3	0.313837	0.369732	0.061269		0.20	0.479107	0.350804	0.059308	
√	Extended exponential		4	0.230896	0.208757	0.037988		0.76	0.319104	0.234114	0.043162	
√	Extended exponential		5	0.202121	0.164202	0.032745		1.35	0.269806	0.203576	0.039457	
√	Extended exponential		6	0.199802	0.161010	0.030917		2.32	0.289425	0.213811	0.039234	
√	Extended exponential		7	0.189665	0.145145	0.028870		4.20	0.291598	0.211848	0.038587	7
√	Extended exponential		8	0.185142	0.138642	0.027956	4	6.45	0.417855	0.249865	0.043857	
√	Nelson Siegel		4	0.302877	0.210378	0.036414	11	2.19	0.314516	0.228739	0.040896	10
√	Svensson		6	0.223184	0.164466	0.031587	7	2.53	0.262959	0.196780	0.037988	5
√	DR spline	0	7.0	0.194930	0.151468	0.029115	5	4.11	0.288776	0.213007	0.039136	8
	DR smoothing spline	10	15.0	0.168588	0.112564	0.021149	1	24.44	0.237018	0.179837	0.036465	3
	DR smoothing spline	100	12.5	0.223531	0.178874	0.030609		18.24	0.329769	0.230586	0.043014	
	DR smoothing spline	500	10.1	0.280808	0.275491	0.046006		18.48	0.489896	0.338864	0.058333	
	YC smoothing spline	50	14.7	0.202985	0.151869	0.027017	2	21.63	0.211408	0.159431	0.035656	2
	YC smoothing spline	500	12.2	0.281020	0.275227	0.044541		17.62	0.392193	0.279850	0.050079	
	FR spline	0	7.0	0.208857	0.169926	0.032447	10	4.05	0.361522	0.252033	0.068126	13
	FR smoothing spline	10	14.2	0.264445	0.252131	0.036725	12	13.60	0.207402	0.156532	0.034372	1
	FR smoothing spline	25	13.5	0.277139	0.276705	0.040137		10.38	0.209951	0.158345	0.034588	
	FR smoothing spline	50	12.9	0.285024	0.292623	0.042368		8.23	0.213242	0.159927	0.034813	
√	FR smoothing spline	100	12.3	0.293960	0.310950	0.044827		6.44	0.218148	0.162775	0.034828	
√	FR smoothing spline	200	11.7	0.300665	0.324242	0.046661		5.02	0.224579	0.167203	0.034741	
√	FR smoothing spline	300	11.3	0.303996	0.332848	0.047839		4.37	0.227564	0.169147	0.034635	
√	FR smoothing spline	500	10.9	0.310040	0.346286	0.049620		3.74	0.232397	0.172067	0.034632	
√	FR smoothing spline	1000	10.3	0.317413	0.364072	0.051953		3.13	0.239409	0.176538	0.034826	
√	FR smoothing spline	2500	9.5	0.326439	0.385638	0.054955		2.62	0.248852	0.183129	0.035470	
√	FR smoothing spline	5000	8.9	0.332243	0.400071	0.057128		2.34	0.256786	0.189441	0.036492	
√	FR smoothing spline	10,000	8.3	0.337745	0.415261	0.059759		2.10	0.266401	0.197605	0.038096	
√	FR smoothing spline	25,000	7.6	0.346297	0.438040	0.064083		1.80	0.279712	0.209355	0.040756	
√	FR smoothing spline	50,000	7.1	0.350269	0.449483	0.067228		1.62	0.290234	0.217883	0.043216	
√	FR smoothing spline	100,000	6.7	0.356057	0.467127	0.071807		1.44	0.303313	0.227996	0.046754	
√	FR Anderson Sleath		9.5	0.293782	0.311373	0.046690	13	2.64	0.242631	0.183304	0.037830	4

Panel 2: Goodness of fit and forward rate curvature statistics for yield curve models - period 2 (27 May 2005 – 10 February 2010)

Fig. 12	Model	Lambda(t)	Parameters	Within sample average (1192 days)					Out of sample average (187 days)			
				RMSE	MAE	RMS(WE)	Rank	Curve	RMSE	MAE	RMS(WE)	Rank
	DR polynomial		5	0.351498	0.542283	0.073635		7.22	0.849434	0.613529	0.093099	
	DR polynomial		6	0.278839	0.375057	0.053119		7.69	0.620550	0.447207	0.068718	
√	DR polynomial		7	0.237450	0.263544	0.042246	3	5.56	0.502226	0.352475	0.060510	9
	YC polynomial		3	1.155953	0.818878	0.102431		0.00	1.024519	0.770264	0.094806	
	YC polynomial		4	0.648372	0.476499	0.064109		0.65	0.629324	0.472791	0.064016	
√	YC polynomial		5	0.417064	0.308791	0.046664		1.32	0.463916	0.344680	0.051672	
√	YC polynomial		6	0.373975	0.276384	0.043349	5	1.61	0.434399	0.317555	0.048336	6
	Fourier		6	0.563547	1.148727	0.142087		>100	2.514273	1.532214	0.184189	13
	Fourier		7	0.550355	1.064510	0.125271	12	>100	4.167879	1.926669	0.245671	
√	Exponential		4	0.272244	0.316212	0.047248		0.97	0.428369	0.331695	0.052374	

Table 5 (*continued*)

Panel 2: Goodness of fit and forward rate curvature statistics for yield curve models - period 2 (27 May 2005 – 10 February 2010)

Fig. 12	Model	Lambda(t)	Parameters	Within sample average (1192 days)					Out of sample average (187 days)			
				RMSE	MAE	RMS(WE)	Rank	Curve	RMSE	MAE	RMS(WE)	Rank
√	Exponential		5	0.232137	0.243228	0.040344		1.06	0.381823	0.287530	0.047142	
√	Exponential		6	0.226354	0.235110	0.039207		1.37	0.378298	0.287202	0.047510	
√	Exponential		7	0.207103	0.195412	0.035306		5.35	0.398248	0.306021	0.046510	
√	Exponential		8	0.201124	0.192087	0.032314		5.38	0.329268	0.253775	0.041690	
√	Exponential		9	0.189634	0.168533	0.031017	1	7.05	0.354901	0.248029	0.044618	2
	Extended exponential		3	0.428955	0.851496	0.108242		0.70	1.049691	0.801251	0.099592	
√	Extended exponential		4	0.295338	0.396660	0.055221		0.61	0.535629	0.397531	0.057243	
√	Extended exponential		5	0.242640	0.272746	0.042599		0.92	0.413865	0.315290	0.049211	
√	Extended exponential		6	0.226434	0.235123	0.039550		1.25	0.380739	0.287265	0.047161	
√	Extended exponential		7	0.223335	0.234933	0.038526		1.75	0.396625	0.301678	0.047994	
√	Extended exponential		8	0.202929	0.197789	0.033883		4.39	0.366656	0.277405	0.043878	
√	Extended exponential		9	0.190527	0.166864	0.031090	2	6.69	0.330321	0.244132	0.042034	3
√	Nelson Siegel		4	0.409292	0.317191	0.051666	7	1.09	0.457923	0.355774	0.055797	8
√	Svensson		6	0.343904	0.259786	0.042984	4	1.66	0.373893	0.286103	0.046315	4
	DR spline	0	7.0	0.391426	0.456090	0.055185	9	8.72	0.958639	0.515708	0.069270	11
	DR smoothing spline	10	16.1	0.362326	0.340261	0.046535	6	17.65	0.857787	0.426483	0.064208	
	DR smoothing spline	100	13.4	0.429885	0.483216	0.062014		13.50	0.818329	0.480563	0.064167	10
	DR smoothing spline	500	10.9	0.483581	0.677227	0.080124		12.89	0.948166	0.660384	0.080797	
	YC smoothing spline	50	16.1	0.695112	1.377054	0.146659	13	43.12	2.942440	1.603997	0.180882	12
	YC smoothing spline	500	13.7	0.794033	1.860778	0.190348		24.04	3.188717	1.849011	0.199573	
√	FR spline	0	7.0	0.313442	0.402618	0.057962	10	5.11	0.339623	0.263165	0.051307	7
	FR smoothing spline	10	16.0	0.357609	0.476063	0.053299	8	8.89	0.306731	0.220569	0.040150	
	FR smoothing spline	25	15.3	0.371786	0.513293	0.057050		7.17	0.315516	0.225975	0.039600	
√	FR smoothing spline	50	14.8	0.380255	0.537181	0.059715		6.12	0.323676	0.231938	0.039526	1
√	FR smoothing spline	100	14.2	0.388818	0.562412	0.062546		5.29	0.335724	0.240705	0.040041	
√	FR smoothing spline	200	13.7	0.397274	0.589033	0.065480		4.64	0.351323	0.250544	0.040746	
√	FR smoothing spline	300	13.3	0.402054	0.604314	0.067184		4.32	0.361454	0.257373	0.041239	
√	FR smoothing spline	500	12.8	0.407504	0.621589	0.069143		3.98	0.374589	0.267163	0.042023	
√	FR smoothing spline	1000	12.2	0.413898	0.645352	0.071783		3.58	0.393280	0.282444	0.043400	
√	FR smoothing spline	2500	11.4	0.421670	0.674491	0.075473		3.06	0.415616	0.301517	0.046137	
√	FR smoothing spline	5000	10.7	0.427769	0.698794	0.079096		2.67	0.432412	0.316333	0.049124	
√	FR smoothing spline	10,000	10.1	0.432071	0.718581	0.083006		2.30	0.451701	0.334222	0.052679	
√	FR smoothing spline	25,000	9.2	0.439624	0.749014	0.088497		1.92	0.473966	0.353703	0.056711	
√	FR smoothing spline	50,000	8.6	0.444249	0.767239	0.091876		1.73	0.489820	0.365768	0.059231	
√	FR smoothing spline	100,000	8.1	0.449126	0.787172	0.095378		1.60	0.505753	0.378494	0.061699	
√	FR Anderson Sleath		11.1	0.369804	0.529695	0.063605	11	3.04	0.358563	0.274375	0.047523	5

The curvature statistic is then calculated as the mean of the absolute value of the curvature in the range $t = \{1.0, 1.01, 1.02, \ldots, m\}$, where m is the longest maturity:

$$C = \frac{1}{100(m-1)} \sum_{i=100}^{100m} \left| \frac{d^2 f(i/100)}{dt^2} \right| \tag{26}$$

The forward rate curvature is calculated for the various yield curve models and the results are shown in Table 5 (reported numbers are $\times 10^{-4}$). Note that for the 3 parameter yield curve polynomial the curvature is zero. This is because by definition, in this case, the forward rate curve will be a straight line.

To understand how the forward rate curves for the various models behave, the yield curve and the forward rate on one particular date (5 January 1998) are graphed for each important model (see Figs. 1–9) of the fixed coupon gilt yield curve. As in Anderson and Sleath (2001) the yields of gilt strips are also plotted. The gilt strip yields give useful reference points for each yield curve and make it easier to see the subtle differences between them. It needs to be stressed that the gilt strips have no part in the estimation of the various yield curves, which are estimated purely from the fixed coupon gilts. In addition, any departure of a yield curve from the gilt strip yields does not mean that the curve is in any way wrong; it may be that the gilt strips themselves are not correctly priced. Excessive curvature of the forward rate curve is not a valuable attribute of a yield curve model. Therefore, also included with each graph is the measure of curvature for the forward rate curve defined in Eq. (26).

In Fig. 1 we see that the Svensson model has a yield curve which tends, asymptotically, to a fixed yield value. This curve can be specified either as a yield curve or a forward rate curve (see Section 3.1) and it is therefore little surprise that the forward rate curve is well behaved. Indeed on the day in question this curve had the lowest curvature. Figs. 2 and 3 are the two polynomial models, of the discount curve and the yield curve respectively. Here the yield curve polynomial exhibits less forward rate curvature. Figs. 4–6 are the three discount curve models, the Exponential, the extended exponential and the Fourier respectively. The forward curves are similar with the extended exponential exhibiting the least curvature and the Fourier showing extreme curvature in the longest maturities. Figs. 7–9 are three smoothing splines of the discount rate, yield curve (actually $r(t) \cdot t$) and the forward rate curve respectively. The discount rate and yield curve curves exhibit the "knuckles" associated with the knot points of the cubic spline. The forward rate spline is smoother and this dramatically demonstrates the benefit of actually fitting cubic splines to the forward rate curve.

In Table 5, it is notable that some models have low out of sample errors, but high curvature (e.g. the forward rate smoothing spline with $\lambda = 10$), whilst others (e.g. the Svensson and the exponential models) have both low errors and low curvature. It is apparent that the smoothed forward rate splines have the required combination of both low out of sample errors and low forward rate curvature. The smoothed discount rate spline and the smoothed yield curve splines do not have the required out of sample fit and forward rate curvature properties.

There are various possibilities for the smoothing function $\lambda(t)$; indeed the identification of the appropriate function could be the subject of a considerable study itself. We evaluated three different

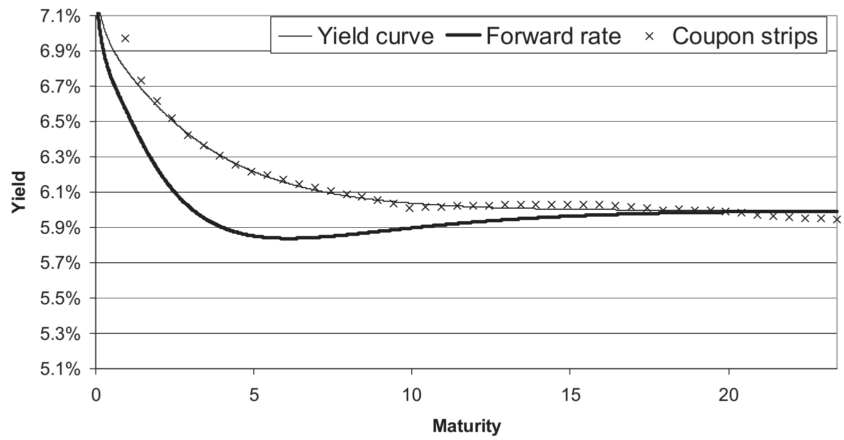

Fig. 1. Svensson model (forward rate curvature = 1.8). UK yield curve estimates on 5th January 1998.

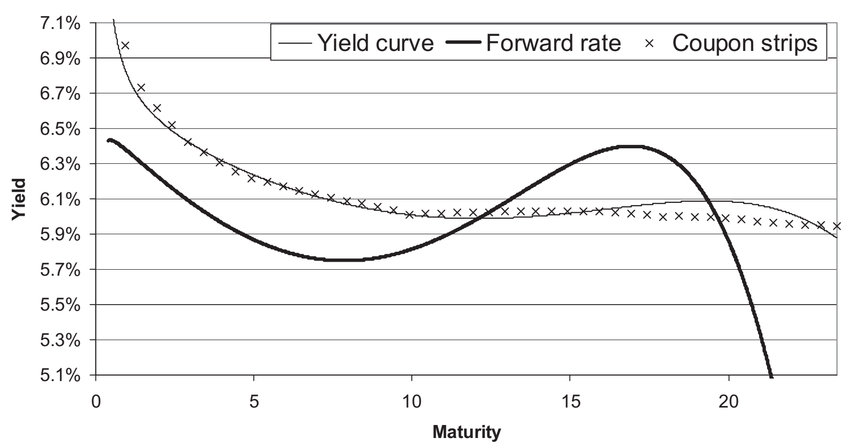

Fig. 2. Discount rate polynomial model (forward rate curvature = 7.8). UK yield curve estimates on 5th January 1998.

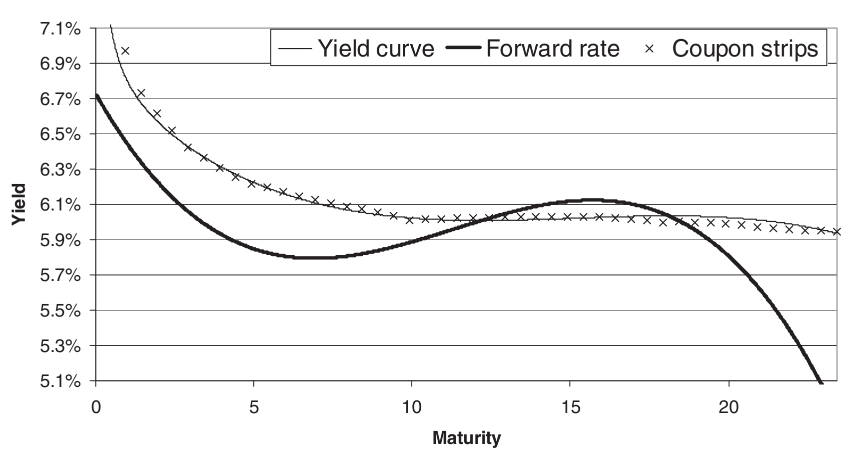

Fig. 3. Yield curve polynomial model (forward rate curvature = 3.3). UK yield curve estimates on 5th January 1998.

potential models for $\lambda(t)$. To do this we calculated the curvature and out of sample error statistics on a sample of 30 days taken at random from the entire period. The sample errors and forward rate curvature for these three different models for $\lambda(t)$ are shown in Fig. 10. The models were the FNZ: $\lambda(t) = x \; \forall \; t$, the VRP: $\lambda(t) = x + xt$ and the Anderson Sleath: $\log \lambda(t) = x - x[\exp(-t/1.44)]$. We can see that the simplest function for the smoothing parameter (the FNZ type model $\lambda(t) = x \; \forall \; t$) gives the best combination of forward rate curvature and out of sample error statistics and therefore this is the model used in the empirical research. Similar results (not reported) were also observed for the smoothing spline of the discount rate and the yield curve.

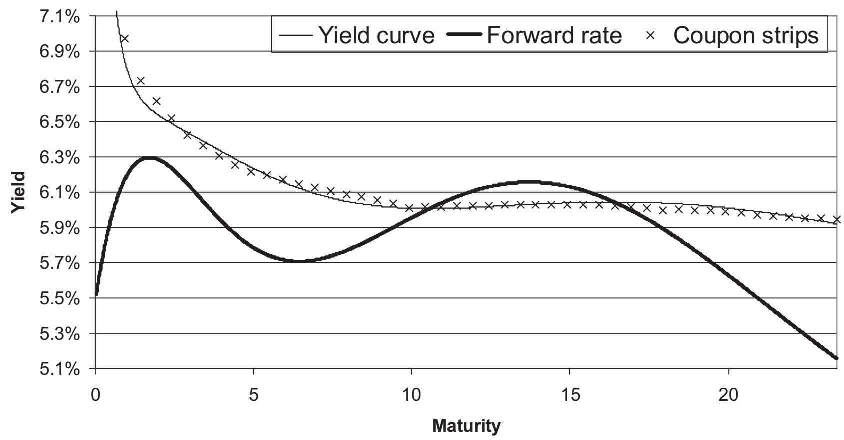

Fig. 4. Exponential model (forward rate curvature = 4.8). UK yield curve estimates on 5th January 1998.

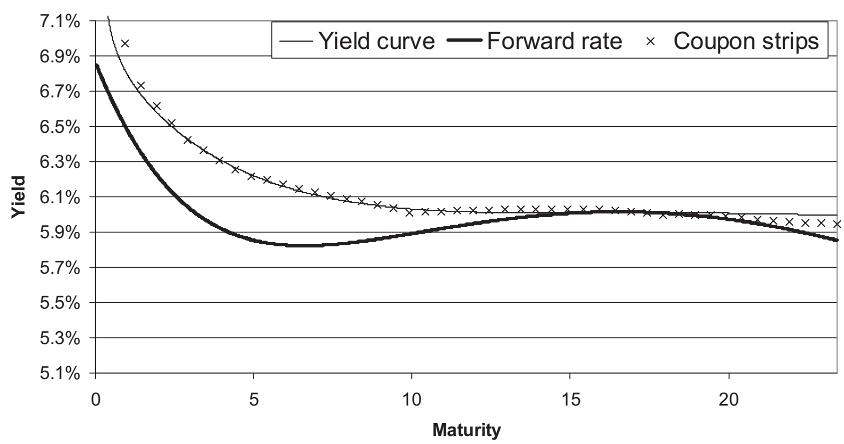

Fig. 5. Extended exponential model (forward rate curvature = 1.9). UK yield curve estimates on 5th January 1998.

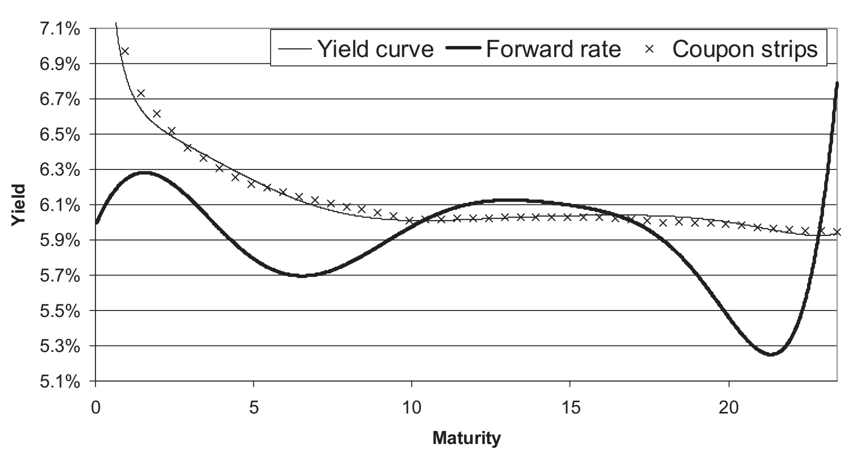

Fig. 6. Fourier model (forward rate curvature = 6.6). UK yield curve estimates on 5th January 1998.

In Figs. 11 and 12 the curvature and out of sample error statistics are plotted for the best models in Periods 1 and 2 respectively. The smoothing spline models of the forward rate curve are denoted by a curve. This is because, by varying the smoothing parameter, any of these combinations of errors and curvature could be achieved. The other models are denoted by a fixed point. With the appropriate amount of smoothing, the smoothing spline models of the forward rate curve perform well as they have the lowest out of sample errors. If the curvature of the forward rate is taken into consideration, then some of the other models also perform

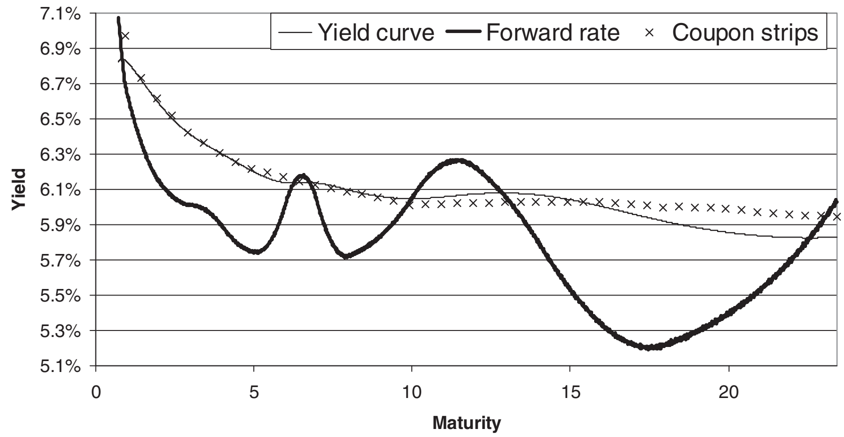

Fig. 7. Smoothed spline of discount rate (forward rate curvature = 20.8). UK yield curve estimates on 5th January 1998.

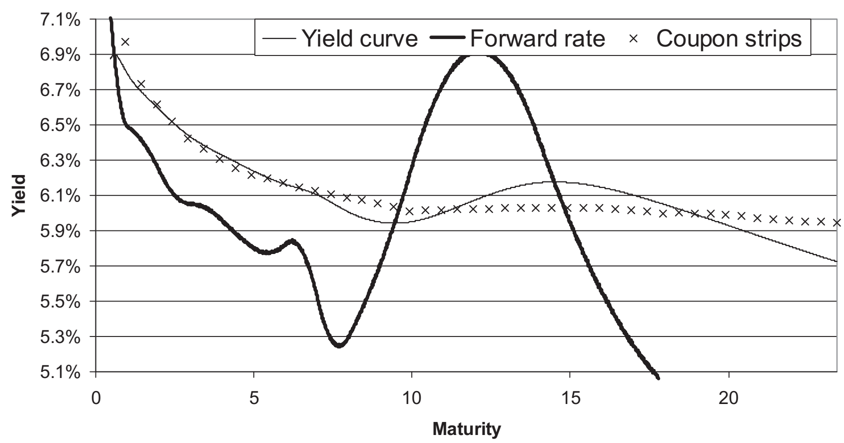

Fig. 8. Smoothed spline of yield curve (forward rate curvature = 24.9). UK yield curve estimates on 5th January 1998.

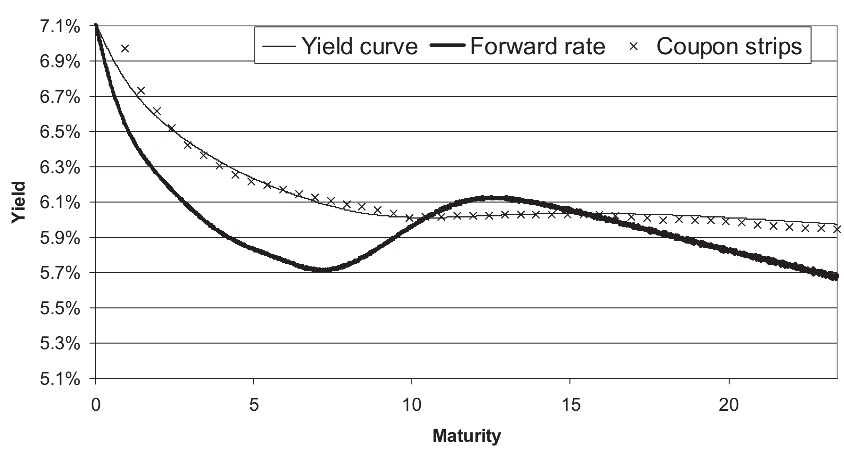

Fig. 9. Smoothed spline of forward rate curve (forward rate curvature = 8.2). UK yield curve estimates on 5th January 1998.

well. In period 1, the 4 and 5 parameter exponential and extended exponential models and the 4 parameter yield curve polynomial exhibit lower forward rate curvature than the equivalent smoothing spline models.

In period 2, some smoothing spline models of the forward rate curve also provide the lowest out of sample errors. In addition, it is notable that both the 8 parameter exponential and the 9 parameter extended exponential models have out of sample errors that are comparable with the best forward rate smoothing spline models. If forward rate curvature is also taken into account, 4–6 parameter exponential; 5–7 parameter extended exponential models, Nelson Siegel, Svensson, and 5 and 6 parameter yield curve polynomial

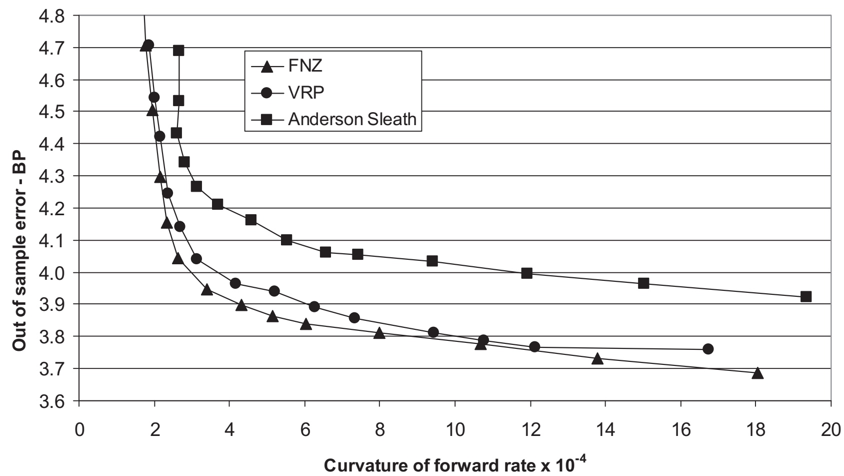

Fig. 10. Choice of smoothing function, $\lambda(t)$, for forward rate smoothing spline models FNZ: $\lambda(t) = x \ \forall \ t$ VRP: $\lambda(t) = x + xt$ Anderson Sleath: $\log \lambda(t) = x - x[\exp(-t/1.44)]$.

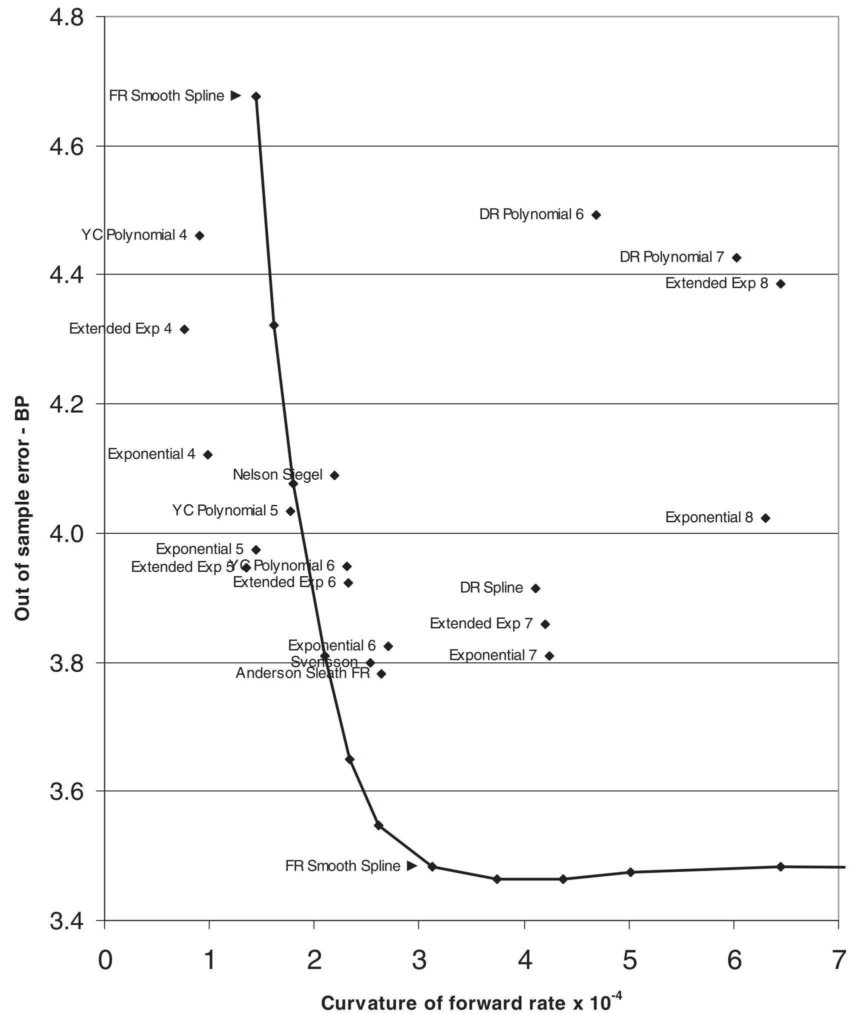

Fig. 11. UK yield curve models out of sample errors and forward rate curvature for period 1 (12 July 1996–26 May 2005).

models all exhibit lower forward rate curvature than the equivalent smoothing spline models.

10. Conclusions

In this research we examine alternative models of the yield curve and select the best models according to two criteria. The first criterion is that a superior yield curve model should have low out of sample errors in pricing bonds; the second criterion is that a superior yield curve model should imply a forward rate curve with low curvature.

We suggest that the GCV criterion of Fisher et al. (1995) used in selecting the smoothing parameter in smoothing spline models is too arbitrary and it is therefore better to explicitly measure the actual shape (curvature) of the forward rate curve.

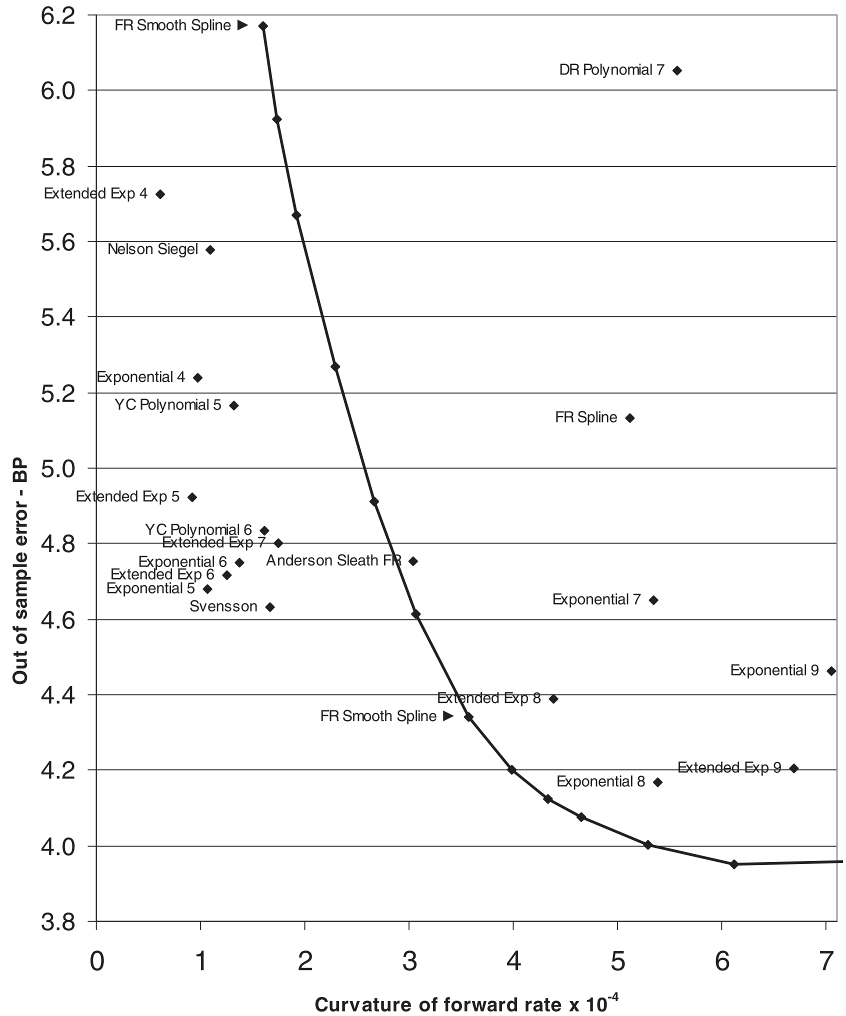

Fig. 12. UK yield curve models out of sample errors and forward rate curvature for period 2 (27 May 2005–10 February 2010).

We cannot support the conclusions of Ioannides (2003) who found that the Svensson (1994) and the Nelson and Siegel (1987) models out performed the various spline alternatives. This research suggests that the smoothing spline model of the forward rate can give superior results.

This study shows that the smoothing spline models of the forward rate curve also have superior properties to smoothing splines of either the yield curve or the discount rate. However the effective use of smoothing spline model requires considerable adroitness. The user needs to choose the positioning rules for the knot points, select a suitable smoothing function and use an appropriate amount of smoothing. The Anderson and Sleath (1999, 2001) model (a smoothing spline of the forward rate) used by the Bank of England, also performs very well. Ironically the results for the Anderson Sleath model (1999, 2001) are not dissimilar to those of the Svensson (1994) model, the technique that was previously used by the Bank of England.

Polynomial models of the yield curve have not received much attention since Chambers et al. (1984). In this study the yield curve polynomial model included a t^{-1} term which is not present in the Chambers et al. (1984) polynomial models. Our results show that the yield curve polynomial models (with the added t^{-1} term) work well and can perform slightly better than comparable forward rate spline models.

The Exponential model of Li et al. (2001) and Bolder and Gusba (2002) as used by Bank of Canada also performs well. The addition of an intercept term results in a model (the extended exponential) with similar error statistics but much reduced forward rate curvature. These models can also perform well compared to the forward rate spline models, particularly in the period since 27 May 2005 when the UK gilt market added its first 50 year maturity. The exponential model also has the advantage of being particularly easily estimated using GLS methods. Overall we favour the exponential model for estimating the UK gilt yield curve due to its combination of accuracy and parsimony in the period of the study.

Acknowledgements

I am grateful for helpful comments from Ser-Huang Poon, Richard Stapleton and the anonymous referee.

Appendix A. Estimation of yield curve models using Taylor series iteration

A Taylor series approximation to $\widehat{\underline{P}}(\underline{\theta})$ can be given by:

$$\widehat{\underline{P}}(\underline{\theta}) \approx \widehat{\underline{P}}(\underline{\theta}^0) + \underline{X}(\underline{\theta}^0)(\underline{\theta} - \underline{\theta}^0)$$

where

$$\underline{X}(\underline{\theta}^0) = \left.\frac{\partial \underline{\widehat{P}}(\underline{\theta})}{\partial \underline{\theta}^T}\right|_{\underline{\theta}=\underline{\theta}^0} \tag{A.1}$$

Let

$$\underline{Y}(\underline{\theta}^0) = \underline{P} - \underline{\widehat{P}}(\underline{\theta}^0) + \underline{X}(\underline{\theta}^0)\underline{\theta}^0 \tag{A.2}$$

Then using $\underline{P} = \underline{Y}(\underline{\theta}^0) + \underline{\widehat{P}}(\underline{\theta}^0) - \underline{X}(\underline{\theta}^0)\underline{\theta}^0$ from (A.2) and $\underline{\widehat{P}}(\underline{\theta}) \approx \underline{\widehat{P}}(\underline{\theta}^0) + \underline{X}(\underline{\theta}^0)(\underline{\theta} - \underline{\theta}^0)$ from (A.1) we get:

$$\underline{P} - \underline{\widehat{P}}(\underline{\theta}) \approx \underline{Y}(\underline{\theta}^0) - \underline{X}(\underline{\theta}^0)\underline{\theta} \tag{A.3}$$

We can express the objective function to minimise:

$$[\underline{Y}(\underline{\theta}^0) - \underline{X}(\underline{\theta}^0)\underline{\theta}]^T \underline{W}[\underline{Y}(\underline{\theta}^0) - \underline{X}(\underline{\theta}^0)\underline{\theta}] \tag{A.4}$$

For which the generalised least squares solution is given by:

$$\underline{\theta}^1 = \frac{\underline{X}(\underline{\theta}^0)^T \underline{W}\underline{Y}(\underline{\theta}^0)}{\underline{X}(\underline{\theta}^0)^T \underline{W}\underline{X}(\underline{\theta}^0)} \tag{A.5}$$

We then use $\underline{\theta}^1$ in place of $\underline{\theta}^0$ in Eqs. (A.1) and (A.5) and repeat the optimisation to obtain $\underline{\theta}^2$. When $|\underline{\theta}^i - \underline{\theta}^{i-1}|$ is small the optimum has been found. Provided $\underline{\theta}^0$ is sufficiently close to $\underline{\theta}$ the sequence will converge to the value of $\underline{\theta}$, see Bolder and Gusba (2002).

References

Aït-Sahalia, Y., Kimmel, R., 2010. Estimating affine multifactor term structure models using closed form likelihood expansions. Journal of Financial Economics 98, 113–144.

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

Anderson, N., Sleath, J., 2001. New Estimates of the UK Real and Nominal Yield Curves. Working Paper No. 2001-126. Bank of England.

Bank for International Settlements, 2005. Zero Coupon Yield Curves: Technical Documentation. Working Paper No. 2005-25. Bank for International Settlements.

Bliss, R.R., 1986. Testing Yield Curve Estimation Methods. Working Paper No. 96-11. Federal Reserve Bank of Atlanta.

Bolder, D.J., Gusba, S., 2002. Exponentials, Polynomials and Fourier Series: More Yield Curve Modelling at the Bank of Canada. Working Paper No. 2002-29. Bank of Canada.

Broyden, C.G., 1970. The convergence of a class of double-rank minimization algorithms 1. General considerations. Journal of the Institute of Mathematics and its Applications 6, 76–90.

Chambers, D.R., Carlton, W.T., Waldman, D.W., 1984. A new approach to estimation of the yield curve of interest rates. Journal of Financial and Quantitative Analysis 19, 233–252.

Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D., 2011. The affine arbitrage-free class of Nelson–Siegel term structure models. Journal of Econometrics 164, 4–20.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the yield curve of interest rates. Econometrica 53, 385–407.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and their Applications. Cambridge University Press, Cambridge, England, UK.

de La Grandville, O., 2001. Bond Pricing and Portfolio Analysis. MIT Press, Cambridge, Massachusetts, USA.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364.

Fernández-Rodríguez, F., 2006. Interest rate term structure modelling using free-knot splines. Journal of Business 79, 3083–3099.

Fisher, M., Nychka, D., Zervos, D., 1995. Fitting the Yield Curve of Interest Rates with Smoothing Splines. Finance and Economics Discussion Series, No. 95-1. Federal Reserve Board.

Fletcher, R., 1970. A new approach to variable metric algorithms. Computer Journal 13, 317–322.

Goldfarb, D., 1970. A family of variational methods derived by variational means. Mathematics of Computation 24, 23–26.

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the yield curve of interest rates: a new methodology for contingent claims valuation. Econometrica 60, 77–105.

Ioannides, M., 2003. A comparison of yield curve estimation techniques using UK data. Journal of Banking and Finance 27, 1–26.

Jarrow, R., Ruppert, D., Yu, Y., 2004. Estimating the interest rate yield curve of corporate debt with a semiparametric penalized spline model. Journal of the American Statistical Association 99, 57–66.

Joslin, S., Singleton, K.J., Zhu, H., 2011. A new perspective on Gaussian dynamic term structure models. Review of Financial Studies 24, 926–970.

Laurini, M.P., Moura, M., 2010. Constrained smoothing B-splines for the term structure of interest rates. Insurance: Mathematics and Economics 46, 339–350.

Li, B., DeWetering, E., Lucas, G., Brenner, R., Shapiro, A., 2001. Merrill Lynch Exponential Spline Model. Working Paper, Merrill Lynch.

McCulloch, J.H., 1971. Measuring the yield curve of interest rates. Journal of Business 44, 19–31.

McCulloch, J.H., 1975. The tax adjusted yield curve. Journal of Finance 30, 811–829.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modelling of yield curves. Journal of Business 60, 473–489.

Press, W.H., Teukolsky, S.A., Vettering, W.T., Flannery, B.P., 2001. Numerical Recipes in C++: The Art of Scientific Computing, second ed. Cambridge University Press, Cambridge, England, UK.

Schaefer, S.M., 1982. Tax-induced clientele effects in the market for British government securities. Journal of Financial Economics 10, 121–159.

Shanno, D.F., 1970. Conditioning of quasi-newton methods for function minimization. Mathematics of Computation 24, 647–656.

Shea, G.S., 1984. Pitfalls in smoothing interest rate yield curve data: equilibrium models and spline approximations. Journal of Financial and Quantitative Analysis 19, 253–269.

Siegel, A.F., Nelson, C.R., 1988. Long term behavior of yield curves. Journal of Financial and Quantitative Analysis 23, 105–110.

Steeley, J.M., 1991. Estimating the gilt-edged yield curve: B-splines and confidence intervals. Journal of Business Finance and Accounting 18, 513–529.

Svensson, L.E., 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992–1994. Working Paper No. WP/94/114. International Monetary Fund, Washington, DC.

Waggoner, D., 1997. Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices. Working Paper No. 97-10. Federal Reserve Bank of Atlanta.