New estimates of the UK real and nominal yield curves

Nicola Anderson
and
John Sleath

The views expressed are those of the authors and do not necessarily reflect those of the Bank of England. The authors wish to thank Robert Bliss and Dan Waggoner for many helpful comments and suggestions at the outset of this work. Thanks also go, without implication, to Creon Butler, Roger Clews, John Lumsden and Jim Steeley.

Copies of working papers may be obtained from Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH; telephone 020-7601 4030, fax 020-7601 3298, e-mail mapublications@bankofengland.co.uk.

Working papers are also available at www.bankofengland.co.uk/workingpapers/index.htm

The Bank of England's working paper series is externally refereed.

ISSN 1368-5562

Contents

Abstract 5

1 Introduction 7

2 Model selection and overview 11
- 2.1 Parametric models 12
- 2.2 Spline-based models 13

3 Estimation and results 18
- 3.1 Smoothness and flexibility 18
- 3.2 Stability of the curves 21
- 3.3 Summary 25

4 Improving estimates of the short end of the yield curve 26
- 4.1 Choice of additional data 26
- 4.2 Results of incorporating GC repo rates 27

5 Estimating the real and implied inflation term structures 30
- 5.1 Evans' theoretical framework 31
- 5.2 Two modifications 32
- 5.3 Stability of the estimated real curve 35
- 5.4 Comparison of new approach with the ITS technique 38

6 Conclusions 41

Appendix: Smoothing cubic spline models 42

References 43

Abstract

This paper presents some new estimates of the UK real and nominal yield curves. These estimates are derived using a spline-based technique put forward by Waggoner (1997), modified for the UK government bond markets. At the short end of the nominal yield curve, additional data are included from the GC repo market. Estimates of the real yield curve are derived from the prices of index-linked gilts within a modified version of the framework put forward by Evans (1998). It is found that the new yield curves outperform existing methods on a number of criteria that are designed to examine the suitability of estimates for the purpose of assessing monetary conditions. In particular, the estimates are found to be smooth across maturity while having sufficient flexibility to describe the shape of the curve at shorter maturities where expectations are relatively precise. The curves are also robust to small errors in the data.

1 Introduction

In this paper, we present some new estimates of the UK real and nominal yield curves, for the purpose of assessing monetary conditions. These estimates differ from those presented in previous studies in a number of ways. First, the yield curves are estimated using a method put forward by Waggoner (1997) for the United States, adapted by us for the UK government bond market. Second, data from the generalised collateral (GC) repo market are used to provide improved estimates of the nominal yield curve at shorter maturities. Third, estimates of the real yield curve are extracted from the prices of index-linked gilts within a modified version of the framework suggested by Evans (1998).

Before describing each of these developments in detail, it is worth briefly reviewing the main problems involved in extracting useful information from the prices of conventional and index-linked gilts. The most basic type of information we are interested in estimating is the implied forward rates of interest at various horizons. These are important in their own right as they reflect, albeit imperfectly, the market's expectations about the future path of interest rates. They also provide the building-blocks for calculating other term structure variables, such as zero-coupon yields, as well as enabling comparisons between the returns on government bonds and other assets, for example, the credit spread embodied in the prices of corporate bonds.

Implied forward rates of interest are defined as the marginal rates of return that investors require in order to hold bonds of different maturities. The set of 'instantaneous' forward rates, $f(m)$, are related to the price, $B(t)$, of a t-maturity zero-coupon bond by:[1]

$$B(t) = \exp\left[-\int_0^t f(m)dm\right] \qquad \textbf{(1)}$$

Equation **(1)** shows that to measure these forward rates directly from the market requires a set of observable zero-coupon bond prices across a

[1] We focus on instantaneous forward rates of interest as these can be used to derive forward rates and yields between any two maturities of interest. For example, the *s*-period forward rate at some point t in the future is given by $f(t, t+s) = \int_t^{t+s} f(m)dm$.

continuum of maturities (the 'discount function'). But in practice, the discount function is not directly observable – we only observe the prices of coupon-bearing bonds.[2] We can, however, write the price of each observable bond in terms of this discount function. So letting c_i denote the cash flow due on the bond at time t_i and n refer to the number of such payments outstanding, we express the price of the bond, $P(c_i, t_i, i=1,\ldots n)$, as:

$$P(c_i, t_i, i = 1,\ldots, n) = \sum_{i=1}^{n} c_i B(t_i) \qquad \textbf{(2)}$$

Together with equation **(1)**, this shows that there is a direct relationship between the bond prices we observe and the instantaneous forward rates we want to measure. Moreover, equation **(2)** can be defined for cash flows denominated in either real or nominal terms, with the forward rates identified accordingly.

The identification of real and nominal forward curves via this relationship means overcoming three sets of problems. These are:

- Both conventional and index-linked bonds are issued for only a finite set of maturities. We therefore need a method of disentangling the cash flows on the bond and 'filling in the gaps' to give a continuous curve.

- The real value of cash flows on index-linked bonds is not known with certainty. Hence, to estimate the real yield curve we need a method for approximating the value of these cash flows.

- For both the real and nominal yield curves, there is a lack of data at shorter maturities. Forward rates along this portion of the curve cannot therefore be identified.

(2) In fact, zero-coupon gilts have existed since the introduction of the strips market in December 1997. These separate the two components of a coupon-bearing gilt to give a principal strip with maturity equal to its redemption date and a series of coupon strips related to each payment date. The market in strips is, however, still small relative to coupon-bearing gilts, with thin trading. We therefore do not use strips prices to estimate the yield curve.

Methods for addressing the first of these problems have been available for more than 30 years, and these have been used in the Bank of England to estimate the nominal yield curve. For the past five years, in common with many other central banks, we have used the estimation method proposed by Svensson (1994, 1995). This is a parametric method, with the entire forward curve characterised by a single set of parameters representing the long-run level of interest rates, the slope of the curve and humps in the curve. Other competing models in the literature include the more parsimonious functional form of Nelson and Siegel (1987), and the spline-based methods of McCulloch (1971, 1975) and Fisher, Nychka and Zervos (1995).

Estimation of the real yield curve is a more recent innovation, made possible by the introduction of index-linked bonds in the United Kingdom in 1981. As noted above, additional problems arise because the value of the cash flows on these bonds is not known with certainty – this is because they are indexed only imperfectly to the price level. We therefore have to use information from the nominal yield curve to extract the real risk-free rates of interest embodied in their prices. Until now the Bank of England has been using an iterative technique developed by Deacon and Derry (1994) in which the real yield curve is described by a restricted version of Svensson's model.

For maturities of less than two years, estimates of both the real and nominal yield curves have not been thought reliable, and as a result have not been used by the Bank's Monetary Policy Committee, nor published in the *Inflation Report* or *Quarterly Bulletin*. This is mainly because there are few gilts at the short end (with terms to maturity of two years or less), where expectations may be relatively precise, and where the curve may be expected to have quite a lot of curvature. This is a particular problem with the index-linked gilt market in which there are relatively few observations across the entire curve.

In this paper, we reassess each of these problems in turn. In the first case, experience has led us to question whether the Svensson technique we currently use provides the best estimate of the 'true' yield curve, even at the longer maturities. We therefore examine four alternative methods of estimation, according to the following criteria:

(1) Smoothness – the technique should give relatively smooth forward curves, rather than trying to fit every data point, since the aim is to supply a market expectation for monetary policy purposes, rather than a precise pricing of all bonds in the market. Nonetheless, subject to the former, a better fit to the data would be preferred.

(2) Flexibility – the technique should be sufficiently flexible to capture movements in the underlying term structure. More flexibility is likely to be needed at shorter maturities (where expectations are better informed and more subject to revision as news reaches the market) than at the longer end.

(3) Stability – estimates of the yield curve at any particular maturity should be stable in the sense that small changes in the data at one maturity (such as at the long end) do not have a disproportionate effect on forward rates at other maturities.

The aim is to find the yield curve model that will provide us with the most reliable and useful estimates, not only on any particular day, but also over time. Based on these criteria, we find that a technique developed by Waggoner (1997), but modified for the United Kingdom, outperforms the competing methods.

Having chosen this as our basic model, we then turn our focus to the short end of the yield curve, where there is a lack of data in both the conventional and index-linked gilt markets. The challenge is to investigate whether there are alternative sources of data that can feasibly be included to help fill the gaps. In the case of the real yield curve, there is very little we can do – index-linked gilts are the only direct source of real interest rate data, at least in the United Kingdom. For nominal yields, however, we find that data from the GC repo market can be used to supplement the dataset at the short end of the conventional gilt market.

Finally, we re-examine estimates of the real yield curve in two ways. First, we use a modified version of the Waggoner technique to fit a curve to index-linked gilt prices. Second, we improve on the iterative technique used to extract real rates of interest from these prices, within the framework suggested by Evans (1998). This also leads to improved measures of the inflation term structure (ITS), which is the difference between the real and nominal yield curves.

The rest of the paper is structured as follows. In Section 2 we set out the four methods of yield curve estimation that form the basis of our comparison. In Section 3 we describe the results of our comparison based on the set of criteria outlined previously. Section 4 introduces the use of GC repo rates to improve estimates at the short end of the curve, and Section 5 illustrates how the Waggoner-based curve can be adapted to estimate real and inflation term structures from the prices of index-linked bonds, specifically within the framework put forward by Evans (1998). Finally, Section 6 concludes.

2 Model selection and overview

We investigate four alternative methods for estimating the term structure. Two of these are parametric: Nelson and Siegel (1987) hereafter NS, and Svensson (1994, 1995) hereafter SV; and two are based on cubic splines: Fisher, Nychka and Zervos (1995) hereafter FNZ, and a modified version of Waggoner's variable roughness penalty model (1997) hereafter VRP. An overview of these models, highlighting the main differences between them is given in Table A.

Table A: Overview of the four models

	Parametric models		**Spline-based models**	
Property	***SV***	***NS***	***FNZ***	***VRP***
Nature of forward rate curve	Forward rate is single function defined at all maturities		Forward rate is piecewise cubic polynomial with segments joined at knot points	
No. of parameters	6	4	Approx. 14 depending on number of bonds	
Objective function	Minimise residual sum of squares	Minimise residual sum of squares	Minimise residual sum of squares plus roughness penalty	Minimise residual sum of squares plus roughness penalty
Pre-specified parameters	None	None	Number of knot points	Number of knot points, smoothing function
Constraints	Long-run asymptote	Long-run asymptote	None	None

Notice that the only property all these models have in common is that forward rates are modelled as a function of a set of underlying parameters. The models then differ in terms of the way in which this function is specified and the criterion used to derive optimal estimates of the parameters. These differences, which are discussed further below, are particularly important in understanding the trade-off between the flexibility of the functional form, and the way in which the resulting yield curve estimates are made to be smooth across maturity.

2.1 Parametric models

Parametric models offer a conceptually simple and parsimonious description of the term structure of interest rates. Smooth yield curve estimates are ensured by modelling the instantaneous forward rate curve as a function, $f(m, b)$, of a relatively small vector of parameters, b. The degree of flexibility of the curve is then largely determined by the number of parameters to be estimated. Of the two curves we estimate, the least flexible is the NS model. In this case, forward rates are defined as:

$$f\left(m, ß\right) = b_0 + b_1 \exp\left(\frac{-m}{t_1}\right) + b_2 \frac{m}{t_1} \exp\left(\frac{-m}{t_1}\right) \qquad \textbf{(3)}$$

so that there are four parameters to be estimated, $b = (b_0, b_1, b_2, t_1)$. These parameters can be interpreted as being related to the long-run level of interest rates, the short rate, the slope of the yield curve and a hump in the curve. The SV model then extends this parameter set to $b = (b_0, b_1, b_2, b_3, t_1, t_2)$, where the additional parameters can be interpreted as allowing an additional hump in the curve. Forward rates are now given by:

$$f\left(m, ß\right) = b_0 + b_1 \exp\left(\frac{-m}{t_1}\right) + b_2 \frac{m}{t_1} \exp\left(\frac{-m}{t_1}\right) + b_3 \frac{m}{t_2} \exp\left(\frac{-m}{t_2}\right) \qquad \textbf{(4)}$$

As noted in Table A, both models are constrained to converge to a constant level. The rationale for this constraint is based on the assumption that forward rates reflect expectations about future short interest rates, or equivalently that the unbiased expectations hypothesis holds. It then seems implausible that agents will perceive a different path for the future short rate in 30 years' time compared with, say, 25 years. Hence, we should expect to see constant expectations and forward rates at the long end.

To estimate the two yield curves, the functional forms can be used via equations **(1)** and **(2)** to derive a fitted value for each bond price, given the set of underlying parameters. The parameters are then estimated to minimise an objective function that compares these fitted values with observations from the gilt market. A variety of objective functions is available to us; we choose to minimise:

$$X_P = \sum_{i=1}^{N} \left[\frac{P_i - \Pi_i(b)}{D_i} \right]^2 \quad \textbf{(5)}$$

where P_i is the observed price of the i^{th} bond, D_i is its modified duration and $\Pi_i(b)$ is the fitted price. This is approximately equal to minimising the sum of squared yield residuals (although it is much quicker to calculate) and so implies roughly equal yield errors, irrespective of maturity.

2.2 Spline-based models

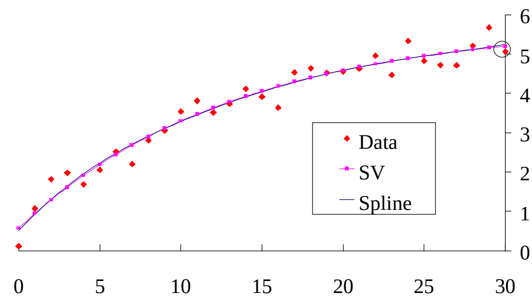
The spline-based techniques model forward rates as a piecewise cubic polynomial, with the segments joined at so-called knot points. The coefficients of the individual polynomials are restricted so that both the curve and its first derivative are continuous at all maturities. Typically, this results in 14 parameters (two more than the number of knot points) that have to be estimated.(3) This approach allows for a much higher degree of flexibility than either of the two parametric models, and accounts for the principal advantage of spline-based techniques over parametric methods. Specifically, the individual curve segments can move almost independently of each other (subject to the continuity constraints), so that separate regions of the curve are less affected by movements in nearby areas. This is in contrast to the parametric forms for which a change in the data at any one point can affect the entire curve, as estimates at any maturity are a function of all the parameters to be estimated.

---

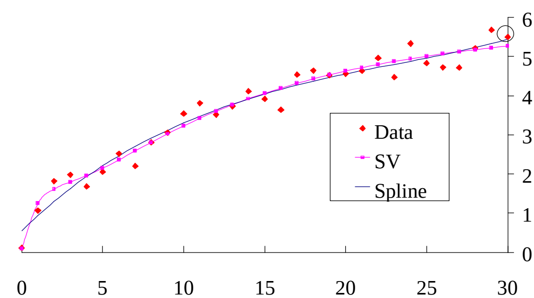
(3) To ensure numerical stability, we choose to represent the splines using basis functions, which means that the estimated parameters cannot be related in a simple manner to the coefficients of the polynomials – see the appendix for details.

Chart 1: Svensson method versus cubic spline[(a)]

(i) Original set of data points



(ii) Change of single data point



[(a)] The data used in this example are purely illustrative chosen at random, and hence the curves should not be interpreted as yield curves.

This is clearly illustrated in Chart 1, which shows a simple non-linear least squares regression to an arbitrary set of data points, using both the Svensson functional form and a cubic spline.[(4)] When a single data point is changed at the *long* end, the Svensson curve changes dramatically, particularly at the

[(4)] This is a much simpler problem than estimating term structures from coupon bond prices, but illustrates the point more clearly. The spline has been chosen to have the same number of degrees of freedom as the Svensson curve.

short end, whereas the spline moves only slightly to accommodate the new data, and only at the long end. This is because the cubic spline is much more flexible than the parametric functional form, and hence is better able to accommodate different patterns in the data.

The downside of this result is that an unconstrained spline, such as the one used in Chart 1, would be far too flexible to generate yield curves appropriate for monetary policy purposes. To achieve the required degree of smoothing, we therefore modify the objective function, as in the FNZ and VRP models. As with the parametric models, the main objective is to choose the parameters so as to minimise the difference between actual and fitted values, as described by X_P in equation **(5)**. But to control the trade-off between goodness-of-fit (flexibility) and the smoothness of the curve, a roughness penalty is also included to penalise excessive curvature (measured by the square of the second derivative) of the forward curve. The size of this penalty is determined by a 'smoothness parameter', $l_t(m)$, so that the modified objective function is:[5]

$$X_S = X_P + \int_0^M l_t(m)\left[f''(m)\right]^2 dm \qquad \textbf{(6)}$$

The two methods differ according to how the value of the smoothing parameter, $l_t(m)$, is chosen and how it varies across maturity. FNZ assume that the smoothing parameter is invariant to maturity but variable over time, so that $l_t(m) = l_t$. They then use a procedure known as generalised cross-validation to choose l_t on a daily basis. In the VRP model, on the other hand, the smoothing parameter is allowed to vary across maturity, but is chosen to be constant over time, hence $l_t(m) = l(m)$. In this case, as noted in Table A, the smoothing function $l(m)$, has to be pre-chosen.

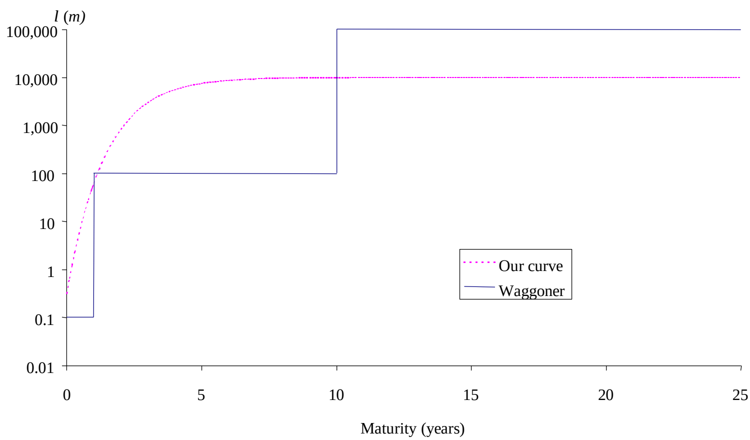
Waggoner chose a three-tiered step function for his smoothing parameter, with steps at one and ten years to maturity. This was based on the natural

---

(5) Note that this modified objective function represents the major difference between the spline-based techniques considered in this paper and earlier methods such as McCulloch (1971,1975). In these models, the objective function was specified in a similar way to those for the parametric form, that is without a roughness penalty. The degree of smoothing in the yield curve estimates was then determined by the number and positioning of the knot points. As described in the next section, these factors are less of a concern for the two methods considered in this paper as smoothing is guaranteed, at least to some extent, via the roughness penalty described in the main text.

segmentation of the US market into bills, notes and bonds, with the three levels chosen roughly to maximise the *out-of-sample* goodness-of-fit of the model. Unfortunately, the UK market cannot be naturally divided in the same way. Hence, to use Waggoner's step function approach would have meant choosing five parameters (two for the maturities at which there was a step in the function, and three for the step levels). We chose instead to define $l(m)$ as a continuous function of only three parameters.[6] Following Waggoner, the main criterion for choosing these parameters was to maximise the out-of-sample goodness-of-fit averaged over our sample period. However, it was found that many combinations of these parameters gave similar goodness-of-fit measures. We therefore opted for the set of parameters that corresponded to the highest level of smoothing among these combinations. The resulting function is plotted alongside Waggoner's step function in Chart 2. Note that we would not necessarily expect the levels to be the same (although they are close), since the smoothing functions are optimised for different markets.

Chart 2: Smoothing functions used by Waggoner (US market) and this paper (UK market)



[6] In particular, we specify the following function: $\log\ l(m) = L - (L - S)\exp(-m/m)$, where L, S and m are the three parameters to be estimated.

For both FNZ and VRP models, we also have to choose the placement of the knot points. In principle, this can have a very influential effect on the resulting yield curve estimates: in particular, when there is no roughness penalty, more knot points imply a less smooth curve with a greater in-sample goodness-of-fit. Thus the choice of knot points in this case determines the trade-off between the goodness-of-fit of the curve and the smoothness of the estimates. But when, as in the two spline-based models considered in this paper, excess curvature in the curve is penalised, the choice of knot points is significantly less crucial. However, in choosing their placement, we should clearly aim to ensure that there are sufficient knot points to allow the curve to reflect the underlying data.

Table B: Sensitivity of VRP forward rates to the number of knot points

Basis points	**2 year**	**10 year**	**20 year**
Average difference from knots at maturity of every third bond			
Knots at maturity of every bond	0.01	-0.03	0.08
Knots at maturity of every sixth bond	0.07	-0.04	1.58
Max. absolute difference from knots at maturity of every third bond			
Knots at maturity of every bond	0.38	0.22	0.29
Knots at maturity of every sixth bond	0.99	0.60	7.06

FNZ found that placing knots at the maturity of roughly every third bond gave the same results as placing a knot at the maturity of every bond, but took considerably less computational time. We adopt a similar rule. As the results in Table B show for the VRP method, increasing the number of knot points to every bond would have a negligible effect on the instantaneous forward rate curve. Halving the number of knot points, however, produces slightly different results, particularly at the long end. In fact, the long end of the curve is always higher than our ‘default case’, suggesting that it is insufficiently flexible to capture any downward slope at longer maturities. By placing a knot at the maturity of every third bond we ensure that the degree of smoothing is determined by the roughness penalty, and not the number of knot points.

Intuitively, there are a number of reasons to suspect that the Waggoner curve will provide us with more reliable estimates of the yield curve. First, by constraining the smoothing parameter to be maturity-invariant, the FNZ curve assumes that there is the same degree of curvature along the length of the term structure. But there are strong reasons to believe that this is not the case. In particular, investors are likely to be more informed about the precise path of the term structure at shorter and medium-term maturities (when interest rates are determined by monetary policy and business cycle conditions) than at longer maturities. Thus the curve may be too stiff at the short end and/or too flexible at the long end. Furthermore, by allowing the smoothing parameter to vary over time, FNZ estimates may be unstable in the sense that changes are driven not only by the underlying data, but also by movements in this parameter.

3 Estimation and results

We estimated each of the four curves over the sample period from 1 May 1996 to 31 December 1998.[7] In this section, we compare the results of this estimation in two ways. We begin by examining how well each method appears to fit the data, focusing in particular on the trade-off between the smoothness of the yield curve estimates and their ability to fit the data. We then provide a rigorous assessment of the stability of the estimates by looking at the condition numbers associated with each method, as suggested by Waggoner (1996).

3.1 Smoothness and flexibility

In the previous section, we described how there was likely to be a trade-off between the flexibility of the four yield curve models and the way in which estimates were made to be smooth across maturity. In this section, we attempt to make these ideas more concrete by assessing how well each method captures the shape of the underlying term structure and fits the data. We do this in two ways: first, we examine statistics for the goodness-of-fit of the four methods; and second, we compare the shape of the estimated curves with data from the strips market.

As Bliss (1997) notes, the appropriate measure of goodness-of-fit is an out-of-sample statistic. Each method will produce a high in-sample fit, but this may not be indicative of the underlying term structure. The important test is to see whether or not the estimated curve can accurately price a bond which has not been used to estimate the curve. Bliss calculates an

---

[7] The beginning of the sample period is chosen to coincide with changes to the taxation of coupon and capital gains income from gilts. These changes effectively eliminated the distortion which had previously existed as a result of the differential treatment of the two types of income. See Anderson *et al* (1994) for further detail.

out-of-sample statistic by fitting the curves using alternative bonds and then examining the residuals relating to the bonds that were not used in the estimation. This approach, however, is unsuitable for the United Kingdom since we have a relatively small number of securities in the set of data used to estimate the curve (typically 30-40). Hence, we adopt instead an approach known as 'leave one out cross-validation' (Davison and Hinkley (1997)). This involves estimating the curves many times, leaving out a different bond on each occasion. The cross-validation statistic is then the average absolute residual of the bonds omitted from the fit. The results are presented in Table C.

Table C: Out-of-sample goodness of fit statistics (£)

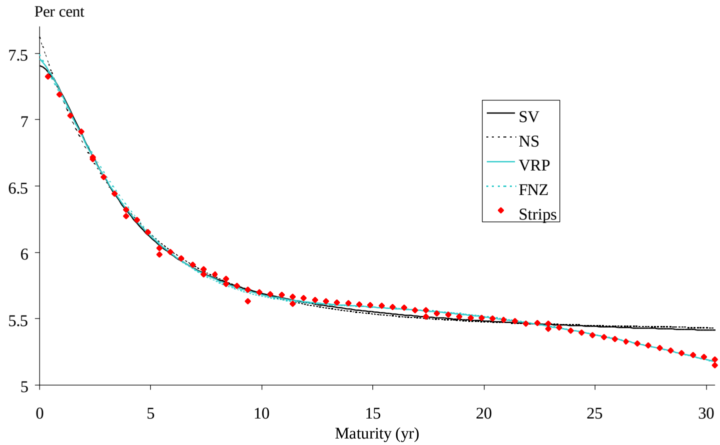
	Mean	Std dev
SV	0.090	0.024
NS	0.101	0.025
VRP	0.088	0.024
FNZ	0.116	0.043

We see that the average price errors are small for all methods (between 9 and 11 pence), although the NS and FNZ curves both have a slightly worse fit than the Svensson method. The fit of the VRP and Svensson curves is almost identical, with the former appearing to perform slightly better. On the basis of these statistics, therefore, we would be indifferent between the Svensson and VRP methods.[8]

But now consider the shape of the yield curves. In theory, we should be able to obtain a direct reading on the 'true' shape of the underlying term structure using observations from the strips market. Chart 3 compares each of the estimated curves with the yields on strips on a day chosen at random, 21 July 1998.

(8) We should note, of course, that out-of-sample goodness-of-fit was actually used as a criterion for choosing the optimal smoothing parameters in the case of the VRP curve. Hence, we would expect it to do well in this test. However, it is worth noting that when we calculated the optimal smoothing parameters over a variety of different sample periods we found them to be very stable. We would therefore not expect to observe significantly different results were we to conduct an entirely out-of-sample test for the VRP method, ie by choosing the smoothing parameters on the basis of data not used to calculate the test statistics in Table C.

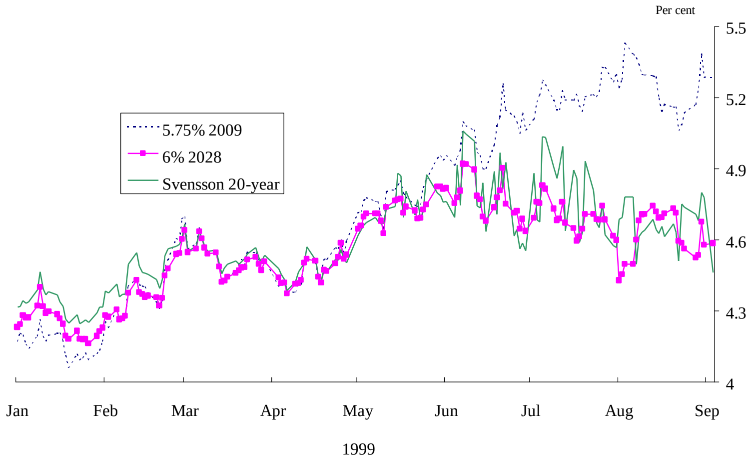
Chart 3: Comparison of all methods with strips prices (21 July 1998)



The strips' yields clearly indicate a downward-sloping term structure at the long end. And this shape is captured by both of the spline-based models. As noted previously, however, both the Nelson and Siegel and Svensson functional forms are constrained to lie flat at longer maturities – hence there is some divergence between these estimates and the strips prices themselves. In the context of our earlier discussion, one interpretation of this result is that, assuming that expectations about future interest rates do converge at longer maturities, the unbiased expectations hypothesis does not hold. In other words, strips' prices are determined by factors other than just expectations about future interest rates. Candidate explanations include the presence of risk premia and convexity terms (see, for example, Brown and Schaefer (2000)).

Of course, the strips may be mispriced (although, note that since the strips are not included in the dataset for estimation, this would not explain how the two spline-based methods match their shape). Direct evidence from the gilt market (in the form of redemption yields), however, suggests that a downward-sloping yield curve may be justified, at least over the maturity range that we consider. Moreover, it appears that forcing the long end of the curve to converge to a constant level can produce a significant amount of instability in the estimated yield curve. This is shown in Chart 4, where we plot the redemption yields on the ten-year benchmark bond and the longest-maturity bond (with maturity of 29 years), together with 20-year zero-coupon yield estimates derived using Svensson.

Chart 4: Time series of redemption yields at the long end



This illustrates the fact that, as the *observed* bond yields have diverged more and more, the yield curve *estimates* have been increasingly unstable. We attribute this to the parameterised nature of the Svensson curve. Estimates at all maturities rely on a single set of parameters, of which one is the long-run level, determined largely by the yield on the longest bond. But the increasing divergence of the two redemption yield series (the series marked 5.75% 2009 and 6% 2028 in Chart 4) suggests that the level of this asymptote is not well-defined, at least in this maturity range. As a result, the asymptote itself is likely to be unstable. Furthermore, as we saw in Chart 1, this volatility may be transmitted into estimates of the entire yield curve. Thus it appears that the cost of the Svensson curve providing a good fit to the underlying data is a loss in stability of the curve.

3.2 Stability of the curves

This idea of instability is discussed more formally by Waggoner (1996). He assesses the stability of the yield curve in the light of small price changes in the underlying bond data. The idea is that, assuming bond prices are measured with error, the curves should be stable in the sense that changes in this measurement error do not have a significant influence on our estimates. An obvious source of measurement error is the fact that prices are quoted in integer multiples of the 'tick size'.[9] This means that there will generally be an error of up to plus or minus half a tick between the observed price and that implied by the underlying term structure. As a result, a very small change in the term structure could be reflected by a change in the observed bond price of up to half a tick. We refer to an estimation method as being stable if the effect of these changes on the resulting yield curve is small.

Waggoner quantifies this notion of stability by computing what is known as a condition number. This 'measures' the sensitivity of the output of a process to a change in one of its inputs. Suppose, for example, that for a particular yield curve model (or 'process'), when the set of bond prices (the 'input') changes by 2%, the zero-coupon yield curve (the 'output') varies by 20%. We say that this process has a condition number of ten, this being the ratio of the output variation to the change in input. We are interested here in processes (we use the notation $G(\mathbf{x})$) which, from a discrete vector of bond prices ($\mathbf{x}$), compute a continuous yield or forward rate curve ($g(m;\mathbf{x})$, where m is maturity).

To formalise the notion of a change in the vector of bond prices, $\mathbf{x}$, we need to specify a norm over these prices. This effectively measures the 'size' of the vector. We use the standard Euclidean norm such that if $\mathbf{x}=(x_1,x_2,\ldots,x_N)$ is our vector of bond prices, then the norm is defined as:

$$\left\|\mathbf{x}\right\|_X = \sqrt{\sum_{i=1}^{N} x_i^2} \qquad \textbf{(7)}$$

Similarly, we need to describe the size of the process, $G(\mathbf{x})$. In this case, the standard Euclidean norm is not available, as the outputs are curves that are continuous rather than discrete. Instead, we define two different norms (from an infinite family). The first, the average norm (or L^1) is given by:

$$\left\|G(\mathbf{x})\right\|_1 = \frac{1}{M - m}\int_m^M \left|g(s;\mathbf{x})\right| \mathrm{d}s \qquad \textbf{(8)}$$

[9] On 1 November 1998 the gilt tick size was changed from £(1/32) to £0.01. Since our sample runs from 1 May 1996 to 31 December 1998, we use the larger value for this analysis.

and the second, the max (or L^∞) norm, is defined by:

$$\left\| G\left(\mathbf{x}\right)\right\|_\infty = \max_{s\in[m,M]} \left| g\left(s;\mathbf{x}\right)\right| \tag{9}$$

where [*m*, *M*] defines the maturity range of interest. Condition numbers relating to each of these norms are then defined respectively as:

$$CN_1^g\left(G,\mathrm{x}\right)=\sup_{0<\|e\|_X<g\sqrt{N}} \frac{\left\|G\left(\mathrm{x}+e\right)-G\left(\mathrm{x}\right)\right\|_1 / \left\|G\left(\mathrm{x}\right)\right\|_1}{\|e\|_X / \|\mathrm{x}\|_X} \tag{10}$$

and

$$CN_\infty^g\left(G,\mathrm{x}\right)=\sup_{0<\|e\|_X<g\sqrt{N}} \frac{\left\|G\left(\mathrm{x}+e\right)-G\left(\mathrm{x}\right)\right\|_\infty / \left\|G\left(\mathrm{x}\right)\right\|_\infty}{\|e\|_X / \|\mathrm{x}\|_X} \tag{11}$$

In equations **(10)** and **(11)**, the parameter g represents the maximum perturbation, which we set to be equal to half the tick size. The ‘sup’ operator indicates that we take the largest value of the percentage change in the yield curve over all possible perturbations (e).

Table D: Zero-coupon yield curve condition numbers

	CN_1				CN_∞			
	SV	NS	VRP	FNZ	SV	NS	VRP	FNZ
Median	5.0	2.3	1.9	1.6	68.3	16.8	38.2	10.3
90th percentile	27.8	5.4	3.2	3.1	524	55.2	83.6	28.6
95th percentile	37.7	6.9	3.6	3.9	996	79.7	95.9	35.3
Maximum	116	102	6.4	62.1	51100	27900	173	309

Table E: Forward curve condition numbers

	CN_1				CN_∞			
	SV	NS	VRP	FNZ	SV	NS	VRP	FNZ
Median	14.4	5.0	3.7	3.5	81.1	18.1	37.3	11.7
90th percentile	101	13.5	5.8	6.7	554	62.2	82.7	28.0
95th percentile	126	19.9	6.6	8.5	964	92.0	95.0	34.8
Maximum	419	170	11.4	144	58300	27800	173	496

Table D and Table E present summary statistics for the distribution of CN^{g} for each type of norm and for the two types of curve: zero-coupon yield and instantaneous forward rate. The figures are calculated as follows. On each day, the set of bond prices is perturbed by a vector of random numbers within the range $[-g, g]$. For each model and for each type of curve and norm, the sensitivity of the estimates to these perturbations are calculated. This is repeated seven times, and the condition number for that particular model is given by the maximum of these sensitivities. This is repeated on a daily basis. Thus, given that there are 676 days in the sample, the figures reported for each specification (model, curve and norm) are based on 4,732 simulations.[(10)]

To interpret the absolute levels of the condition numbers note that the average percentage change in the norm of the input bond prices was 0.0082% (with a range of 0.0060% to 0.0102%). Hence, a condition number of 120 roughly corresponds to a 1% change in the output level while a condition number of 600 corresponds approximately to a 5% change in output. Since interest rates averaged around 7% during the sample period, it follows that these numbers (120 and 600) correspond to shifts of 7 and 35 basis points respectively in the estimated term structure. Hence, we can broadly characterise condition numbers of less than 120 or so as indicating a stable process, and condition numbers of greater than about 600 as characterising an unstable process.

On this measure the VRP method is stable irrespective of the norm used to calculate the change in the estimated term structure. Moreover, it is more stable than all the other methods. We can also compute condition numbers for segments of the maturity spectrum. Table F gives results on these segments for the L^1 norm calculated for the forward curve. This confirms that the VRP method is stable across the entire maturity spectrum. All of the other methods can occasionally exhibit substantially larger distortions (as shown by the maxima of the condition number distributions) for the same input perturbation.

---

(10) Note that, to ensure comparability, identical perturbed prices are used as input for each method on any particular day.

Table F: CN_1 condition numbers (forward curve) by maturity segment

Range		SV	NS	VRP	FNZ
0 - 2 years	Median	21.5	6.6	11.3	4.9
	Maximum	775	825	49.4	168
2 - 5 years	Median	10.8	4.5	3.4	3.7
	Maximum	192	170	18.5	149
5 - 10 years	Median	9.4	3.6	2.1	2.7
	Maximum	339	115	10.1	182
10 – 20 years	Median	15.3	5.1	2.7	3.3
	Maximum	644	272	9.7	180

3.3 Summary

To summarise, we have conducted a number of tests, based on the criteria set out in Section 1. A striking feature of these tests is that no one alone is conclusive, thus demonstrating the potential trade-off which exists between the three properties we have chosen to focus on: smoothness, flexibility and stability. The NS model, for example, appears to be much more stable than the Svensson technique. But, as we saw in Section 3.1, this is at the cost of a lower goodness-of-fit. At the same time, we know that the spline-based method of FNZ is better able to capture the shape of the underlying term structure, as measured by strips, but again its out-of-sample goodness-of-fit is worse than that of Svensson.

In all cases, the VRP curve appears to perform well. It is by far the most stable approach, defined on the basis of our condition numbers, and is able to capture the shape of the underlying term structure, as measured by strips. The curves are also smooth across maturity, while providing out-of-sample goodness-of-fit results that are at least as good as each of the other competing methods. The least stable method appears to be the Svensson curve. As we have seen, one explanation for this result is that, while it is fairly flexible, the fact that it is constrained to lie flat at the long end means that it is in some way inherently unstable, at least when this constraint is not satisfied by the underlying data.

Intuitively, it is unlikely that these results will change when we include additional data at the short end, or apply similar criteria to estimates of the real yield curve from index-linked gilts. Problems highlighted with the Svensson curve, for example, appear to be related to the parametric nature of the functional form and the constraint applied at the long end, rather than to specific problems in fitting data from the conventional gilt market. For completeness, however, we compare our VRP estimates in what follows with results derived using the Svensson method.

4 Improving estimates of the short end of the yield curve

In this section, we focus on the short end of the yield curve. As mentioned in the introduction, we expect the short end of the yield curve to exhibit the greatest amount of structure – expectations of the future path of interest rates are better informed at shorter maturities, and are more likely to respond to short-term news. This can be confirmed by considering the forward curve derived, for example, from the price of short sterling futures contracts. Yet at the short end of the yield curve, there are relatively few gilts – in relation to the expected structure in the shape of the curve, the sampling frequency is too low. The question is whether or not there are alternative sources of data that we can use to supplement observations from the conventional gilt market at these maturities.

4.1 Choice of additional data

By using gilt prices to estimate the yield curve we aim to measure the risk-free (or default-free) term structure of interest rates. Thus, although there is a wide range of short-term instruments traded in the UK money market, their prices are not consistent with gilt prices as they generally include a credit risk premium. As a result, yield curves estimated using assets of mixed credit rating would be internally inconsistent. Even for Libor (where the participating banks all have high credit rating), this premium can be substantial (about 25-40 basis points for three-month Libor). Moreover, since the premium is likely to be both time-varying and a function of maturity, we cannot simply subtract a constant value to obtain the risk-free rate.

The requirement that data used for estimating the yield curve are (virtually) default-free leaves us with a choice of two possible instruments: treasury bills (T-bills) and general collateral (GC) repo rates. Treasury bills are short-term zero-coupon bonds that are issued by the government, and therefore have the same risk-free nature as gilts. The T-bill discount rate is, however, generally accepted as being unrepresentative of the underlying ‘fundamental’ rate determined by expectations. This is because commercial banks use T-bills for cash management purposes, and their prices are largely determined by the banks' liquidity requirements.

A GC repo agreement is equivalent to a secured loan, and thus the credit risk is much lower than on unsecured Libor. In addition, the repo is marked-to-market on a daily basis, thereby limiting the exposure of either party to large moves in the value of the collateral. The risk premium is further reduced since the collateral comprises gilts or similar instruments, for which there is virtually no chance that the issuer will default during the term of the repo. GC repo therefore provides us with the only widely traded, virtually riskless instrument.

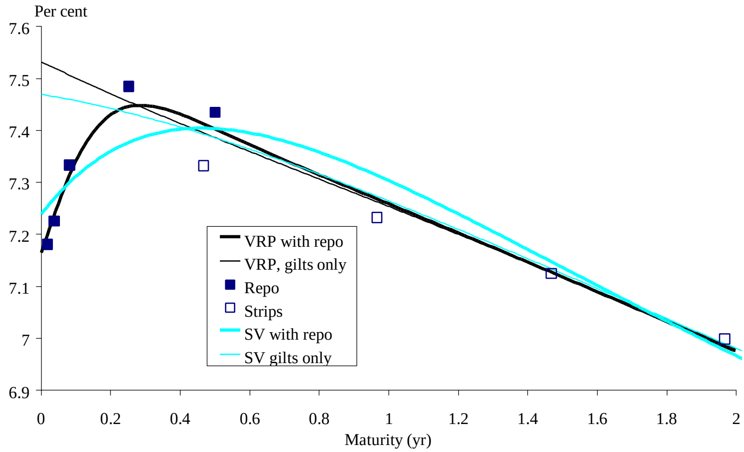
The maturities of GC repo contracts range from overnight to as long as one year. The shortest rates are quite volatile, however, and the longer rates are traded too infrequently for their prices to be considered reliable. Hence we use one-week, two-week, one-month, two-month, three-month and six-month repo rates in our estimation. Since the repo rate is determined by the difference between the sale and re-purchase prices of the collateral, there is no traded asset whose price is a direct reflection of the rate. We therefore create synthetic zero-coupon bonds whose maturity and yield match the corresponding repo rate, and for which the terminal cash flow is £100, so that repo rates are given exactly the same weight in the estimation as a short-dated gilt.

As a result of adding in the extra data, we found it necessary to place knot points at the maturity of every security included in the estimation process, rather than every third, as discussed above. The additional structure implied by the repo rates could not otherwise be captured reliably; with fewer knot points the effective stiffness of the curve was being controlled by the distribution of the knot points, and not, as required, wholly by the roughness penalty.

4.2 Results of incorporating GC repo rates

The effect of including GC repo rates can be seen most clearly by focusing on a single day. Chart 5 shows the short end of the zero-coupon yield curve on a randomly chosen day estimated using both Svensson and VRP methods, with and without the inclusion of GC repo rates. Clearly, without the GC repo rates, the initial upward slope of the yield curve was being missed by both techniques. But we also see that the VRP method is better able to capture the 'true' shape of the curve. In particular note that the Svensson curve is significantly altered by the inclusion of GC repo rates at maturities of around one year, although the longest repo rate has a time to maturity of only six months.

Chart 5: Short end of zero-coupon yield curve on 19 June 1998

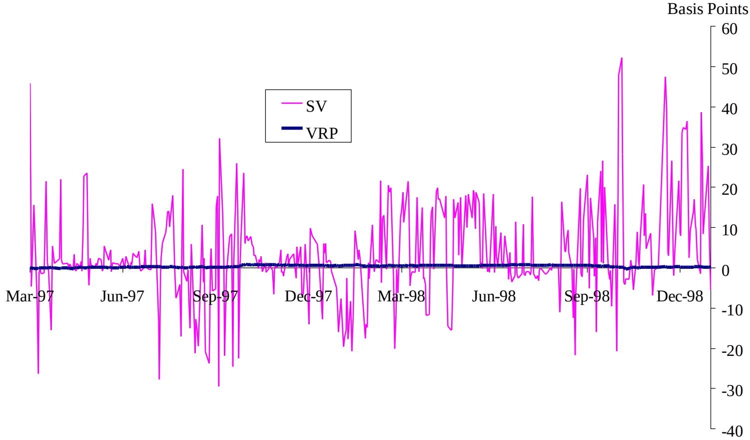


Indeed, the Svensson estimates of the forward curve are significantly altered even at maturities of 20 years or more (see Chart 6). In contrast, VRP estimates of the one-year forward rate are very similar with and without repo rates, and at two years and longer are indistinguishable. This is important as it indicates that, even if there is reason to doubt the reliability of the GC repo data, or if these are not available,[(11)] we can still have confidence in the VRP estimates at longer maturities. Once again this difference between the methods is a consequence of the relative behaviour of spline and parametric curves illustrated in Section 2.2 – estimates using parametric methods can be significantly changed across the whole curve by small alterations to the data in one small region of the curve.

---

(11) For example, the repo market effectively ceased trading for a couple of weeks around the end of 1999.

Chart 6: Differences in 20-year instantaneous forward rates, estimated using SV and VRP, with and without GC repo rates



To ensure that including GC repo rates had no detrimental effect on the performance of the VRP method, we repeated the tests described in the previous section. First, we re-estimated condition numbers, treating the repo rates as bonds to be perturbed in the same manner as gilts.(12) We found that the inclusion of repo rates does not in any way impair the performance of the method on this measure: the curve remains stable irrespective of the norm, curve type (forward or yield) or maturity segment. We then re-calculated the out-of-sample goodness-of-fit of the VRP estimates: a comparison of results, with and without the inclusion of the GC repo data, is given in Table G. This shows that the out-of-sample goodness-of-fit is actually improved substantially when repo rates are included. Intuitively, this is because the greatly reduced maturity spacing at the short end means that the ability of the fitted curve to move when a repo rate is omitted from the fit (and thus give a poor out-of-sample fit) is curtailed by the reduced distance to the neighbouring rates.

(12) Although market participants can in principle specify any value for repo rates, in practice the minimum separation between observed repo rates is 0.005%. This is equivalent to the 'tick' size for bonds. Thus we perturb each repo rate by a random amount drawn from a random distribution, centred on zero and with a half-width of 0.0025%. This perturbed rate is then converted to a perturbed price for the estimation procedure. In this way, the repo rates are treated in an identical fashion to the gilt prices.

Table G: VRP out-of-sample goodness-of-fit (£)

	Mean	Standard deviation
Gilt only	0.088	0.024
With GC repo data	0.014	0.006

5 Estimating the real and implied inflation term structures

In this section, we focus on the particular problems associated with estimating the real yield curve and term structure of inflation expectations, using the prices of both conventional and index-linked gilts (IGs). If payments from IGs were perfectly indexed, and hence were fixed in real terms, we could estimate the real term structure in much the same way as we derive the nominal yield curve. But in practice, IG cash flows are indexed to the level of the retail price index prevailing eight months previously. This means that coupon payments from IGs that occur within the next eight months are known exactly in nominal terms, but are uncertain in real terms. The price of the bond is, therefore, a complicated function of both real and nominal interest rates, and hence, in addition to a curve-fitting methodology, we need a theoretical framework to enable us to disentangle the real yield from the prices of both conventional and index-linked gilts.

Until recently, the Bank of England used a method proposed by Deacon and Derry (1994), referred to as the 'inflation term structure' or ITS model. To implement this technique, an initial assumption is made about market expectations of future inflation. Using this assumption, a real forward rate curve is then fitted to the prices of the index-linked gilts.(13) By comparing the real forward rates with a nominal interest rate curve (derived from the prices of conventional gilts using the Svensson method described above), a revised estimate of the inflation term structure can be derived,(14) which in turn is used to re-estimate the real forward curve. This process is repeated until the real forward rate curve converges.

(13) In practice, a very simple parametric form is used to fit the forward curve, namely $f(m)= b_0 + b_1 exp(-m/t)$. This is basically a truncated version of the Svensson functional form, using only the first two terms (constant and exponential decay). The idea of using a simpler functional form is to prevent overfitting given the small number of IGs.

(14) We implicitly assume that there is no inflation risk premium so that nominal forward rates equal the sum of real rates and inflation expectations.

The aim in this section is to describe a new approach to estimating the real yield curve and inflation term structure. We have already seen the advantages offered by a spline-based technique over parametric approaches. Recent work by Evans (1998) has further provided us with a more elegant and transparent framework for dealing with the indexation lag. The method described below incorporates both of these advances, together with some modifications to Evans' approach necessary to allow us to estimate the real and inflation term structures on a daily basis.

5.1 Evans' theoretical framework

Evans' main innovation was to make the relationship between IG prices, real yields and nominal interest rates explicit by introducing the notion of an index-linked discount function. To understand this concept, we first need to develop some notation. In particular, let M_t denote the nominal stochastic discount factor or pricing kernel. This is defined so that the current value, V_t, of a future nominal cash flow, V_{t+h}, payable at time $t+h$, is given by:

$$V_t = \mathrm{E}_t\left[M_{t+h}V_{t+h}\right] \qquad \textbf{(12)}$$

Now let $Q_t(h)$ denote the time t price of a nominal zero-coupon bond, paying a certain £1 at time $t+h$. According to equation **(12)**, this is given by:

$$Q_t(h) = \mathrm{E}_t\left[M_{t+h}.1\right] \qquad \textbf{(13)}$$

Similarly the nominal price of a real (perfectly indexed) bond which pays £(P_{t+h}/P_t) at time $t+h$, where P_t is the general price level, is given by:

$$Q_t^*(h) = \mathrm{E}_t\left[M_{t+h}\frac{P_{t+h}}{P_t}\right] \qquad \textbf{(14)}$$

But, as discussed above, IGs are not perfectly indexed – cash flows are linked to the price level with a lag of length l. Hence, we also define the price of a zero-coupon index-linked bond as:

$$Q_t^+(h) = \mathrm{E}_t\left[M_{t+h}\frac{P_{t+h-l}}{P_t}\right] \qquad \textbf{(15)}$$

Given these definitions, it is reasonably straightforward to show that there is a unique relationship between the yields on the three types of bond: nominal, real and index-linked (see Evans (1998) for further details). Denoting these by y, y^* and y^+ respectively:

$$y_t^*(\tau) = \frac{h}{\tau} y_t^+(h) - \frac{1}{\tau}\left[h y_t(h) - \tau y_t(\tau)\right] + \frac{\gamma_t(\tau)}{\tau}, \qquad \textbf{(16)}$$

where $\tau = h\text{-}l$, and $\gamma(\tau)$ is a risk premium term, which we assume in what follows to be equal to zero. This assumption is supported by Evans who found that the risk premium term contributed only approximately 1.5 basis points to the annualised real yield.

All that remains to do is to describe the relationship between the theoretical index-linked zero-coupon bonds and the prices of IGs that we observe in practice. So let $Q^{c+}(H)$ denote the time t price of an IG, maturing at time H, and paying an annual (real) coupon rate of c. Assuming that its price is equal to the present discounted value of its future cash flows, we can write:

$$Q_t^{c+}(H) = \frac{c}{2}\sum_{h=0}^{l} \Im_t(h)\frac{P_{t+h-l}}{P_i} Q_t(h) + \frac{c}{2}\frac{P_t}{P_i}\sum_{h=l}^{H} \Im_t(h) Q_t^+(h) + \frac{P_t}{P_i} Q_t^+(H) \qquad \textbf{(17)}$$

Here $\Im_t(h)$ is an indicator function, which is equal to one if a cash flow occurs at time h, and zero otherwise. Coupon payments are indexed from a base price level, P_i, which is set when the bond is issued. Note that the nominal values of cash flows occurring within the next eight months (the indexation lag) are known, and are therefore discounted at the nominal rate.

5.2 Two modifications

In his paper, Evans estimated the zero-coupon nominal and index-linked yield curves from the prices of conventional and index-linked gilts respectively, and then used equation **(16)** to back out the real yield curve. But by implementing this approach he implicitly assumed that the current price level is known at all times while, in practice, this is never the case.[15]

---

[15] Evans did some tests using a 'certainty equivalent' value of the price level, which he estimated as an additional parameter. He found, using end-of-month data, that this was an unnecessary refinement. We wish to estimate the curves on a daily basis, however, which means we have to take the publication schedule of the RPI into account.

The retail prices index (RPI), to which IGs are indexed, is published monthly, and the data indicate the level of the index on the last day of the month. This creates two problems:

- Since the data are published with an approximate two-week delay, the current price level, P_t, is never known with certainty at any time t.

- Cash flows are indexed to the level of the RPI prevailing eight months previously. But months are not equal in length; moreover, whether a cash flow is due on 1 November or 30 November, say, it will be uplifted with reference to the same RPI value (that prevailing the previous March). Thus the indexation lag, l, does not have a constant value.

Both of these problems are best illustrated with actual examples. In the first case, suppose that t corresponds to 29 February 2000. The price level P_t is not known; the latest RPI value available corresponds to 31 January 2000. Moreover, this remains the latest value until mid-March 2000. Hence, if we let $t1$ denote the date of the latest RPI value (in this case 31 January 2000), then there can be up to around six weeks between t and $t1$. To account for this, we re-define the index-linked discount factor with respect to the latest known RPI value, such that:

$$\hat{Q}_t^+(h) = \mathrm{E}_t\left[M_{t+h}\frac{P_{t+h-l}}{P_{t1}}\right] \qquad \textbf{(18)}$$

Denoting the yield on this bond by $\hat{y}_t^+(h)$, equation **(16)** then becomes:

$$y_t^*(t) = \frac{h}{t}\hat{y}_t^+(h) - \frac{1}{t}\left[hy_t(h) - ty_t(t)\right] + \frac{\ln(P_t) - \ln(P_{t1}) + g(t)}{t} \qquad \textbf{(19)}$$

In this way we have made explicit the fact that we need to know P_t, and although we still have to make an assumption about its value, we can now see exactly how that assumption impacts on our estimated yields. One simple approach is to base the current price level, P_t, on the latest available value, P_{t1}, increasing it by the current annual rate of inflation, such that:

$$P_t = \left[1 + (t - t1)\Pi_{t1}\right]P_{t1} \quad \text{where} \quad \Pi_{t1} = \frac{P_{t1} - P_{t1-1year}}{P_{t1-1year}} \qquad \textbf{(20)}$$

The second problem is more subtle. In deriving the relationships above, we have denoted the date on which a cash flow occurs as $t+h$, and the date to which the cash flow is indexed as $t+t$, so that the indexation lag is $l=h-t$. Now consider a specific example. Suppose today is 24 August 1999, and also that there is a coupon payment due from a specific IG on 15 November 2001. This implies that h equals 814 days. The cash flow will be uplifted with reference to the March 2001 RPI figure, giving a t of 585 days and therefore an indexation lag, l, of 229 days. If we now consider the subsequent cash flow of the IG, due six months later on 15 March 2002, we find that its indexation lag is 227 days. And in general, each cash flow of every IG will have a unique indexation lag.

As a direct consequence, we are unable to fit an index-linked curve and then back out a real curve in the same way as Evans. Instead, we fit a VRP spline directly to the real curve, and then use equation **(19)**, together with a separately estimated nominal curve, to calculate the index-linked discount factors applicable to each individual cash flow. A fitted price for each IG can then be calculated using a modified version of equation **(17)** which takes into account the new definition of the index-linked discount factor:

$$Q_t^{c+}(H) = \frac{c}{2}\sum_{h=0}^{l} \Im_t(h)\frac{P_{t+h-l}}{P_i}Q_t(h) + \frac{c}{2}\frac{P_{t1}}{P_i}\sum_{h=l}^{H}\Im_t(h)\hat{Q}_t^{+}(h) + \frac{P_{t1}}{P_i}\hat{Q}_t^{+}(H) \qquad \textbf{(21)}$$

This, however, raises a further issue. Suppose today's (settlement) date is 1 September 1999, which means that (because of the publication delay) the latest RPI value corresponds to July 1999. Now consider a cash flow from an IG which is due on 24 April 2000, and which will be indexed to the August 1999 value of the RPI. This cash flow is not known in nominal terms, and thus we need to calculate its index-linked discount factor. This cannot be done, however, because the corresponding value of t is equal to -1, and hence $Q^+(t)$ is undefined.

To resolve this problem, note that if $t<0$ then, in the same way that we assume we know the market's assumption for P_t, we can equally well consider P_{t+t} to be known, and given by:

$$P_{t+t} = \left[1 + \left(t + t - t1\right)\Pi_{t1}\right]P_{t1} \qquad \textbf{(22)}$$

Hence, we can modify our definition of the index-linked zero-coupon bond as:

$$\hat{Q}_t^+(h) = Q_t(h)\frac{P_{t+t}}{P_t} \quad \text{if} \quad t < 0 \qquad \textbf{(23)}$$

In this way, we are able to calculate the present value of any index-linked cash flow, and thus fitted prices for all IGs.

5.3 Stability of the estimated real curve

In Section 3.2 we saw that the VRP method offers significant advantages over parametric techniques in terms of the stability of nominal yield curve estimates. The stability of the real curve is more complex as it will be affected by movements in both conventional and index-linked gilts. Moreover, since in equation **(19)** t appears in the denominator of the right-hand side, the estimates of the real curve will become increasingly sensitive to curve estimation errors at shorter maturities. In this section, we examine this sensitivity in purely theoretical terms using the condition number framework described previously.

The aim is to determine the shortest maturity for which the real yield curve is stable with respect to small changes in the underlying index-linked and nominal term structures. A small change in the estimated nominal yield, $\Delta y_t(x)$, can be written to a first-order approximation as:

$$\Delta y_t(x) \approx \frac{-\Delta Q_t(x)}{xQ_t(x)} \qquad \textbf{(24)}$$

Similar relationships hold for the index-linked and real yields. Now consider a small perturbation to the price of index-linked gilts, such that $\hat{y}_t^+(h) \to \hat{y}_t^+(h) + \Delta\hat{y}_t^+(h)$. Keeping nominal yields fixed, it follows from equation **(19)** that there is a change in the real yield curve of:

$$\Delta y_t^*(t) = \frac{-\Delta \hat{Q}_t^+(h)}{t\hat{Q}_t^+(h)} \qquad \textbf{(25)}$$

The effect of a small perturbation to the nominal curve, meanwhile, will depend on whether this occurs at a maturity of h or t, in which case the effect on the real yield curve will be given by respectively:

$$\Delta y_t^*(t) = \frac{\Delta Q_t(h)}{tQ_t(h)} \quad \text{or} \quad \Delta y_t^*(t) = \frac{-\Delta Q_t(t)}{tQ_t(t)} \qquad \textbf{(26)}$$

Realistically, both index-linked and conventional gilts will be subject to errors at the same time. We can summarise the effect this has on the estimated real yield by defining a condition number as previously. In particular, using the L^1 norm, we define this condition number as:[16]

$$CN = \frac{\frac{1}{y_t^*(t)}\left[\frac{\Delta \hat{Q}_t^+(h)}{t\,\hat{Q}_t^+(h)} + \frac{\Delta Q_t(h)}{t\,Q_t(h)} + \frac{\Delta Q_t(t)}{t\,Q_t(t)}\right]}{\left[\frac{\Delta \hat{Q}_t^+(h)}{\hat{Q}_t^+(h)} + \frac{\Delta Q_t(h)}{Q_t(h)} + \frac{\Delta Q_t(t)}{Q_t(t)}\right]} = \frac{1}{t\,y_t^*(t)} \qquad \textbf{(27)}$$

This describes the sum of percentage changes to the real curve arising from three possible changes in its underlying components: the index-linked yield at maturity h, and nominal yields at maturities of h and t.

(16) Note that this is equation **(10)** with $G(\boldsymbol{x}+\boldsymbol{e})$-$G(\boldsymbol{x})$ representing changes to the real yield curve arising from perturbations to the index-linked curve (as in equation **(25)**) or in the nominal curve, with the associated changes described in **(26)**.

To assess whether the technique is stable we need to determine a threshold, or critical value, for the condition number, CN_{crit}, below which the curve is felt to be stable. Recall that the tick size used in the gilt market (both conventional and index-linked) is £(1/32), for a notional redemption payment of £100.(17) In the spirit of the condition number analysis we performed earlier, we require the method to be stable to perturbations of less than or equal to half of the tick size, since bond prices inevitably have errors of this magnitude. Suppose then that we want to restrict the response of the real yield curve to perturbations of this magnitude to 1% (eg from 3% to 3.03%). Focusing on the short end (where prices of both nominal and index-linked zero-coupon bond prices are approximately unity) this implies a critical condition number of:(18)

$$CN_{\text{crit}} \approx \frac{1\times 10^{-2}}{\frac{1}{64}\times 10^{-2}} \approx 64 \qquad \textbf{(28)}$$

Similarly, if we require the curve to shift by only 0.5%, then we have the more stringent case, $CN_{\text{crit}}=32$. For the method to be considered stable we then require $CN<CN_{\text{crit}}$, which from equation **(27)** implies:

$$t > \frac{1}{CN_{\text{crit}}\, y_t^*(t)} \qquad \textbf{(29)}$$

For illustration, if we assume that the real yield curve is a constant 3%, and take $CN_{\text{crit}}=64$, then we conclude that the method is only stable for t greater than approximately half a year.

In practice, we assure stability at the short end of the curve by omitting IGs from the estimation with less than 16 months to maturity, corresponding to a t of around 8 months. In fact, as Chart 7 shows for the period May 1996

(17) On 1 November 1998, gilt prices changed from being quoted in multiples of £(1/32) to £0.01 ticks. This reduced the pricing error, and hence the sensitivity of the curve to the non-continuous price distribution. In the analysis presented here, this reduces the minimum value of t for which the curve is stable, and thus we could extend the estimated curve to shorter maturities. We continue, however, to truncate the curve at the higher value to provide a continuous time series.

(18) Equation **(28)** is derived by definition of the condition number – equations **(10)** and **(11)** – replacing the change in the yield curve with 1×10^{-2} and a change in one of the underlying bond price components of $(1/64)\times 10^{-2}$, where this is the proportion of half the tick size attributed to a bond with face value of £1.

to February 2000, this constraint is rarely binding. Over the sample period, the only time that there was an index-linked gilt with maturity greater than 8 months (so that at least one cash flow is unknown in nominal terms) but less than 16 months was between January 1997 and August 1997.

Chart 7: Maturity spectrum of index-linked gilts

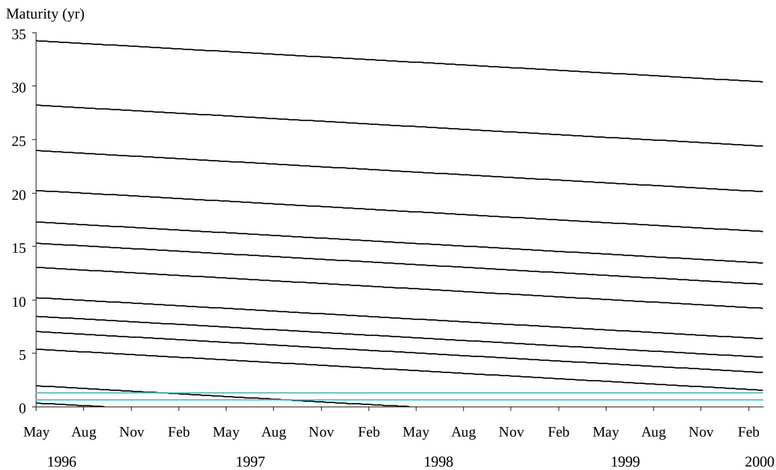


Chart 7 also has strong implications for how we interpret our real yield curve results. Estimates at maturities less than the shortest available gilt must be treated with caution, as the curve is not well-defined in such a region. Unfortunately there are no alternative index-linked assets which we could incorporate into the estimation procedure as we did for the nominal curve. In the following discussion we consider only the portion of the real curve for which there is data.

5.4 Comparison of new approach with the ITS technique

In this section, we compare our new estimates with those using the ITS technique, across the criteria used previously for the nominal yield curve. We find that estimates of the real yield curve using the two methods track each other quite closely, with the largest differences at the long end where the ITS curve is constrained to lie flat. Inflation forwards differ more markedly, reflecting differences in the nominal curve estimates. For example, the fact that the Svensson model is unable to capture any downward slope in the nominal yield curve means that the 20-year ITS inflation forward estimates always lie above the rates estimated using the VRP approach. It is also worth noting that the volatility of the ITS inflation forward rates is considerably greater than that of the VRP estimates (see Table H). Since the contrast is not so striking when we consider real forward rates, we conclude that much of this instability again arises from the nominal estimation method.

Table H: Volatility of real and inflation forward rates[(a)]

	Volatility of instantaneous forward rates (%)		
	3-year	10-year	20-year
VRP real	3.2	2.6	2.1
ITS real	3.4	2.5	2.5
VRP inflation	5.8	5.0	5.8
ITS inflation	6.3	9.9	16.6

(a) Measured as the standard deviation of daily changes in the level.

As before, we measure the stability of the estimates more formally by calculating the condition numbers associated with small perturbations in the underlying price. But note that we only perturb the prices of the index-linked bonds, keeping the respective nominal curves fixed. This is because in these tests we wish to isolate the relative performance of a spline or parametric function fitted to the real forward rate. Nonetheless, in practice measurement errors clearly affect both conventional and index-linked bonds. And since we know the ITS nominal curve (Svensson) is less stable than the VRP nominal curve, then if we were to perturb all bond prices, we expect the ITS condition numbers to increase more than the VRP numbers.

To interpret the condition numbers in absolute terms, note that the average fractional change in the bond prices was 0.0050%; hence a condition number of 200 corresponds to a 1% change in the level of the real rate (for example from 3.00% to 3.03%). This is higher than the critical threshold of 120 used to assess the nominal curves since the average price of an IG is higher than that of a conventional bond, but the tick size is the same. Similarly, a condition number of 1,000 corresponds to a 5% shift in the output curve. Loosely, we can say that a condition number of below 200 implies that the curve is stable.

Table I: Forward curve condition numbers

	Average norm		Maximum norm	
	ITS	VRP	ITS	VRP
Median	6.5	7.6	55.7	36.8
90th percentile	21.0	13.4	732	77.3
95th percentile	53.2	15.5	3320	89.0
Maximum	733	25.2	525000	141

Table I shows summary statistics for the distribution of condition numbers calculated for the forward curve using the two methods across all maturities. This shows clearly that the VRP method outperforms the ITS estimates. Moreover, estimates derived using the VRP method are always stable using the calculation guide outlined above. A similar story is found comparing the goodness-of-fit of the two sets of estimates, in Table J below. This shows that the VRP method not only fits the data better than the ITS technique, but also the quality of the fit is less variable from day to day.

Table J: Out-of-sample goodness-of-fit statistics

	Mean	**Std dev**
ITS	0.78	0.71
VRP	0.55	0.54

6 Conclusions

We have presented a new method for estimating the term structure of interest rates based on cubic smoothing splines. The important innovation of this model, initially suggested by Waggoner, is that the degree of smoothing applied to the curve is a function of maturity. Specifically, we allow the curve to be more flexible at the short end, where we expect the yield curve to exhibit the greatest amount of curvature, than at the long end, where expectations are likely to be constant.

In a detailed comparison with other widely used techniques, we find that the new VRP method outperforms the commonly used Svensson, Nelson and Siegel, and Fisher, Nychka and Zervos models on all our criteria, which are based on the suitability of the resulting yield curves for monetary policy purposes. By including GC repo rates (which are effectively riskless), we can significantly improve the VRP estimates of the short end of the curve. We cannot use the same procedure to improve Svensson estimates, since the parametric nature of this technique means that rates at the very long end are also affected. We have also combined the VRP technique with an enhanced version of the framework introduced by Evans (1998) to allow us to estimate real and inflation term structures from the price of index-linked bonds. Again the new method significantly outperforms the ITS technique used previously.

Appendix: Smoothing cubic spline models

A generic spline is a piecewise polynomial, ie a curve constructed from individual polynomial segments, joined at 'knot points'. By construction, both the curve and its first derivative are continuous at all points. This places constraints on the coefficients of the individual polynomial segments. Concentrating on cubic functions (which result in a cubic spline), the continuity constraints mean that every cubic spline can be written in the form,

$$S(x) = ax^3 + bx^2 + gx + d + \sum_{i=1}^{N-1} h_i \left| x - k_i \right|^3$$

for some constants, a, $ß$, g d, h_i, where k_i, i=[0,N] is the set of knot points.

Although this is the simplest expression for a cubic spline, it is numerically unstable,[19] and so instead we prefer to represent our splines as a linear combination of cubic B-splines. This is a completely general transformation – any spline can be written as such a combination of B-splines of the appropriate order – which cures the numerical problems. B-splines of order n are most simply represented by the following recurrence relation:

$$B_{i,n}(x) = \frac{x - k_i}{k_{i+n-1} - k_i} B_{i,n-1}(x) + \frac{k_{i+n} - x}{k_{i+n} - k_{i+1}} B_{i+1,n-1}(x)$$

with $B_{I,1}(x) = 1$ if $k_i \le x < k_{i+1}$ and zero otherwise. For more details see Lancaster and Šalkauskas (1986).

[19] Computers have only finite accuracy, and the calculation of a spline using this expression typically involve the subtraction of very large, similar numbers, resulting in (potentially) large errors.

References

Anderson, N, Breedon, F, Deacon, M, Derry, A and Murphy, G (1994), 'Estimating and interpreting the yield curve', John Wiley & Sons.

Bliss, R R (1997), 'Testing term structure estimation methods', in Boyle, P, Pennacchi, G and Ritchken, P (eds), *Advances in Futures and Options Research*, Vol. 9, JA1 Press, pages 197-231.

Brown, R H and Schaefer, S M (2000), 'Why long term forward interest rates (almost) always slope downwards', unpublished manuscript.

Davison, A C and Hinkley, D V (1997), 'Bootstrap methods and their applications', Cambridge University Press, Cambridge, pages 292-93.

Deacon, M and Derry, A (1994), 'Deriving estimates of inflation expectations from the prices of UK government bonds', *Bank of England Working Paper no. 23*.

Evans, M D D (1998), 'Real rates, expected inflation and inflation risk premia', *Journal of Finance*, Vol. 53, pages 187-218.

Fisher, M, Nychka, D and Zervos, D (1995), 'Fitting the term structure of interest rates with smoothing splines', *Finance and Economics Discussion Series*, 95-1, Federal Reserve Board.

Lancaster, P and Šalkauskas, K (1986), 'Curve and surface fitting: an introduction', Academic Press, London.

McCulloch, J H (1971), 'Measuring the terms structure of interest rates', *Journal of Business*, Vol. 44, pages 19-31.

McCulloch, J H (1975), 'The tax-adjusted yield curve', *Journal of Finance*, Vol. 30, pages 811-30.

Nelson, C R and Siegel, A F (1987), 'Parsimonious modelling of yield curves', *Journal of Business*, Vol. 60, pages 473-89.

Svensson, L (1994), 'Estimating and interpreting forward interest rates: Sweden 1992-94', *IMF working paper series*, No 114.

Svensson, L (1995), 'Estimating forward interest rates with the extended Nelson & Siegel method', *Sveriges Riksbank Quarterly Review*, 1995:3, pages 13-26.

Waggoner, D (1996), 'The robustness of yield and forward rate curves extraction methods', unpublished manuscript.

Waggoner, D (1997), 'Spline methods for extracting interest rate curves from coupon bond prices', *Federal Reserve Bank of Atlanta, Working Paper series*, 97-10.