The Ex Ante Predictive Accuracy of Alternative Models of the Term Structure of Interest Rates

Authors: Basma Bekdache, Christopher Baum

Boston College Working Papers in Economics, 1997

The Ex Ante Predictive Accuracy of Alternative Models of the Term Structure of Interest Rates

Basma Bekdache[*]
Wayne State University

Christopher F. Baum
Boston College

August 19, 1997

Abstract

This paper compares six term structure estimation methods in terms of actual ex ante price and yield prediction accuracy. Specifically, we examine the models' ability to price Treasuries for one to five trading days ahead. The models' performance differs markedly between in- and out-of-sample predictions. Their relative success also depends on time, the forecast horizon, and whether price or yield errors are compared. We examine the degree of loss in accuracy the modeler incurs by not using the best method: in particular, we compare the more complex splining methods and the parsimonious Nelson-Siegel model.

Keywords: term structure, ex ante forecasts, spline models, Nelson-Siegel approach

JEL codes: E43, E47

[*]We acknowledge suggestions offered by Christian Gilles, Robert Bliss, and participants in the Third Annual Conference of the Society for Computational Economics, June 1997. The standard disclaimer applies.

Please address correspondence to Bekdache, FAB 2101, Dept of Economics, Wayne State University, Detroit, MI 48202; phone: 313.577.3231, fax: 313.577.0149, email: bbekdac@econ.wayne.edu

1. Introduction

Modeling the term structure of interest rates has been the subject of a vast literature in economics and finance. Characterizing interest rates as a function of maturity embodies information about future movements in interest rates and has direct implications for real economic activity. In financial markets, the term structure is crucial to the pricing of interest-rate-contingent claims and fixed-income derivative securities. Although many analytical interest rate models imply particular shapes for the term structure, much of the empirical literature has been concerned with fitting model parameters to the data in order to faithfully reflect the interrelations among observed prices and yields. A wide gulf exists between the specific predictions of theory and the complex empirically derived forms of many term structure estimation models.

In this paper, rather than focusing on the adequacy of a specific empirical approach, we consider the relative performance of a number of competing models. Our contribution to this literature is our emphasis on ex ante predictive accuracy. Since the empirical methods commonly applied in this literature differ widely in their degree of parsimony, ex post comparisons are often flawed. Other researchers have attempted to address this issue by estimating from a subsample and predicting over the remainder of the sample, but this out-of-sample approach tests a methodology that would hardly be attractive to practitioners.[1] Since one of the major goals of term structure estimation is the prediction of near-term interest rate levels, changes, and spreads, we consider that the evaluation of short-term ex ante forecasts, or actual out-of-sample predictions, is a very realistic approach for the comparison of competing methodologies. This consideration is tempered by the fact that the observed term structure inevitably

[1] The split-sample approach may be employed for a purely econometric rationale, such as determining whether a method is overfitting the in-sample data.

shifts during the forecast period, confounding identification of prediction errors; but in a very liquid market, the shifts are in the nature of common factors, affecting all similar bonds in a common fashion. Our approach to these issues focuses more sharply on how we might reasonably compare these competing methodologies, using measures of ex ante predictive accuracy, than on the specific scores attained by each model, or the uncertainty generated by the forecast horizon.

The purpose of this study is to apply ex ante measures of predictive accuracy to six term structure estimation methods and evaluate their relative performance over quarter-end dates for a six-year period. These methods can be considered partial equilibrium, or "no-arbitrage" approaches to the term structure, in that the no-arbitrage condition is used to empirically determine the shape of the term structure from price and yield data on individual securities. Five of the six methods utilize cubic splines to approximate the term structure, differing in the choice of function which is splined and the degree to which the parameterization is determined by the data. The sixth method is the well-known Nelson-Siegel approach, in which the yield curve is derived from a second-order differential equation applied to Treasury coupon STRIPS data. The spline methods are applied to essentially all of the Treasury (bill, note, and bond) issues, so that their performance approximates that achievable by a practitioner with access to the entire Treasury curve. We analyze the models' performance on the quote date and for each of the following five trading days, and generate an array of parametric and non-parametric summary measures on the models'p erformance by date, by tenor of security, and by length of forecasting horizon.

Our main findings are as follows. The out-of-sample relative performance of methods differs, in some cases radically, from that of the in-sample rankings, and those models which excel at price forecasts are not necessarily those which generate the best yield forecasts. Several methods show marked differences in their ability to generate accurate forecasts when the horizon of the forecast is varied from one to five trading days. There is considerable time variation in the relative performance of the models considered; although one of the spline models stands out, it dominates the other models in fewer than half of the cases, and is often dominated by the parsimonious Nelson-Siegel approach. The correlations between models' forecasts and movements in the yield curve are low at long horizons; this suggests that the models' relative performance is not strongly affected by shifts in the yield curve. A general statement on the superiority of any of these competing models would not be justified as their relative performance is quite sensitive to the level of rates, to the forecast horizon, and to the tenor of the underlying securities.

The plan of the paper is as follows. Section 2 surveys the literature. In subsection 2.1 some basic notation and terminology are introduced followed by a brief description of the various no-arbitrage methodologies in subsection 2.2. Section 3 presents the data and estimation methods to be compared. The in- and out-of-sample evaluation of the methods, over the full sample and tenor subsamples, is presented in section 4. Finally, we summarize the results and make concluding remarks in section 5.

2. Survey of the Literature

2.1. Notation and Basic Framework

Let $P_d(\tau)$ denote the present value at time t of \$1 repayable in $\tau = T - t$ periods. Alternatively, $P_d(\tau)$ is the time t market price of a discount bond with τ periods to maturity and whose principal is \$1, commonly known as the discount function. Further, define $r_d(\tau)$ by

$$P_d(\tau)\exp(\tau r_d(\tau)) = 1. \tag{2.1}$$

Alternatively, $r_d(\tau) = -\log(P_d(\tau))/\tau$ is the rate of growth of $P_d(\tau)$, or the yield to maturity. The term structure of interest rates is the function relating the τ-period spot rate, $r_d(\tau)$, to τ. In addition to the discount and spot rate functions, the term structure can also be represented with forward rates, which are essentially implicit future spot rates. The forward rate for the time interval t' to T observed at t is given by[2]:

$$f(t,t',T) = -\log(P_d(t,T)/P_d(t,t'))/(T-t'). \tag{2.3}$$

Where $P_d(t,t')$ denotes the time t price of a discount bond maturing at time t' and $t < t' < T$. In the absence of arbitrage opportunities, the value of the ith coupon bond at time t, maturing in $\tau = T - t$ periods, equals the sum of the present values of its stream of cash flows,

$$P_i(\tau) = \sum_{j=1}^{n} C_i(t+j)P_d(t+j) + P_d(\tau), \tag{2.4}$$

where $P_i(\tau)$ denotes the theoretical price of the bond and $C_i(t+j)$ denotes the bond's promised cash flow at time $t+j$. $n = 2\tau$ is the number of semiannual coupon payments over the life of the bond.

In general equilibrium models of the term structure, the form of the discount function $P_d(\tau)$–and hence the term structure–is derived theoretically[3]. Alterna-

[2]Assuming that the discount function is differentiable, the instantaneous forward rate is computed as:

$$f(\tau) = -d\log(P_d(\tau))/d\tau. \tag{2.2}$$

[3]Cox, Ingersoll and Ross (CIR hereafter) (1985a, 1985b) developed the first general equilibrium model of the term structure of interest rates. Prior attempts to model the term structure in a partial equilibrium setting include Vasicek (1977) and Brennan and Schwartz (1979).

tively, the no-arbitrage approaches generally fit observed bond prices to the pricing constraint in (2.4) to estimate the parameters of a given function (the discount, forward rate or the spot rate function) that relates interest rates to term-to-maturity. We now turn attention to a more detailed description of the various no-arbitrage approaches to term structure modeling.

2.2. No-Arbitrage Methodologies

The main focus of this second branch of the literature is to use the no-arbitrage condition in (2.4) to find a functional form that can best approximate the term structure. McCulloch (1971, 1975) introduced the practice of using approximating functions such as polynomials and splines to empirically approximate the discount function $P_d(\tau)$.[4] Since these papers, various approximating functions have been proposed to fit either the discount function or one of its variants. In this section we discuss only those methods which we implement in this study.

2.2.1. McCulloch's Splining Method

Recall the bond price equation from (2.4)

$$P_i(\tau) = \sum_{j=1}^{n} C_i(t+j)P_d(t+j) + P_d(\tau) + \epsilon_i \quad . \tag{2.5}$$

In (2.5), the actual price of the bond deviates from its theoretical price by the error term ϵ_i. This deviation reflects not only the statistical error due to approximation but also other factors such as transactions cost, measurement error and mispricing.

[4]Methods which were used prior to McCulloch's papers include Durand's hand fitting yield curve technique and the point-to-point or "bootstrap" method (e.g. Carleton and Cooper (1976)). Since none of these techniques are directly relevant to the current literature, we concentrate on studies following McCulloch (1971, 1975).

In McCulloch's methodology, the discount function, $P_d(\tau)$, takes the form of a cubic spline. This means that if we divide the maturity spectrum into a number of intervals, $P_d(\tau)$ is a different third-order polynomial over each interval. These piece-wise polynomials are joined at knots or "break" points so that the spline's first and second derivatives are set equal at those points. A parameterized representation of this approximation is given by [5]

$$P_d(\tau) = b_0 + b_1\tau + b_2\tau^2 + b_3\tau^3 + b_4 z_1^3 + b_5 z_2^3 + b_6 z_3^3 + \cdots \quad , \tag{2.6}$$

$$\text{where} \quad z_j = \begin{cases} \tau - k_j & : \quad \tau > k_j \\ 0 & : \quad \tau \leq k_j \end{cases} \quad ,$$

$$k_j = \text{the knot points of the spline,}$$

$$b_j = \text{the parameters of the estimated discount function.}$$

Substitute (2.6) into (2.5), then the price of the ith bond is

$$P_i(\tau) = b_0 X_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 K_1 + b_5 K_2 + b_6 K_3 + \cdots + \epsilon_i \quad . \tag{2.7}$$

In (2.7), the X_j are functions of the cash flows $C_i(t+j)$ and of term to maturity τ. The K_i are functions of the knots of the spline and the $X's$. The parameter b_0 is set equal to unity since it gives the present value of $1 to be received immediately. In McCulloch's paper, the remaining parameters of the discount function are estimated by minimizing a weighted sum of the squared price errors of (2.7), i.e. via weighted least squares. The residuals are weighted by the inverse of $v_i = (P_i^a - P_i^b)/2 + b$, where P_i^a, P_i^b and b denote the ask price of the bond, the

[5] This discussion of McCulloch's cubic spline is based on Baum and Thies (1992).

bid price and brokerage fees, respectively. This adjustment prevents the estimates from being affected by large errors that are caused solely by transactions costs. McCulloch chooses the knots k_j such that there is an equal number of bonds in each interval. Estimating the discount function as a cubic spline allows for greater flexibility in approximating complex shapes since the parameters of the curve in a given interval are heavily influenced by observations in that interval. However, this flexibility can sometimes result in curve shapes that seem unreasonable, such as the negative forward rates in McCulloch's paper. Shea (1984) discusses placing constraints on the spline and varying the number and placement of break points as potential solutions to these problems.

2.2.2. Fisher et al.: Smoothing Splines

The methodology developed by Fisher et al. (1995, 1996) implements some of these suggestions. Their study introduces two main innovations to the McCulloch framework. First, smoothing splines rather than regression splines are used to approximate the functional form chosen to represent the term structure. Second, they propose placing the spline directly on the forward rate function rather than on the discount function. In smoothing splines, the criterion used to estimate the spline parameters requires a compromise between closeness of fit, as measured by the least squares criterion, and the smoothness of the resulting function. This implies that the number and location of knots is chosen optimally rather than predetermined by the modeler. Assuming, as in McCulloch, that the discount function $P_d(\tau)$ takes the form of a cubic spline, this is achieved by minimizing the criterion function[6]

[6]In Fisher et al.'s study, the discount function is parameterized using a Basis Spline (B-spline) rather than using independent parameters as shown in (2.6). The properties of B-splines (see de Boor (1978) for details) provide for the efficient and numerically stable calculation of spline functions. A cubic spline which is formed as a linear combination of B-splines can be easily evaluated and manipulated.

$$\sum_{i=1}^{N} \epsilon_i^2 + \lambda \int_0^T P_d''(\tau) d\tau \quad , \tag{2.8}$$

where N is the number of bonds in the sample. The second term in (2.8) measures the non-smoothness in the approximating function $P_d(\tau)$ as it will be larger the greater the variation ("wiggliness") in $P_d(\tau)$. Note that if $\lambda = 0$, a knot will be placed at each observation, as this achieves a perfect fit. Thus in choosing the optimal number of knots, the fit as measured by the least squares term is balanced against smoothness according to λ, the weight given to the non-smoothness term. In this method, λ is chosen to minimize the Generalized Cross Validation (GCV) criterion

$$GCV(\lambda) = \frac{1}{N} \frac{\sum_{i=1}^{N} \epsilon_i^2(\lambda)}{(1 - \frac{n}{N})^2} \quad , \tag{2.9}$$

where n is the number of parameters of the spline. Craven and Wahba (1979) developed the GCV approach as an approximation to the Cross-Validation (CV) score $(1/N) \sum_{i=1}^{N} \epsilon_i^{(-i)2}$ with $\epsilon_i^{(-i)}$ equal to the residual for the ith observation obtained from estimates of the model with the ith observation omitted from the sample. The process of omitting an observation and predicting it using the remaining observations is repeated for all data points to generate the average residuals of the omitted observations given by the CV score. Choosing the optimal smoothness parameter λ by minimizing (2.9) and essentially simulating true out-of-sample prediction errors amounts to limiting the influence of each data point in determining the variability in the fitted curve.

Fisher et al. (1995, 1996) apply smoothing splines to the discount function, $P_d(\tau)$, the log of $P_d(\tau)$, and the forward rate function, $f(\tau)$. They compare the pricing accuracy of the fixed-knot specifications of the three functional forms to their GCV counterparts. In addition, they use Monte Carlo simulations to evaluate the ability of each method to produce stable and reasonable estimates of the forward and discount functions. Their findings indicate that splining the forward rate function with GCV produces the best overall results.

2.2.3. Nelson-Siegel: A more parsimonious approach

In the Nelson and Siegel (1987) (NS hereafter) approach, the relationship between yield and maturity is derived from the assumption that spot rates follow a second-order differential equation and that forward rates, being forecasts of the spot rates, are the solution to this equation with equal roots. Thus we have

$$f(\tau) = \beta_0 + \beta_1 \exp(-\tau/\upsilon) + \beta_2[\tau/\upsilon \exp(-\tau/\upsilon)]. \tag{2.10}$$

By integrating (2.10) from 0 to τ, we get the yield to maturity or discount rate function

$$r_d(\tau) = \beta_0 + (\beta_1 + \beta_2)[1 - \exp(-\tau/\upsilon)]/(\tau/\upsilon) - \beta_2 \exp(-\tau/\upsilon). \tag{2.11}$$

Following NS, we parameterize (2.11) in order to fit the yield curves. The equation to be estimated is given by

$$r_d(\tau) = a + b[1 - \exp(-\tau/\upsilon)]/(\tau/\upsilon) + c\ \exp(-\tau/\upsilon) + \eta. \tag{2.12}$$

Where $r_d(\tau)$ denotes yield as a function of maturity τ, and υ is a time constant associated with the differential equation. The parameters a, b and c are to be estimated. NS estimate the model using U.S. Treasury bills and find that it can adequately characterize the shape of the Treasury bill term structure. Cecchetti (1988) uses this method to construct estimates of the term structure from coupon bearing bonds for the period 1929-1949, since it provides for a broad set of alternative shapes while requiring estimation of only a few parameters.

The method has been used in a slightly different manner by Baum and Thies (1992) in that they apply it to smooth or post-filter their cubic spline term structure estimates, obtained from applying McCulloch's methodology to U.S. railroad bond quotations. In a recent paper, Bliss (1997) applies an extended form of the Nelson-Siegel approximating function (2.11) directly to Treasury bond prices. In this context, the estimation problem is rendered highly nonlinear by the inclusion of all of the security's cashflows in a single observation of the dataset.[7] In contrast, the use of Treasury STRIPS, each with a single cashflow, greatly simplifies the estimation problem.

3. Data and Estimation Methods

In this section, we briefly describe the data and outline the estimation methods which will be compared in the empirical analysis of the next section. We do not present the estimation results in this paper since our focus is on the forecasting accuracy of the various models. Details on the estimation results are available upon request.

3.1. The Data

In order to estimate the NS model for longer maturities, we use data on U.S. Treasury STRIPS from quotes in the *Wall Street Journal* on the last trading day of the month every three months starting in June 1989 (when the *Journal* started publishing these quotes) and ending in September 1995, a total of 26 quarter-end dates. The number of Treasury STRIPS quotations used in the Nelson-Siegel approach is between 116 and 120, which include all coupon strips from quoted issues. Although data on Treasury principal STRIPS are also available for this

[7] Bliss' approach also takes the recorded bid and ask prices into account; in his treatment, predicted prices within the observed bid-ask spread are considered to have zero prediction errors.

period, we have chosen to exclude them from the analysis on the advice of practitioners. Many of the heavily traded strippable Treasuries are "on special," implying that they may be used as specific collateral in repurchase agreements (RPs) at an RP rate significantly below market rates. This raises the price of such a security above that which would otherwise prevail, by an amount related to the lower future borrowing costs (for details on the workings of this relationship, see Duffie, 1996, pp. 494-496). To link this discrepancy (which has been found to explain a significant portion of the "on-the-run" effect) to coupon and principal STRIPS, we need only consider practitioners'claims (e.g. Gilles, 1997) that the "specialness" of the coupon security is fully embodied in the principal STRIP, rather than allocated evenly throughout the security's decomposed cash flows. Given the large number of relatively small coupon cash flows, the concentration of "specialness" in the principal STRIP is surely a reasonable assumption. Restricting our analysis to coupon STRIPS generates a more homogeneous data set for the estimation of the zero-coupon (spot) curve.[8]

For the spline models, we used between 226 and 245 quotes of Treasury bills, notes, and bonds, drawn from the CRSP Daily Government Bond Master data set[9]. A detailed description of the data for each date is available upon request.

[8]We also estimated the models in this paper with coupon and principal STRIPS. Although the mean forecast errors are larger at all forecast horizons when principal STRIPS are included, our conclusions regarding the relative rankings of the models remain unchanged.

[9]All available quotations on bills, notes and bonds (with the exception of "flower bonds") were extracted from the CRSP Daily Government Bond Master dataset. Noncallable securities were flagged as such; callable securities priced at a discount to par were given their stated maturity, while (by "Street" convention) callable securities priced at a premium were given their next call date as their maturity. Our analyses can readily be executed omitting all callable securities.

3.2. Splining Methods: Fixed and Endogenous Knots

Following Fisher et al. (1995, 1996), we employ both fixed knots as well as adaptive knot placement specifications to estimate the following five models[10]:

DF1: The discount function, $P_d(\tau)$, is specified as a cubic spline with the number of interior knots set equal to one-third of the number of bonds in the sample.

LDFGCV: The log of the discount function is modelled as a cubic spline with adaptive parameters chosen with GCV.

FGCV: The forward rate function, $f(\tau)$, is modelled as a cubic spline with adaptive parameters chosen with GCV.

DFGCV: The discount function, $P_d(\tau)$, is specified as a cubic spline with adaptive parameters chosen with GCV.

DF2: The discount function, $P_d(\tau)$, is specified as a cubic spline with the number of interior knots set equal to 10.

In the implementation of the GCV methods (the second, third and fourth), a large number of knots is initially specified (again, equal to one-third of the number of observations) and the effective number of parameters is determined according to the optimal value of the smoothness parameter λ in (2.8)[11]. The models are estimated for the 26 sample dates from June 1989–September 1995 using the average of bid and ask prices quoted for the Treasury securities in the CRSP Daily Government Bond Master dataset.

[10] We are indebted to Mark Fisher for providing us with his *Mathematica* programs to perform the empirical analysis for this section.

[11] See Fisher et al. (1995, 1996) for more details on the implementation of the splining techniques using B-Splines.

3.3. The Nelson-Siegel Estimates

The Nelson-Siegel methodology is applied to data on U.S. Treasury coupon STRIPS to obtain estimates of the term structure from the approximating function (2.12) on the 26 quarter-end dates between June 1989 and September 1995. $r_d(\tau)$ is calculated from the average of the quoted bid and ask prices as the continuously compounded yield from settlement to maturity date annualized to a 365-day year. In our implementation, τ is the number of years to maturity. Following NS, we search over a grid of values for υ to obtain the best fitting values for a, b and c and υ.

4. Evaluation of Alternative No-Arbitrage models of the Term Structure

The no-arbitrage approaches differ in two main respects: first, in terms of the functional form fitted to the bond price data, and second, with respect to the criterion function to be minimized to obtain the parameter estimates. In this section, we assess the performance of the various treatments of these aspects cross-sectionally as well as over time in terms of in- and out-of-sample price and yield errors. We concentrate our attention on out-of-sample performance, given the focus of this study. Although the out-of-sample errors include the effects of day-to-day shifts of the yield curve, the comparisons of methods' performance on this basis is a more realistic evaluation of their varying abilities.

The price error for each bond is calculated from (2.5) as:

$$\hat{\epsilon}_i = P_i(\tau) - \sum_{j=0}^{\tau} C_i(t+j)\hat{P}_d(t+j), \tag{4.1}$$

with $\hat{P}_d(\tau)$ as the estimated discount function. For the five splining models described in section 3.3, these are identical to the estimation errors. For the NS approach, we calculate $\hat{P}_d(\tau)$ from the predicted spot rate, $\hat{r}_d(\tau)$ using (2.1) and (2.12).

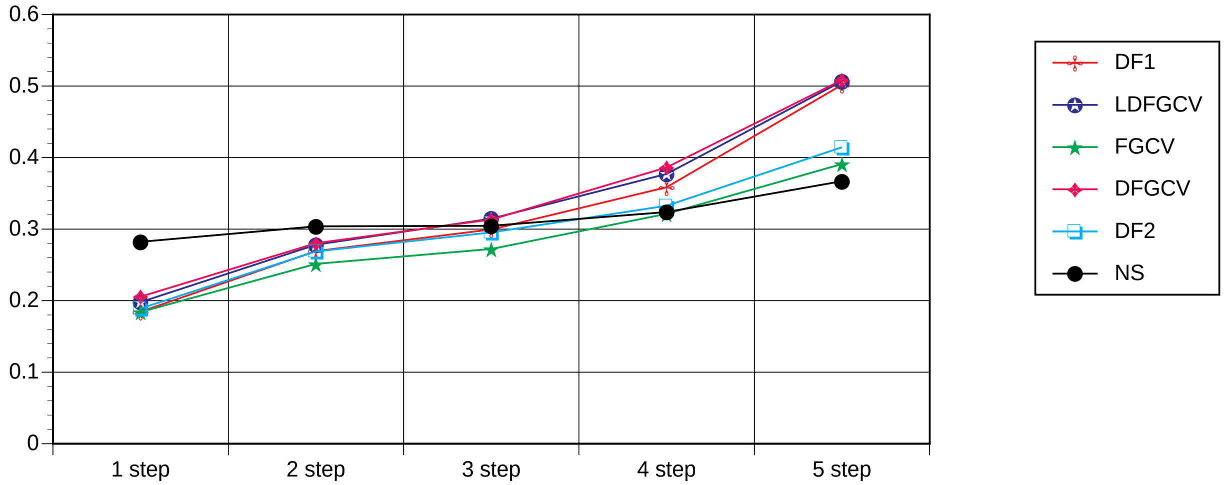
4.1. Price Errors

In order to evaluate the overall performance of the models, we start by examining summary statistics of the price and yield forecast errors, evaluating the models' abilities to predict bond prices and yields both in- and out-of-sample. The price and yield prediction errors may be compared in terms of root mean squared error (RMSE), mean absolute error (MAE), or mean absolute percentage error (MAPE). The average price errors for all maturities for each method (averaged over the 26 quotation dates) are given in Table 1 for the quotation date and one-, two-, three-, four- and five-step-ahead forecasts, while the corresponding median price errors are presented in Table 2. It is obvious from both tables that the results differ greatly between in- and out-of-sample. We focus on the out-of-sample results since the more highly parameterized model will almost always do better in sample. Using median MAPE as the preferred criterion, we find that model FGCV (splining the forward rate curve with GCV) is best for one, two and three steps ahead and is second to model NS (the Nelson-Siegel approach) at five steps ahead. Somewhat similar results are obtained using median MAE, in that model FGCV is the best model for two to four steps ahead. The results appear to be different if we use median RMSE as the criterion, in that model FGCV is best for three to five steps ahead and NS does not appear as either of the two best models.[12] It should be noted that the rankings do not vary over criteria when we use the mean over dates rather than the median.

Figure 1 plots the median MAPE for each model over the forecast horizon

[12] We focus on the MAPE criterion, as we are in agreement with Bliss (1997) who states that there is "...no particular economic rationale for utilizing a squared-error loss function other than econometric convenience."

Figure 1
Median MAPE for Price Errors over Forecast Horizon



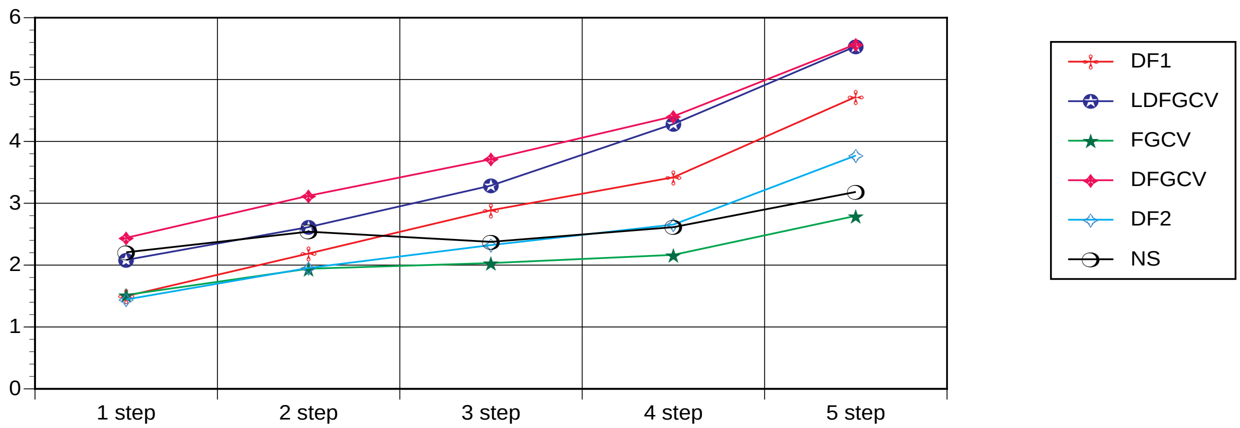
for price errors. The median MAPE increases steadily for all models except the NS model where the increase is less pronounced. We can see from Figure 1 that the NS median MAPE relatively improves at three steps ahead and becomes lower than all other models at five steps ahead.

4.2. Yield Errors

The average yield errors for all maturities for each method (averaged over the 26 quotation dates) are given in Table 3, with median yield errors over the quotation dates in Table 4. RMSE and MAE figures are given in basis points. Using median MAPE as a criterion, FGCV is best for two to five steps ahead; similar results are obtained using RMSE and MAE. NS is best only at five steps ahead using MAE while it's second best at the same horizon using MAPE.

Figure 2 plots the median MAPE for each model over the forecast horizon for yield errors. As with price errors, the NS median MAPE holds fairly steady for yields, although in this case, NS remains second best to model FGCV. The

Figure 2
Median MAPE for Yield Errors over Forecast Horizon

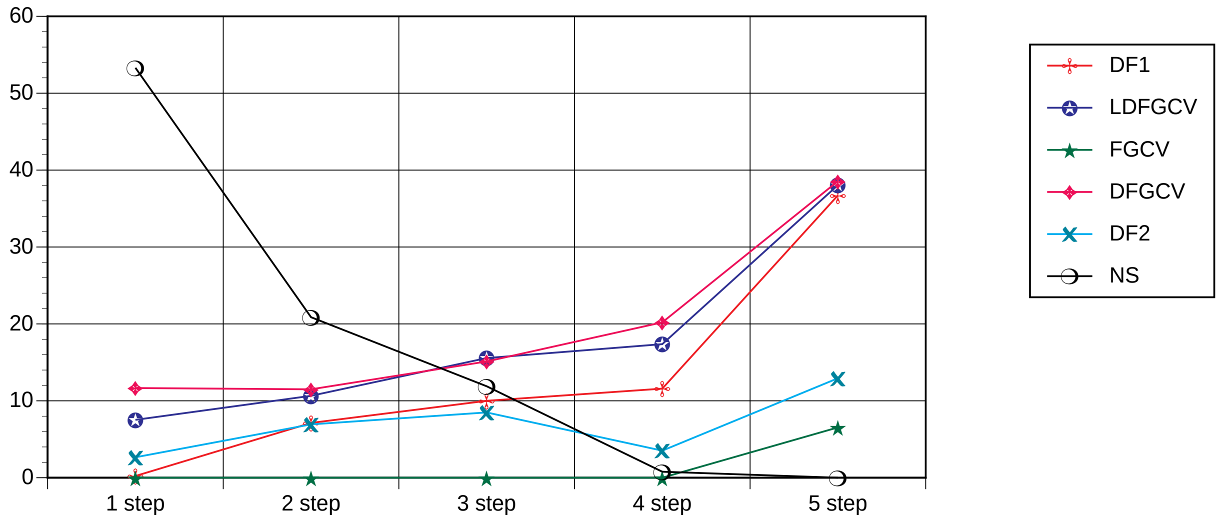


other models'forecasts worsen with the forecast horizon. Overall, model FGCV is the best model both in terms of price and yield errors. NS appears to perform best when applied to longer-horizon forecasts.

4.3. Cost of the Second-Best Model

In this section, we ask how well does each of the competing models track the best model. We compute the percentage increase in median MAPE for each model as compared to the best model. This measure is meant to capture the "penalty" incurred (in percent of MAPE) by the modeler when the best model is not used. Figures 3 and 4 plot the penalty for price and yield errors respectively. The yield penalty is larger than the price penalty for all models at all forecasting horizons. Model DFGCV (splining the discount function with GCV) has the largest yield penalty at all forecasting horizons followed by model LDFGCV (splining the log of the discount function with GCV) and model DF1 (splining the discount function with a large number of knots). With the exception of the NS method, the penalty for all models increases with the forecast horizon. In terms of yield

Figure 3
Percentage Penalty over Best Method based on MAPE of Price Errors

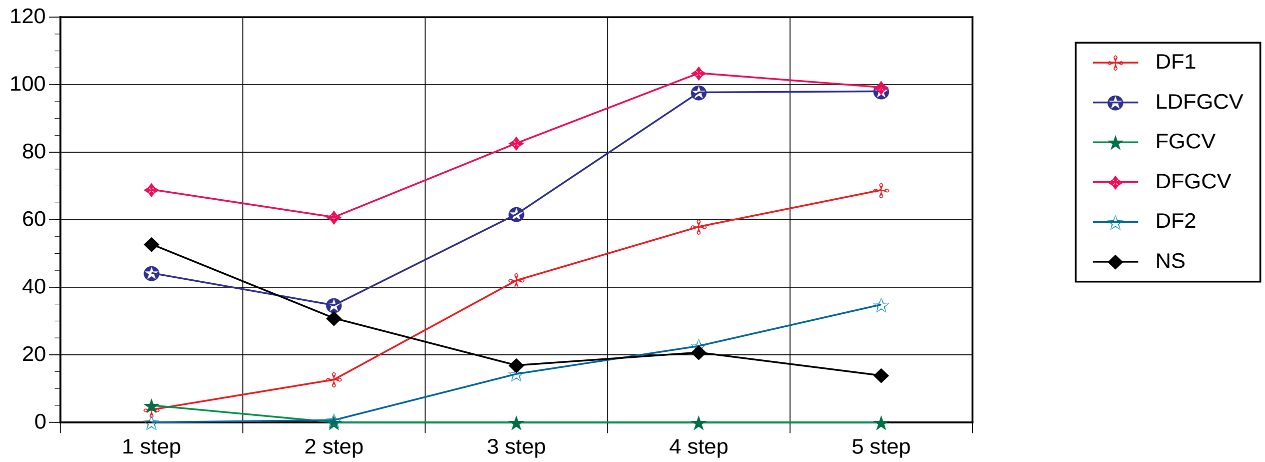


error, the parsimonious model DF2 (splining the discount function with a fixed number of knots: McCulloch's method) has a very low penalty at one to three steps ahead (less than 20%) while the NS model has the lowest penalty at four and five steps ahead.

4.4. Time Series Analysis of Results

Tables 5 and 6 show the best and runner-up models based on the MAPE for price and yield errors respectively at each date and for all forecasting horizons. By looking at the performance of the models over time, we can see that the average results over all dates may be misleading. For example, for price errors, at the one-step ahead forecast horizon, the overall best model, FGCV, has the lowest MAPE only 7 out of 26 times, while model DF1 is best on 11 out of 26 days, and model NS is best for 4 out of 26 dates. At two steps ahead, model FGCV is best on 10 of 26 dates, model NS on 7 of 26 dates, and model DF1

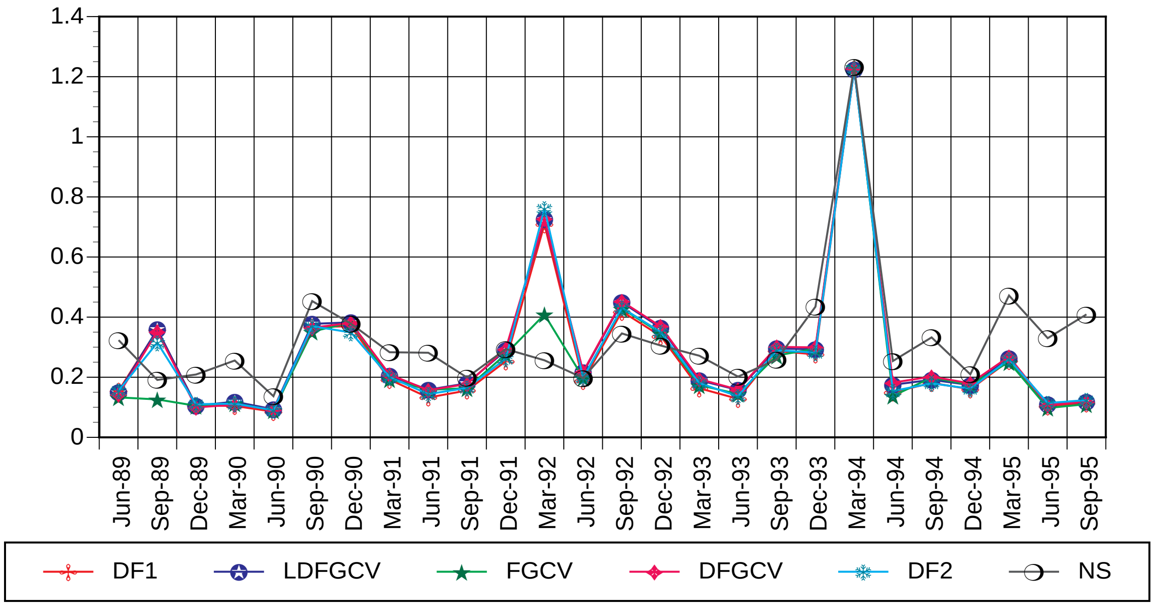
Figure 4
Percentage Penalty over Best Method based on MAPE of Yield Errors



on 6 of 26 dates. The percentage of times that each model is best (in terms of MAPE) is given in Tables 7 and 8 for price and yield errors respectively. The most striking result is that the performance of the NS model closely trails that of the best model, FGCV, at the two- to four-step-ahead horizon for price errors. At five steps ahead, model NS is best 42 per cent of the time as compared to only 31 per cent for the overall best model FGCV. Model NS's performance is even more impressive in terms of yield errors where it outperforms model FGCV at three to five steps ahead, approaching a high of 54 per cent as compared to 31 per cent for model FGCV. Model DF2 does relatively well at short horizons in terms of yield errors, whereas the much less parsimonious model DF1 also does better in terms of price errors at short horizons (1-3 steps ahead). Given that the overall best model (model FGCV) fails to excel more than 50 per cent of the time suggests the need for further investigation of the performance of the models over time.

Figures 5 and 6 plot the time series of the median MAPE for price errors

Figure 5
Time Series of Median MAPE of Price Errors from One-Step-Ahead Forecasts



at one- and five-steps-ahead respectively. We can see that the performance of the models varies a great deal over the sample period. Figure 5 illustrates that the one-step-ahead MAPEs for all models seem to move together over time, except for that of model NS, which does not closely track the other models on a number of dates. For example, on March 1992, all models' MAPEs increase except for that for model NS. After June 1994, model NS's performance appears to deteriorate relative to all other models. The picture looks similar at five steps ahead but with NS performing considerably better on all dates especially on March 1992. Although all models' MAPEs increase on that date, model NS's performance is much better than that of the other models. Next, we examine whether the time series performance of the models is related to movements in the yield curve over the forecast horizon.

Figure 6
Time Series of Median MAPE of Price Errors from Five-Step-Ahead Forecasts

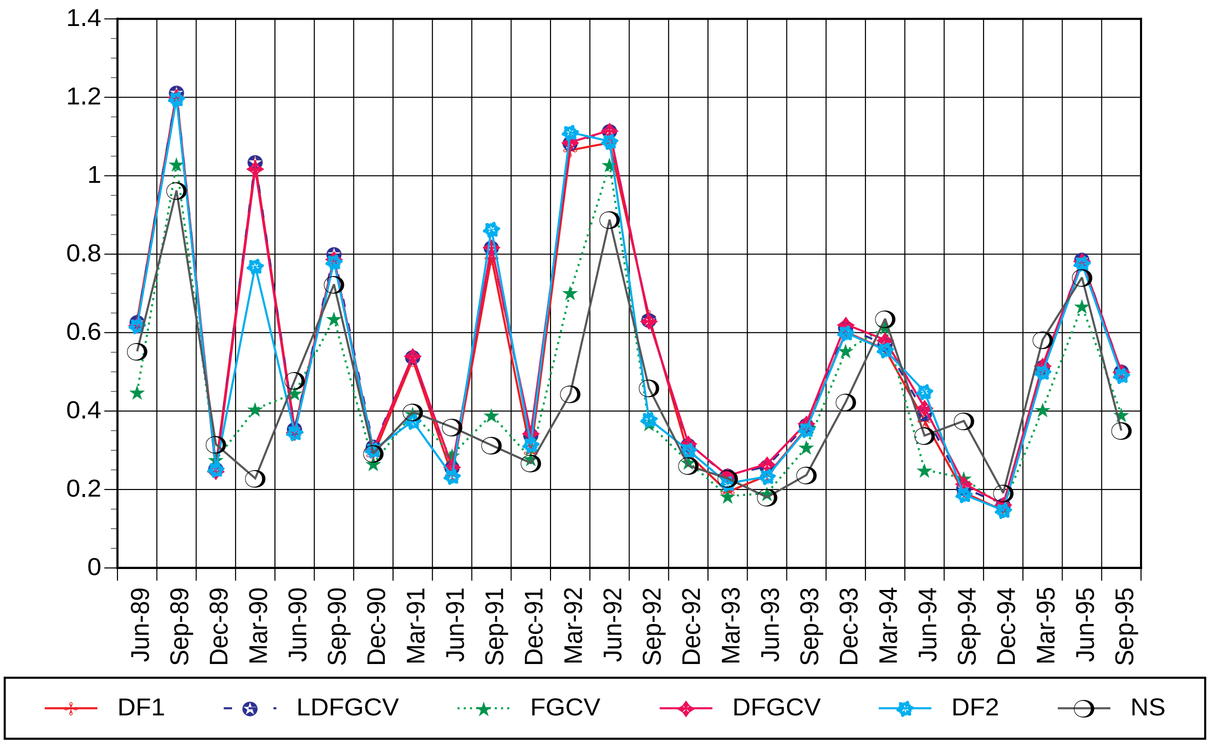
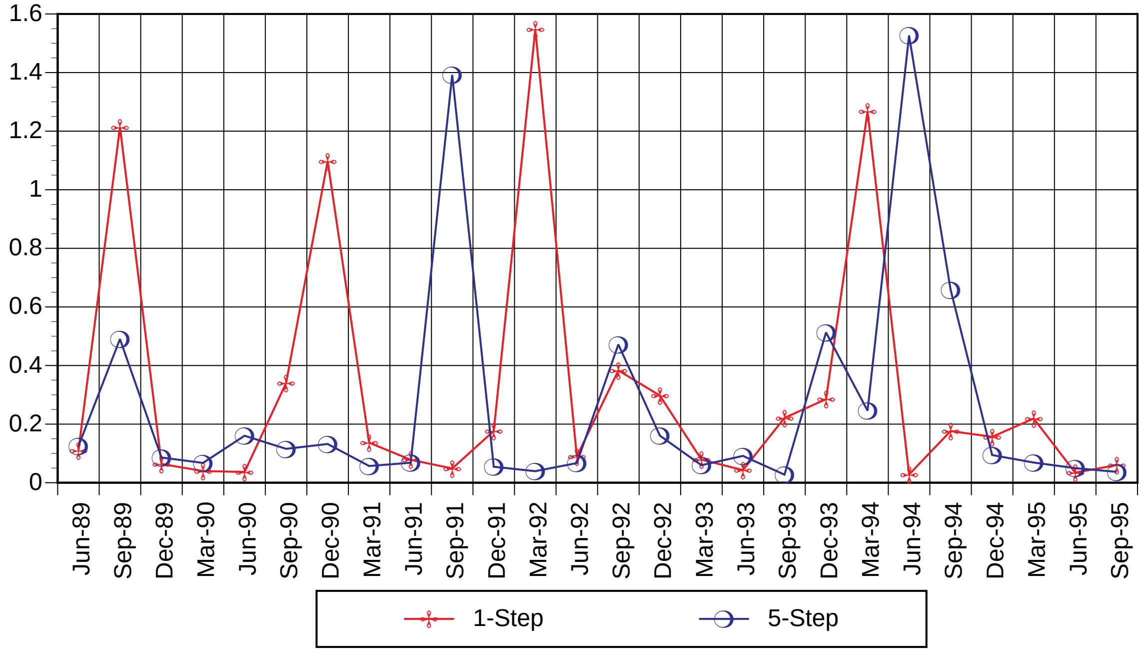


Figure 7

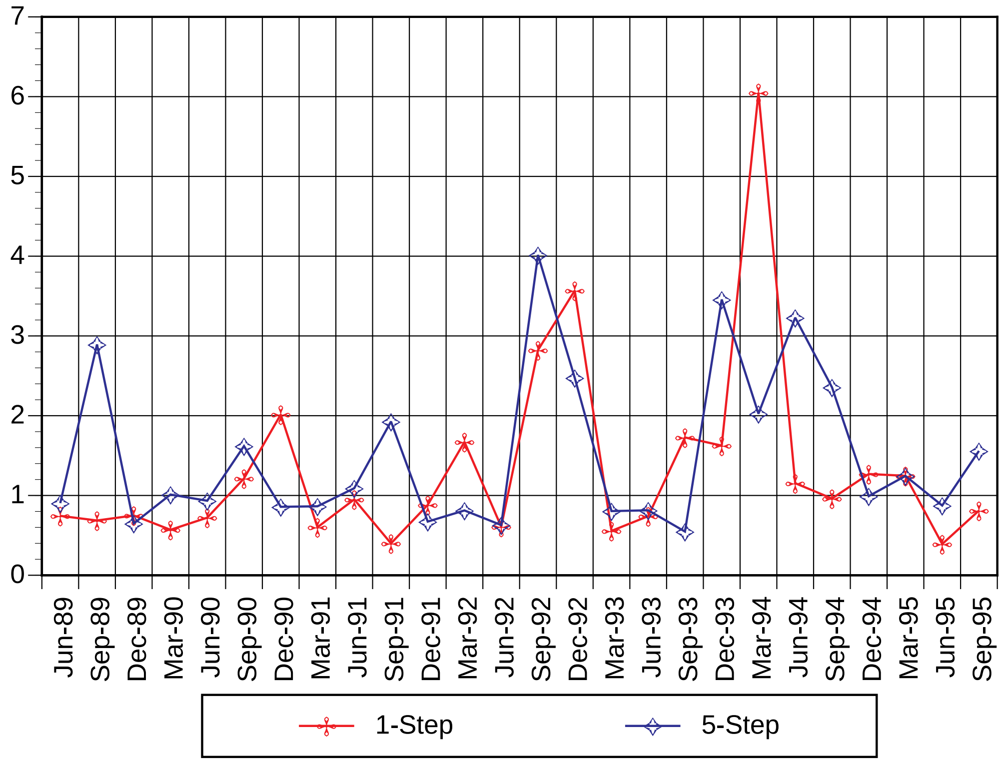
Average Absolute Percentage Change in Bond Prices at One and Five Steps Ahead



We compute the average absolute percentage change in bond prices and bond yields at each date of our sample in an attempt to measure shifts in the yield curve between the quote and the forecast dates. Figures 7 and 8 show plots of these series for bond prices and yields, respectively, at the one- and five-steps-ahead forecast horizons.

On several dates, including March 1992, bond prices on average change by more than one per cent one day after the estimation date, whereas changes in bond yields can be as large as six per cent. The largest one-step-ahead average price increase occurs in March 1992, when all models' MAPE increased (except for model NS, as Figure 5 indicated). One question is the extent to which movements in the yield curve, as measured by average absolute percentage changes, are related to the change in forecast MAPE of the models. Tables 9 and 10

Figure 8
Average Absolute Percentage Change in Bond Yields at One and Five Steps Ahead



report simple correlation coefficients of the price and yield change and the price MAPE series over our 26 sample dates. The correlations between price MAPE and average price and yield changes are positive and quite high at one-step-ahead for all models; in general they are much lower at two- to five-steps-ahead. In fact, they become negative at the four-step-ahead horizon, suggesting that models' forecasting performance improves the larger the absolute change in the yield curve. Note that the overall best model, FGCV, and model NS, have the lowest correlations at one-step-ahead when we express movements in the yield curve in terms of price changes.[13] This reflects the results illustrated in Figures 5 and 6 and is also apparent in Figure 9, which plots the average price change series against the median price MAPE for models FCGV and NS. Figure 9 indicates that model FGCV's and NS's MAPEs do not increase commensurately with the price change on those dates with large price changes, such as Sep. 1989, Dec. 1990, March 1992, and March 1994.

4.5. Analysis of Whole Sample Results

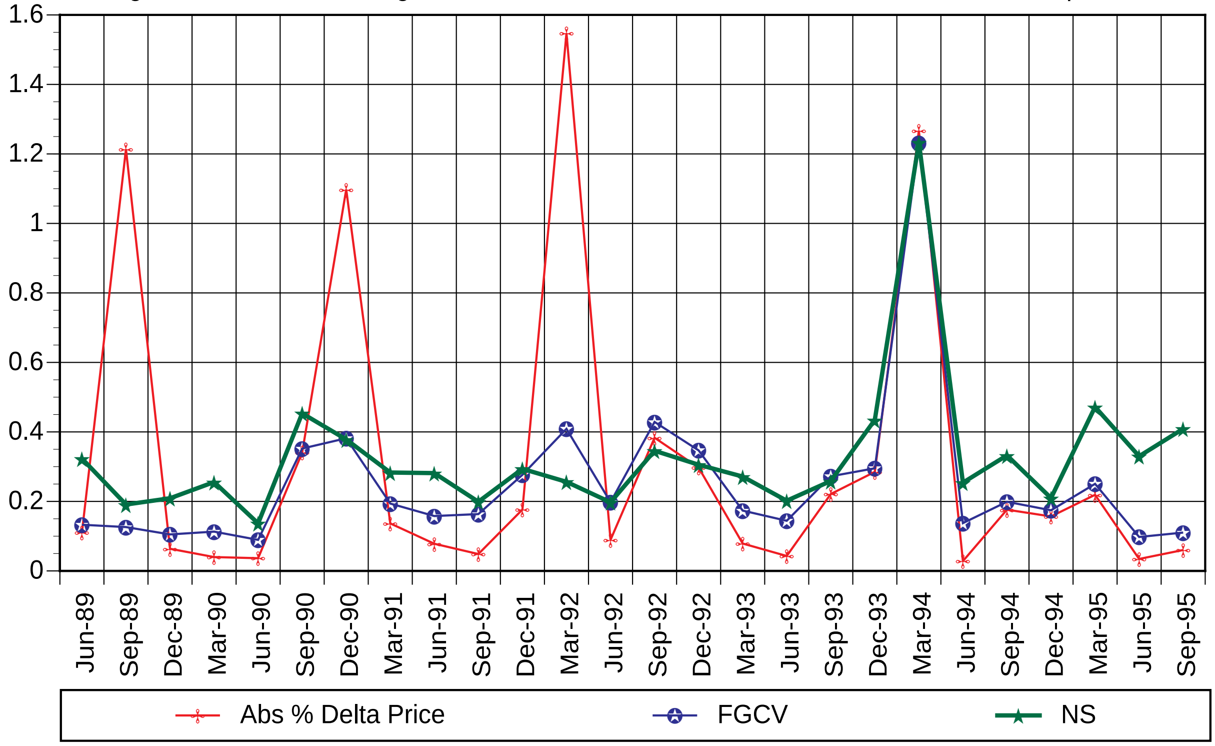
Before we turn to evaluating the models'forecast performance in various maturity ranges, let us summarize the results we have so far for the whole sample (all maturities). As previously stated, our comparison centers on the models' ability to predict bond prices and yields up to five days ahead, thus testing the models in an actual out-of-sample setting. Several main conclusions emerge from this analysis.

First, as expected, the performance of the models varies a great deal between in- and out- of sample. Models' relative rankings depend on the forecast horizon and on whether price or yield errors are compared. On average over the 26 sam-

[13]When yield curve shifts are expressed in terms of yield changes, the NS model has the lowest correlation at one-step-ahead, while model FGCV has the highest correlation.

Figure 9

Average Absolute Price Change and MAPE of Models FGCV and NS Forecasts, One Step Ahead



ple dates, the highly parameterized model DF1 (splining the discount function with a large number of knots) is always best in sample. In terms of yield errors, model FGCV (splining the forward rate curve with GCV) excels out-of-sample (from two to five steps ahead) but only for one to three steps ahead for price errors. Interestingly, the NS model, estimated from STRIPS data, is best for the longest horizon price forecast (five steps ahead) and second best at the five-step horizon in terms of yield errors. At long horizons, the loss in forecast accuracy (as measured by percentage increase in MAPE) from using the "wrong" model is as high as 40% for price errors and 100% for yield errors. Overall, models DFGCV (splining the discount function with GCV) and LDFGCV (splining the log of the discount function with GCV) appear to be the least reliable at all forecasting horizons.

Second, comparing the models over time suggests that relying solely on measures of average performance may overstate the case for the overall best model (FGCV). Particularly, model FGCV is best at most in 42% and 38% of the cases for price and yield errors respectively. In terms of price errors, models DF1 and NS closely track the overall best model for short and longer horizons respectively. In terms of yield errors, DF2 is best at least 26% of the time for short horizons while NS is best in up to 54% of the cases at the five-step horizon.

Third, there is a high positive correlation between movements in the yield curve (as measured by average price and yield change from the quote date) and models' MAPEs at the one-step horizon for all models. The correlations are considerably lower at two and three steps ahead and become very small and negative at four and five steps ahead. This suggests that, at least for long-horizon forecasts, the performance of the models cannot be attributed to movements in the yield curve. The fact that the NS model has some of the lowest correlations of all models may be viewed as beneficial in hedging against unexpected large movements in the yield curve.

4.6. Performance by Tenor Category

A common difficulty with any term structure estimation method is presented by the convexity of the price/yield relation, which causes price and yield errors of very different magnitudes to be generated along the yield curve. Many estimation techniques have attempted to account for this relation by duration weighting, minimizing the product of price and yield errors, and so on. Our goal in this study is not to add to this arsenal of methodology, but to evaluate the suitability of several techniques. It is likely that a technique which works quite well at short maturities may not exhibit the same performance at long maturities, and vice versa. To consider the relative suitability of the several techniques for securities of short, medium and long tenors, we summarize ex ante forecast performance for four tenor categories: 0-3 years, 3-7 years, 7-15 years, and greater than 15 years to maturity, where the category for a particular security is defined by its current remaining term.

More than half of the price quotations used in the analysis are associated with securities with no more than three years to maturity. Between 49% and 52% of the quotations fall in this range, with an additional 23% (22-26%) associated with securities in the 3-7 year tenor range. On average, 15% (12-17%) of the quotations are associated with medium-term tenors (from 7 to 15 years to maturity) while the remaining 11% (10-14%) are associated with longer term securities. The heavy weight given to short-term (0-3 year) securities in our analysis suggests that forecasting reliability must be associated with reasonable performance at the short end of the yield curve.

Tables 11 and 12 show the median MAPE for each tenor category for price and yield errors respectively, while Tables 13 and 14 present the percentage penalty (in terms of median MAPE) incurred by choosing a method other than the best method for each forecast horizon and method. In the short-term tenor bracket, model FGCV is superior at all but the one-step-ahead horizon, while model NS is the least accurate method at each horizon. In the 3–7 year bracket, model NS performance steadily improves, surpassing model FGCV by small margins at four and five steps ahead. Models DF1 and LDFGCV are also contenders in this bracket; in the 7-15 year bracket, model DF1 is marginally superior to FGCV at two steps ahead, with model NS superior at four and five steps ahead. A somewhat similar pattern is realised for long-term securities, where model DF1 is superior at one and four steps ahead, model FGCV only at the two-step-ahead horizon, and model NS taking the remaining horizons. Thus, there are few clear lessons to be learned at medium to long tenors; model FGCV's performance never bears more than a ten per cent penalty over the best model, but it is not dominant in those three brackets. For short-term securities, the picture is clear, with model FGCV superior by a considerable margin. The three- to five-step-ahead price error in that 0–3 year category is in the range of 10 to 12 cents per $100.00. Substantially greater errors are realised at longer horizons; five-step-ahead price errors range from 43 cents (at 3–7 years) to 79 cents (at 15-30 years) per $100.00, reflecting the convexity of the price-yield relation.

Turning to yield errors, model FGCV stands out, with superior performance at three to five steps ahead in the short tenor range, at one to three steps ahead in the 3–7 year range, and at all ex ante horizons for the 7–15 year range. Its performance is weaker for long-term securities, with model NS superior at three and five steps ahead and model DF2 best at two steps ahead. Model DF1 also has some power in the long-term bracket. Of considerable interest is the weak in-sample performance of model FGCV for 0–3 and 15–30 year brackets;

from the in-sample results, one would not consider this method a success, but the out-of-sample forecasting performance is much stronger. The penalty for unconditional use of model FGCV is somewhat higher for the 3–7 year and 15–30 year categories, with model FGCV underperforming model NS by up to 17 per cent and 16 per cent, respectively. The yield errors (measured by median MAPE) decline with tenor, amounting to 2–4 per cent for short-term securities, but only about one per cent for long-term securities. At a 7% yield, this would amount to about 30 basis points on a short-term security, and under 10 basis points for a long-term security.

4.7. Performance for Treasury Bills

All of the methods applied here make use of Treasury Bill quotations to anchor the short end of the estimated term structure. Some researchers (e.g. Bliss, 1997) have found that term structure estimation methods which use both bill quotations and coupon security quotations perform very poorly in predicting Treasury bill prices and/or yields. Since Treasury bill pricing is an essential element of many strategies in the interest rate derivatives markets, poor performance would be of considerable concern. Accordingly, we investigate the reliability of the six methods applied in this paper over the set of Treasury bill quotations in terms of ex ante forecast accuracy. There are between 30 and 33 Treasury bills in each quotation day's dataset: roughly half of the quotations available in the 0-3 year tenor category considered above. Table 15 presents the median MAPE for Treasury bill price and yield prediction over the 26 quote dates, while Table 16 presents the percentage penalty (in terms of median MAPE) suffered by choosing a method other than the best method for each forecast horizon and method.

In this comparison, model FGCV is the most accurate price predictor at horizons of three, four and five days ahead, while model DF2 (using a fixed number of knots) is either first- or second-best at all horizons, with similar MAPE statistics for all but the five-day horizon. The Nelson-Siegel (NS) model is never better than third, with sizable penalties over the best model at all horizons. This same pattern of findings holds for yield errors, in which model FGCV is again superior for three-step-ahead or longer forecasts (while exhibiting much weaker performance on the quote date). Model DF2 is also quite reliable, with percentage penalties of no more than 35% over model FGCV in terms of median MAPE.

In summary, then, it appears that model FGCV performs quite well in terms of ex ante forecasts of Treasury bill prices and yields. At the three-, four- or five-day forecasting horizon, this model achieves percentage pricing errors of two to three cents per $100.00 and percentage yield errors of 1.5 to 3.0 per cent–equivalent to less than 20 basis points at rates of 5–7%. Considering that the models generating these errors have not been fit over bills alone, this would seem to be very respectable performance indeed. A serious effort to generate ex ante forecasts of Treasury bill prices and yields would presumably consider only bill quotations, but would have limited ability to improve upon the forecasting results summarized here.

5. Conclusions

We compare six term structure estimation methods in terms of their ex ante price and yield prediction accuracy. Our contribution centers on comparing the various models based on actual out-of-sample price and yield forecasts from one to five days ahead. Relative to in-sample and split-sample comparisons made in other studies, this approach is more realistic and does not depend on arbitrary division of the sample. Further, this approach enables us to evaluate whether the models' price and yield forecasting accuracy depends on the forecast horizon, an issue of likely interest to many practitioners. We evaluate five splining methodologies that differ in the choice of function which is splined and the degree to which the parameterization is determined by the data. The sixth method is the well-known Nelson-Siegel approach, in which the yield curve is derived from a second-order differential equation applied to Treasury coupon STRIPS data.

On average over the 26 sample dates, model FGCV (splining the forward rate curve with generalized cross-validation) produces the lowest out-of-sample price and yield forecast errors for most horizons. The model appears to perform particularly well for securities with 0-3 years to maturity in terms of price errors and for all securities up to 15 years to maturity in terms of yield errors. We find that model FGCV is best in terms of price and yield forecasts for Treasury bills at the three-to five step ahead horizons. This contradicts findings reported by Bliss (1997) indicating that this splining methodology is not appropriate for modeling short-maturity securities' prices or yields.

When we compare the models over time, we find that the overall best model (FGCV) is best at most in 42% and 38% of the cases for price and yield errors respectively. Surprisingly, the highly parsimonious NS model excels at long forecast horizons: up to 42% of cases for price errors and 54% for yield errors. Of the splining methodologies, models LDFGCV (splining the log of the discount function with GCV) and DFGCV (splining the discount function with GCV) are always dominated, whereas model DF2 (splining the discount function with a fixed number of knots: McCulloch's method) is best in up to 38% of the short horizon forecasts of yield errors.

In summary, we conclude that a general classification or ranking of the models is not reasonable, since a model's relative performance varies with time, level of rates, forecast horizon and tenor category. However, models FGCV, NS and

DF2 emerge as the most promising candidates for the generation of short-term out-of-sample forecasts, depending on the modeler's particular interest.

Table 1: Average Price Errors for All Maturities

(a) Root Mean Squared Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.247152**	0.280955	0.279429	0.276835	0.282603	0.415137
1-S	0.467394	0.489547	**0.450985**	0.486379	0.483983	0.550584
2-S	0.530724	0.549203	**0.502758**	0.546171	0.543202	0.589320
3-S	0.615749	0.632083	**0.538717**	0.629263	0.598356	0.592420
4-S	0.644560	0.660581	**0.578643**	0.657753	0.630593	0.623941
5-S	0.821705	0.834508	**0.683226**	0.831851	0.784347	0.706681

(b) Mean Absolute Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.118641**	0.141526	0.138952	0.142651	0.136663	0.266981
1-S	0.306624	0.321723	**0.287469**	0.322428	0.312000	0.363949
2-S	0.370173	0.385591	**0.338654**	0.385897	0.374609	0.393880
3-S	0.442600	0.457564	**0.370442**	0.457432	0.425520	0.401244
4-S	0.474783	0.489388	**0.406943**	0.489024	0.459231	0.426527
5-S	0.607245	0.620319	0.487448	0.620403	0.583327	**0.481311**

(c) Mean Absolute Percentage Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.104275**	0.126006	0.12348	0.127389	0.120736	0.242216
1-S	0.268347	0.28327	**0.252339**	0.284039	0.273830	0.327563
2-S	0.324722	0.340017	**0.297143**	0.34034	0.329516	0.353276
3-S	0.391217	0.406237	**0.326710**	0.406186	0.376842	0.360041
4-S	0.421863	0.436579	**0.360812**	0.436256	0.408825	0.384419
5-S	0.536388	0.549447	**0.429919**	0.549629	0.516639	0.431550

Note: minimum entry for each horizon displayed in **bold**.

Table 2: Median Price Errors for All Maturities

(a) Root Mean Squared Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.249058**	0.288407	0.277639	0.278205	0.292936	0.403516
1-S	**0.328411**	0.365883	0.344186	0.361772	0.368867	0.494385
2-S	**0.425066**	0.439028	0.425831	0.435771	0.437237	0.494093
3-S	0.512558	0.530665	**0.499083**	0.526479	0.51079	0.550869
4-S	0.550141	0.571718	**0.515831**	0.565973	0.544487	0.552410
5-S	0.739052	0.749959	**0.586355**	0.748821	0.669604	0.667095

(b) Mean Absolute Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.114763**	0.138124	0.136912	0.14277	0.131931	0.265084
1-S	**0.204302**	0.223907	0.210373	0.228109	0.218591	0.325144
2-S	0.312434	0.321912	**0.28549**	0.321525	0.312286	0.341022
3-S	0.351165	0.369439	**0.325941**	0.372604	0.344615	0.344603
4-S	0.407639	0.418785	**0.361164**	0.426605	0.384389	0.362819
5-S	0.564271	0.568839	0.439222	0.570685	0.478575	**0.391155**

(c) Mean Absolute Percentage Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.100066**	0.124848	0.121723	0.126517	0.115488	0.230192
1-S	0.184534	0.197977	**0.184126**	0.205595	0.189017	0.282279
2-S	0.269329	0.278195	**0.25142**	0.280321	0.268853	0.303894
3-S	0.299513	0.314646	**0.272283**	0.31344	0.295393	0.304658
4-S	0.358618	0.377209	**0.321386**	0.386312	0.332687	0.323885
5-S	0.501725	0.506632	0.390983	0.508492	0.414452	**0.366991**

Note: minimum entry for each horizon displayed in **bold**.

TABLE 3: AVERAGE YIELD ERRORS FOR ALL MATURITIES

(a) Root Mean Squared Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**11.0772**	17.4104	16.2896	19.2617	13.7724	26.8950
1-S	17.5138	20.8659	16.5483	21.6529	**15.9735**	26.9479
2-S	23.5325	26.8094	**18.0754**	27.2927	18.7184	28.3773
3-S	31.2476	33.8326	**20.2626**	34.3368	23.0829	30.0170
4-S	38.6712	41.7911	**26.6707**	42.2327	30.0947	34.6741
5-S	54.6455	56.255	**36.2815**	56.6181	40.3668	39.5553

(b) Mean Absolute Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**5.16757**	7.97423	7.3566	9.14992	5.98373	15.1124
1-S	10.2749	12.7797	**9.86506**	13.3184	9.95255	16.2564
2-S	14.1437	16.9475	**11.5189**	17.245	12.7283	17.2894
3-S	18.4208	21.3589	**12.7975**	21.6328	15.6478	17.7902
4-S	22.2695	25.3247	**14.6303**	25.6235	18.2163	19.1162
5-S	29.3406	32.3426	**17.4074**	32.6549	23.3715	20.2587

(c) Mean Absolute Percentage Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**1.00716**	1.68634	1.47453	1.9552	1.15061	2.88174
1-S	1.80512	2.51811	1.83103	2.70974	**1.77939**	2.96308
2-S	2.49961	3.29901	**2.12751**	3.46735	2.25112	3.17733
3-S	3.2499	4.12509	**2.37139**	4.29548	2.77784	3.27099
4-S	3.89844	4.81413	**2.60819**	4.99136	3.16912	3.49264
5-S	5.07493	6.0789	**3.13233**	6.25878	4.08668	4.04212

Note: minimum entry for each horizon displayed in **bold**.

Table 4: Median Yield Errors for All Maturities

(a) Root Mean Squared Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**10.419**	16.5444	16.5616	17.6965	13.2542	19.7471
1-S	16.2256	19.4071	**14.315**	20.2888	15.4738	17.6396
2-S	22.5143	26.2283	**15.871**	26.3681	17.6228	20.2990
3-S	25.4352	33.8697	**17.331**	33.412	19.3144	18.0887
4-S	34.9253	43.3593	**24.0774**	41.465	28.2569	33.4522
5-S	51.7297	59.3482	30.3817	57.3714	34.9706	**28.8863**

(b) Mean Absolute Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**4.61009**	7.62482	7.14827	8.67079	5.7854	10.7950
1-S	8.30869	11.5518	8.64949	12.2121	**8.1234**	11.2180
2-S	12.8922	16.3565	11.0510	15.4412	**11.039**	12.3801
3-S	15.1336	17.9323	**11.7972**	19.4698	11.8903	12.5821
4-S	19.4305	21.7975	**14.0802**	23.4312	14.7297	14.7815
5-S	25.6167	29.0451	15.0490	29.4610	19.5363	**14.8385**

(c) Mean Absolute Percentage Error

	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.855686**	1.44069	1.18917	1.7846	0.970327	2.07960
1-S	1.49863	2.08346	1.51754	2.44103	**1.44465**	2.20664
2-S	2.1876	2.61492	**1.94167**	3.12146	1.95434	2.54066
3-S	2.88721	3.28566	**2.03273**	3.71294	2.32398	2.37584
4-S	3.42008	4.28164	**2.16549**	4.40533	2.6552	2.61363
5-S	4.72008	5.53727	**2.79604**	5.5704	3.77156	3.18339

Note: minimum entry for each horizon displayed in **bold**.

Table 5: Best and Runner-Up Models by Date and Horizon, MAPE of Price Errors

Date	Quote	1-Step	2-Step	3-Step	4-Step	5-Step
Jun-89	DF1,DF2	FGCV,DFGCV	FGCV,DFGCV	FGCV,DFGCV	FGCV,DF2	FGCV.NS
Sep-89	DF1,LDFGCV	FGCV,NS	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV
Dec-89	DF1,LDFGCV	DFGCV,FGCV	DF1,DFGCV	DFGCV,DF1	DFGCV,LDFGCV	DFGCV,DF2
Mar-90	DF1,DF2	DF1,DFGCV	FGCV,DFGCV	NS,FGCV	NS,FGCV	NS,FGCV
Jun-90	DF1,FGCV	DF1,FGCV	FGCV,DFGCV	FGCV,DFGCV	DFGCV,DF2	DF2,DFGCV
Sep-90	DF1,DF2	FGCV,DFGCV	FGCV,DFGCV	FGCV,DFGCV	FGCV,DF2	FGCV,NS
Dec-90	DF1,FGCV	DF2,DF1	DF1,DFGCV	FGCV,NS	FGCV,DF1	FGCV,DF1
Apr-91	DF1,DF2	DF1,FGCV	FGCV,DF1	DF2,FGCV	DF2,FGCV	DF2,FGCV
Jun-91	DF1,DF2	DF1,DF2	DF1,DF2	DF1,FGCV	DF1,DF2	DF2,DF1
Sep-91	DF1,DF2	DF1,FGCV	DF1,FGCV	FGCV,NS	NS,FGCV	NS,FGCV
Dec-91	DF1,DF2	DF1,DF2	DF2,DF1	DF1,DF2	NS,FGCV	NS,FGCV
Mar-92	DF1,FGCV	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV
Jun-92	DF1,DF2	DF1,NS	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV
Sep-92	DF1,DF2	NS,DF1	NS,FGCV	NS,FGCV	NS,FGCV	FGCV,DF2
Dec-92	DF1,DF2	NS,DF1	NS,DF1	NS,FGCV	DF1,DF2	NS,FGCV
Mar-93	DF1,FGCV	DF1,FGCV	DF1,DF2	DF1,FGCV	FGCV,DF1	FGCV,DF1
Jun-93	DF1,DF2	DF1,DF2	NS,FGCV	FGCV,DF2	FGCV,DF2	NS,FGCV
Sep-93	DF1,FGCV	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV
Dec-93	DF1,DF2	DF1,DF2	DF1,DF2	DF1,DF2	FGCV,DF1	NS,FGCV
Mar-94	DF1,DF2	LDFGCV,DF2	DF2,DF1	DF1,DF2	DF1,DF2	DF2,DF1
Jun-94	DF1,DF2	FGCV,DF1	FGCV,DF1	FGCV,NS	FGCV,NS	FGCV,NS
Sep-94	DF1,DF2	DF2,DF1	DF1,DF2	DF2,DF1	DF2,DF1	DF2,DF1
Dec-94	DF1,DF2	DF1,DF2	FGCV,DF1	DF1,DF2	FGCV,DF2	DF1,DF2
Mar-95	DF1,LDFGCV	FGCV,DF1	FGCV,DF1	FGCV,DF1	FGCV,DF1	FGCV,DF2
Jun-95	DF1,DF2	FGCV,DF1	FGCV,DF2	FGCV,DF2	FGCV,NS	FGCV,NS
Sep-95	DF1,LDFGCV	FGCV,DF1	FGCV,DF2	FGCV,DF2	FGCV,NS	NS,FGCV

Table 6: Best and Runner-Up Models by Date and Horizon, MAPE of Yield Errors

date	Quote	1-Step	2-Step	3-Step	4-Step	5-Step
Jun-89	DF1,DF2	DFGCV,FGCV	FGCV,NS	FGCV,NS	FGCV,NS	FGCV,NS
Sep-89	DF1,DF2	FGCV,NS	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV
Dec-89	DF1,DF2	DF2,FGCV	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV
Mar-90	DF1,DF2	DFGCV,FGCV	FGCV,DFGCV	NS,FGCV	NS,FGCV	NS,FGCV
Jun-90	DF1,DF2	FGCV,DFGCV	FGCV,NS	FGCV,NS	NS,DF2	NS,FGCV
Sep-90	DF1,DF2	FGCV,DFGCV	FGCV,NS	FGCV,NS	FGCV,NS	NS,FGCV
Dec-90	DF1,DF2	NS,DF1	NS,DF1	NS,FGCV	NS,FGCV	NS,FGCV
Apr-91	DF1,DF2	FGCV,DF2	FGCV,DF2	FGCV,DF2	NS,FGCV	NS,FGCV
Jun-91	DF1,DF2	DF2,DF1	DF2,DF1	DF2,DF1	DF2,DF1	NS,FGCV
Sep-91	DF2,DF1	DF2,FGCV	FGCV,NS	NS,FGCV	FGCV,DF2	NS,FGCV
Dec-91	DF1,DF2	DF1,NS	DF2,DF1	NS,FGCV	NS,FGCV	NS,FGCV
Mar-92	DF1,FGCV	NS,FGCV	NS,FGCV	NS, FGCV	NS,FGCV	NS,FGCV
Jun-92	DF1,FGCV	FGCV,DF1	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV
Sep-92	DF1,DF2	NS,FGCV	NS,FGCV	NS,FGCV	NS,FGCV	FGCV,DF2
Dec-92	DF1,DF2	NS,DF1	NS,DF1	NS,DF1	NS,DF2	DF1,FGCV
Mar-93	DF1,FGCV	DF1,FGCV	DF2,FGCV	FGCV,DF2	FGCV,NS	FGCV,NS
Jun-93	DF1,DF2	DF1,DF2	NS,DF2	DF2,FGCV	FGCV,DF2	NS,FGCV
Sep-93	DF1,FGCV	FGCV,DF2	FGCV,DF2	FGCV,DF2	FGCV,NS	FGCV,NS
Dec-93	DF1,DF2	DF2,DF1	DF2,DF1	DF2,FGCV	DF2,FGCV	NS,FGCV
Mar-94	DF1,DF2	FGCV,DF1	DF2,DF1	DF2,DF1	DF2,DF1	DF2,FGCV
Jun-94	DF1,DF2	DF1,DF2	DF2,DF1	DF2,DF1	DF2,FGCV	FGCV,DF2
Sep-94	DF1,DF2	DF2,DF1	DF2,DF1	DF2,DF1	DF2,DF1	DF2,FGCV
Dec-94	DF1,DF2	DF2,DF1	DF2,DF1	DF2,DF1	DF2,FGCV	DF2,FGCV
Mar-95	DF1,DF2	DF2,FGCV	FGCV,DF2	FGCV,DF2	FGCV,DF2	FGCV,DF2
Jun-95	DF1,LDFGCV	FGCV,DFGCV	FGCV,DF2	FGCV,DF2	FGCV,DF2	FGCV,DF2
Sep-95	DF1,LDFGCV	FGCV,DF2	FGCV,DF2	FGCV,DF2	FGCV,DF2	FGCV,DF2

Table 7: Percentage of Cases in which each model is best, MAPE of Price Errors

model	1-Step	2-Step	3-Step	4-Step	5-Step
DF1	42.31	26.92	23.08	11.54	3.85
LDFGCV	3.85	0.00	0.00	0.00	0.00
FGCV	26.92	38.46	38.46	42.31	30.77
DFGCV	3.85	0.00	3.85	7.69	3.85
DF2	7.69	7.69	7.69	7.69	19.23
NS	15.38	26.92	26.92	30.77	42.31

Table 8: Percentage of Cases in which each model is best, MAPE of Yield Errors

model	1-Step	2-Step	3-Step	4-Step	5-Step
DF1	15.38	0.00	0.00	0.00	3.85
LDFGCV	0.00	0.00	0.00	0.00	0.00
FGCV	34.62	38.46	34.62	34.62	30.77
DFGCV	7.69	0.00	0.00	0.00	0.00
DF2	26.92	30.77	26.92	23.08	11.54
NS	15.38	30.77	38.46	42.31	53.85

Table 9: Correlations between Average Price Change and MAPE of Price Errors

model	1-Step	2-Step	3-Step	4-Step	5-Step
DF1	0.80	0.31	0.38	-0.17	0.06
LDFGCV	0.80	0.31	0.37	-0.17	0.07
FGCV	0.62	0.37	0.18	-0.07	-0.06
DFGCV	0.79	0.30	0.36	-0.18	0.08
DF2	0.79	0.30	0.21	-0.14	0.13
NS	0.42	0.32	0.10	-0.09	-0.03

Note: The price change series contains average absolute percentage changes in bond price.

Table 10: Correlations between Average Yield Change and MAPE of Price Errors

model	1-Step	2-Step	3-Step	4-Step	5-Step
DF1	0.85	0.62	0.47	-0.25	0.13
LDFGCV	0.86	0.64	0.48	-0.24	0.12
FGCV	0.92	0.75	0.56	-0.14	0.10
DFGCV	0.86	0.64	0.49	-0.24	0.13
DF2	0.85	0.61	0.52	-0.18	0.09
NS	0.80	0.58	0.38	-0.16	0.16

Note: The yield change series contains average absolute percentage changes in bond yield.

Table 11: Median MAPE of Price Errors by Tenor Category

(a) Securities with 0–3 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**0.041355**	0.057647	0.056517	0.075278	0.045449	0.111705	0.100066
1-S	0.072474	0.094600	0.075782	0.094129	**0.071831**	0.120954	0.184126
2-S	0.100925	0.115865	**0.099629**	0.122959	0.102338	0.121619	0.251420
3-S	0.125254	0.145133	**0.097156**	0.155753	0.116524	0.139872	0.272283
4-S	0.142055	0.160857	**0.125417**	0.169119	0.136017	0.143571	0.321386
5-S	0.171021	0.189539	**0.12643**	0.195266	0.158979	0.164866	0.366991

(b) Securities with 3-7 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**0.0978682**	0.0998163	0.10392	0.10374	0.109199	0.269941	0.100066
1-S	0.235154	0.243411	**0.218118**	0.236971	0.247965	0.285583	0.184126
2-S	**0.267204**	0.270453	0.272475	0.268964	0.278767	0.299114	0.251420
3-S	0.372749	0.382541	**0.328485**	0.379128	0.39006	0.331844	0.272283
4-S	0.448644	0.460271	0.402441	0.460955	0.458951	**0.384470**	0.321386
5-S	0.47986	0.462087	0.450521	0.463972	0.465603	**0.431740**	0.366991

(c) Securities with 7-15 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**0.272192**	0.302664	0.303969	0.296767	0.331185	0.424667	0.100066
1-S	0.429222	0.437532	**0.423399**	0.437647	0.434365	0.501361	0.184126
2-S	**0.519443**	0.534779	0.522022	0.533479	0.534175	0.581360	0.251420
3-S	0.628074	0.637765	**0.580446**	0.641759	0.584812	0.631992	0.272283
4-S	0.629701	0.643622	0.609421	0.640345	0.619902	**0.605199**	0.321386
5-S	0.870721	0.872315	0.743426	0.868834	0.796315	**0.712300**	0.366991

(d) Securities with 15-30 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**0.0541203**	0.154805	0.143715	0.120204	0.149313	0.406867	0.100066
1-S	**0.384072**	0.434007	0.3874	0.411932	0.422089	0.567215	0.184126
2-S	0.613553	0.616728	**0.593907**	0.617579	0.60819	0.595221	0.251420
3-S	0.662924	0.678163	0.617893	0.671576	0.654702	**0.608901**	0.272283
4-S	**0.633187**	0.668914	0.693802	0.65907	0.665908	0.689437	0.321386
5-S	1.01554	1.02524	0.908154	1.0192	1.01229	**0.789631**	0.366991

Note: minimum value over models for each tenor category presented in **bold**.

Value in "Full" is the minimum achieved by any model over all tenors.

Table 12: Median MAPE of Yield Errors by Tenor Category

(a) Securities with 0–3 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**1.32526**	2.31072	1.8487	3.01267	1.41543	2.93978	0.855686
1-S	2.12119	3.33225	2.19015	4.19942	**1.95069**	3.28018	1.44465
2-S	3.24391	3.89878	2.56333	4.83422	**2.50592**	3.48377	1.94167
3-S	4.24714	5.36038	**2.80508**	5.89728	3.16521	3.28293	2.03273
4-S	5.19701	6.87342	**3.42076**	7.69246	3.74546	3.61587	2.16549
5-S	7.29745	9.13318	**3.88385**	9.45022	5.26022	4.11486	2.79604

(b) Securities with 3–7 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**0.395073**	0.40224	0.411712	0.422794	0.417359	1.05745	0.855686
1-S	0.988681	1.0432	**0.903704**	1.02207	1.02594	1.12779	1.44465
2-S	1.09385	1.13516	**1.0878**	1.12339	1.16255	1.16136	1.94167
3-S	1.58999	1.58929	**1.2933**	1.57417	1.61874	1.38380	2.03273
4-S	1.98498	2.01117	1.714	2.003	2.00687	**1.69910**	2.16549
5-S	2.55595	2.46515	1.91168	2.47311	2.37902	**1.63899**	2.79604

(c) Securities with 7-15 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**0.668106**	0.704729	0.715163	0.707078	0.72407	0.93723	0.855686
1-S	0.992218	1.01195	**0.905532**	1.0146	0.997435	1.12664	1.44465
2-S	1.2952	1.30747	**1.15224**	1.3022	1.30886	1.33136	1.94167
3-S	1.43095	1.43157	**1.25853**	1.42719	1.40138	1.54691	2.03273
4-S	1.38948	1.40533	**1.23922**	1.40534	1.34761	1.33109	2.16549
5-S	2.03311	2.0556	**1.44738**	2.05631	1.82028	1.47140	2.79604

(d) Securities with 15-30 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS	Full
Quote	**0.0747147**	0.205923	0.184648	0.155663	0.198777	0.523428	0.855686
1-S	**0.501944**	0.566058	0.511504	0.538426	0.559037	0.718061	1.44465
2-S	0.77826	0.780202	0.782415	0.783337	**0.769142**	0.771668	1.94167
3-S	0.85119	0.875308	0.817194	0.856558	0.837093	**0.754363**	2.03273
4-S	**0.868876**	0.917675	0.923477	0.903982	0.919852	0.932578	2.16549
5-S	1.30013	1.31742	1.18038	1.30755	1.3068	**1.020520**	2.79604

Note: minimum value over models for each tenor category presented in **bold**.

Value in "Full" is the minimum achieved by any model over all tenors.

Table 13: Percentage Penalty of Price Errors by Tenor Category

(a) Securities with 0–3 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	39.3972	36.6643	82.0286	9.90057	170.113
1-S	0.895317	31.6986	5.50057	31.0428	0	68.386
2-S	1.30047	16.2969	0	23.4173	2.71908	22.071
3-S	28.92	49.3804	0	60.3112	19.9348	43.966
4-S	13.266	28.2575	0	34.8453	8.45185	14.475
5-S	35.2694	49.9162	0	54.4467	25.7454	30.410

(b) Securities with 3–7 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	1.99048	6.18355	5.99965	11.5771	175.821
1-S	7.8104	11.5958	0	8.6435	13.6838	30.9307
2-S	0	1.2156	1.97239	0.658413	4.32709	11.9422
3-S	13.4753	16.456	0	15.4171	18.7451	1.0227
4-S	16.6915	19.7157	4.6742	19.8937	19.3723	0
5-S	11.1456	7.0289	4.3501	7.4656	7.8434	0

(c) Securities with 7-15 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	11.195	11.6743	9.02855	21.6732	56.0174
1-S	1.37524	3.33812	0	3.36513	2.59014	18.4134
2-S	0	2.95234	0.49647	2.70207	2.83599	11.9198
3-S	8.20537	9.87485	0	10.563	0.752176	8.8804
4-S	4.0486	6.3489	0.6976	5.8074	2.4295	0
5-S	22.2408	22.4645	4.3698	21.9758	11.7949	0

(d) Securities with 15-30 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	186.038	165.547	122.105	175.892	651.781
1-S	0	13.0014	0.866516	7.25382	9.89833	47.6846
2-S	3.30803	3.84265	0	3.98586	2.40497	0.2213
3-S	8.8722	11.3749	1.4767	10.2931	7.5218	0
4-S	0	5.64236	9.57294	4.08763	5.1676	8.8835
5-S	28.6098	29.8382	15.0099	29.0735	28.1981	0

Table 14: Percentage Penalty of Yield Errors by Tenor Category

(a) Securities with 0–3 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	74.359	39.497	127.326	6.80338	121.826
1-S	8.74013	70.8237	12.2756	115.279	0	68.1549
2-S	29.4497	55.5827	2.29091	92.9119	0	39.0216
3-S	51.4088	91.0952	0	110.235	12.8384	17.0350
4-S	51.9255	100.933	0	124.876	9.49206	5.7036
5-S	87.8921	135.158	0	143.321	35.4383	5.9479

(b) Securities with 3–7 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	1.81411	4.21152	7.01664	5.64088	167.658
1-S	9.40314	15.4365	0	13.098	13.5257	24.7963
2-S	0.557036	4.35391	0	3.27253	6.87253	6.7627
3-S	22.9409	22.8862	0	21.717	25.1636	6.9973
4-S	16.8249	18.3662	0.8764	17.8855	18.1136	0
5-S	55.9464	50.4067	16.6375	50.8926	45.1514	0

(c) Securities with 7-15 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	5.4816	7.04332	5.83323	8.37647	40.2823
1-S	9.57304	11.7522	0	12.0444	10.1491	24.4170
2-S	12.4069	13.4718	0	13.0147	13.592	15.5446
3-S	13.7	13.7493	0	13.4013	11.3509	22.9136
4-S	12.1252	13.4039	0	13.4051	8.74673	7.4138
5-S	40.4684	42.0223	0	42.0712	25.764	1.6597

(d) Securities with 15-30 years to maturity

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	175.613	147.138	108.343	166.048	600.569
1-S	0	12.7732	1.90469	7.26831	11.3745	43.0561
2-S	1.1856	1.4381	1.7257	1.8457	0	0.3285
3-S	12.8355	16.0327	8.3290	13.5471	10.9667	0
4-S	0	5.61639	6.28417	4.04047	5.86693	7.3316
5-S	27.3980	29.0926	15.6638	28.1250	28.0515	0

Table 15: Median MAPE for Treasury Bills

(a) Price Errors

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.00880527**	0.020914	0.0177765	0.0276213	0.00976778	0.052031
1-S	0.015862	0.0313875	0.0203268	0.0353665	**0.0145105**	0.048836
2-S	0.0250686	0.0428039	0.0210726	0.0432806	**0.0196475**	0.047613
3-S	0.0319612	0.0566628	**0.0228137**	0.0552664	0.026548	0.049468
4-S	0.0432801	0.066793	**0.0250907**	0.0661124	0.0308675	0.058193
5-S	0.0638737	0.0809441	**0.0244604**	0.0856178	0.0491018	0.051764

(b) Yield Errors

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	**0.796772**	1.60218	1.49004	2.14318	0.961537	4.80937
1-S	1.44603	2.5043	1.28598	3.25211	**1.18681**	3.93438
2-S	2.28349	3.40034	1.63677	3.81861	**1.58425**	4.59787
3-S	3.30232	4.18825	**1.52563**	5.13248	2.07375	4.26124
4-S	4.33967	5.31338	**2.13098**	6.08526	2.75941	4.91517
5-S	5.964	7.15206	**2.92778**	7.81706	3.69184	4.74682

Note: minimum entry for each horizon displayed in **bold**.

Table 16: Percentage Penalty for Treasury Bill Prediction

(a) Price Errors

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	137.517	101.884	213.691	10.9311	490.904
1-S	9.31417	116.309	40.0836	143.731	0	236.557
2-S	27.5919	117.86	7.2537	120.286	0	142.337
3-S	40.097	148.372	0	142.251	16.3687	116.837
4-S	72.4947	166.207	0	163.494	23.0239	131.931
5-S	161.131	230.919	0	250.026	100.74	111.623

(b) Yield Errors

model	DF1	LDFGCV	FGCV	DFGCV	DF2	NS
Quote	0	101.084	87.0098	168.982	20.679	503.607
1-S	21.8413	111.011	8.35547	174.02	0	231.509
2-S	44.1369	114.634	3.31499	141.036	0	190.224
3-S	116.456	174.525	0	236.417	35.9273	179.310
4-S	103.646	149.339	0	185.561	29.4899	130.653
5-S	103.704	144.282	0	166.996	26.0966	62.130

References

[1] Baum, C.F. and C. F. Thies, 1992, On the construction of monthly term structures of U.S. interest rates, *Computer Science in Economics and Management*, 5, 221-46.

[2] Bliss, Robert R., 1997, Testing term structure estimation methods, *Advances in Futures and Options Research*, 9, 197-231.

[3] de Boor, C., 1978, *A practical guide to splines*, Applied Mathematical Sciences, Vol. 27. Springer, New York.

[4] Brennan, M.J., and E.S. Schwartz, 1979, A continuous time approach to the pricing of bonds, *Journal of Banking and Finance* 3, 133-55.

[5] Carleton, W. and I. Cooper, 1976, Estimation and uses of the term structure of interest rates, *Journal of Finance* 31, 1067-83.

[6] Cecchetti, S., 1988, The case of negative nominal interest rates: new estimates of the term structure of interest rates during the Great Depression, *Journal of Political Economy* 96(6), 1111-41.

[7] Cox, J., J. Ingersoll, and S. Ross, 1985a, An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363-84.

[8] __________, 1985b, A theory of the term structure of interest rates, *Econometrica* 53, 385-467.

[9] Craven, P. and G. Wahba, 1979, Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of Generalized Cross-Validation, *Numerische Mathematik*, 31, 377-403.

[10] Duffie, Darrell, 1996, Special Repo Rates, *Journal of Finance,* 51:2, 493-526.

[11] Fisher, M.E., Nychka, D., and D. Zervos, 1995, Fitting the term structure of interest rates with cubic splines, FEDS discussion paper 95-1, Federal Reserve Board of Governors.

[12] Fisher, M.E. and D. Zervos, 1996, Yield Curve, in *Computational Economics and Finance: Modeling and Analysis with* MATHEMATICA, H. Varian, ed. Springer-Verlag, New York.

[13] Gilles, Christian, 1997, private communication with the second author.

[14] McCulloch, J.H., 1971, Measuring the term structure of interest rates, *Journal of Business*, 44, 19-31.

[15] __________, 1975, The tax-adjusted yield curve, *Journal of Finance*, 30, 811-30.

[16] Nelson, C.R. and A.F. Siegel, 1987, Parsimonious Modeling of Yield Curves, *Journal of Business*, 60, 473-89.

[17] Shea, G.S., 1984, Pitfalls in smoothing interest rate term structure data, *Journal of Financial and Quantitative Analysis*, 19 (3), 253-69.

[18] Vasicek, O.A., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177-88.