Forthcoming in *Advances in Futures and Options Research*

TESTING TERM STRUCTURE ESTIMATION METHODS

Robert R. Bliss
Federal Reserve Bank of Atlanta
Working Paper 96-12a
November 1996

JEL classification: E34, G12, C52

The views expressed here are those of the author and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. The author thanks George Constantinides, George Easton, Eugene Fama, Wayne Ferson, Kenneth French, Merton Miller, Ehud Ronn, Ruey Tsay, Dan Waggoner and seminar participants at the University of Chicago and Indiana University for their comments and suggestions. Any remaining errors and omissions are the author's responsibility.

Please address questions of substance to Robert R. Bliss, Research Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713, 404/521-8757, 404/521-8810 (Fax), rbliss@frbatlanta.org.

Abstract: This paper tests and compares five distinct methods for estimating the term structure. The Unsmoothed Fama-Bliss method is an iterative method by which the discount rate function is built up by computing the forward rate necessary to price successively longer maturity bonds. The Smoothed Fama-Bliss "smooths out" these discount rates by fitting an approximating function to the "unsmoothed" rates. The McCulloch method fits a cubic spline to the discount function using an implicit smoothness penalty, while the Fisher-Nychka-Zervos method fits a cubic spline to the forward rate function and makes the smoothness penalty explicit. Lastly, the Extended Nelson-Siegel method, introduced in this paper, fits an exponential approximation of the discount rate function directly to bond prices.

The tests demonstrate the dangers of in-sample goodness-of-fit as the sole criterion for judging term structure estimation methods. A series of residual analysis tests are introduced to detect misspecification of the underlying pricing equation relating the term structure to bond prices. These tests establish the presence of unspecified, but nonetheless systematic, omitted factors in the prices of long maturity notes and bonds.

Comparisons of the five term structure estimation methods using these parametric and non-parametric tests finds that the Unsmoothed Fama-Bliss does best overall. Differences with some alternatives may not be economically significant given the much larger number of parameters this method estimates. Users seeking a parsimonious representation of the term structure should consider either the Smoothed Fama-Bliss or the Extended Nelson-Siegel methods. One method was found to be unacceptable. The Fisher-Nychka-Zervos cubic spline method performs poorly relative to the alternatives, both in- and out-of-sample. Furthermore, it systematically misprices short maturity issues and suffers from instability in the estimated term structure.

1 Introduction

The term structure of interest rates[1] is a concept central to economic and financial theory and the pricing of interest rate contingent claims. Interest rate models based on the short rate, such as Cox, Ingersoll, and Ross (1985) or Vasicek (1977), imply structures for the shape of the term structure. In practice, however, applications of interest rate models to pricing derivatives first calibrate the parameters of the model to fit the term structure.[2] Other models, for instance Ho and Lee (1986) and Heath, Jarrow, and Morton (1990 and 1992), take the entire term structure as given and provide structure for its evolution. These models are then applied to derivatives pricing, for instance Amin and Morton (1994).

In all these cases an estimate of the term structure of interest rates is the necessary starting point for testing these models or applying them to pricing and hedging interest rate derivatives. Unfortunately, rarely in financial economics is the contrast between theory and reality more starkly visible than in the study of term structure estimation. Efforts to reconcile the apparent conflicts have given rise to a rich literature on market frictions. This literature has generally consisted of hypothesis tests of individual models. Little attention has been paid to resolving conflicting models or evaluating fitted-price errors (residual analysis). This paper seeks to address these gaps.

1.1 Objectives

The focus of this paper is methodological. We first propose and implement a procedure for comparing alternate term structure estimation methods, emphasizing out-of-sample performance. We then propose and apply a series of tests to detect misspecification of the pricing equation by examining the randomness of the fitted-price errors. To illustrate our proposed

[1]The term "term structure" is somewhat loose and includes the discount function, the discount rate function (zero-coupon yield curve) and the forward rate curve. Since each is a transformation of others, "term structure" may be used in place of any of them, except where specificity is necessary.

[2]This is done by permitting the drift parameter, and perhaps others, to evolve deterministically. The time-varying parameters are then selected so as to price pure discount bonds, that is "match the term structure." See, for example, Black, Dermon, and Toy (1990), Hull and White (1993), and Bliss and Ronn (1995). These time-varying parameters are *ad hoc* and outside the context of the underlying model and no structure is imposed on their evolution, but it is considered by many practitioners an essential element of any model that it be able to price vanilla bonds.

procedures, we use four term structure estimation methods, detailed in the next section, based on the simplest possible pricing equation without tax-effects or other refinements.

Past research has used full sample estimates of the term structure to test hypotheses. This tends to reward more highly parametrized models.[3] We will show that in-sample results give a distorted view of performance. There is a clear danger of over-fitting the data. The solution to this problem is to use out-of-sample tests when evaluating pricing equations or estimation methods.

Past research has paid scant attention to residual (fitted-price error) analysis as a means of detecting potential problems.[4] A pricing model, such as equation (2) below, is used under the assumption (or to test the assumption) that it captures all the systematic information embedded in bond prices and that remaining differences between the theoretical and actual prices are noise. A direct test of whether this is so is to look for systematic relations in the fitted-price errors. We do this with a series of three tests:

1. We examine the behavior of fitted-price errors of individual bonds from one period to the next by examining the transition matrix for fitted-price errors classified as positive, zero, or negative. If the errors are indeed noise, the transition probabilities should reflect the frequency of positive, zero, and negative errors in the overall population.

2. To test whether our time series independence results are estimation method-specific or the result of a misspecified pricing equation (which is common to all estimation methods used here), we examine the coincidence matrices of positive, zero or negative errors across estimation methods. It may be the case that all estimation methods are imperfect and each introduces some method-specific errors. It is unlikely that these errors will be similar across estimation methods unless the (common) pricing equation itself is the culprit.

3. Lastly, we regress the fitted-price errors against variables that various theories (tax-clientele, tax-timing, liquidity premia) suggest might be relevant to the pricing process.

[3] See Table 1 and the related discussion below.

[4] Nelson and Siegel (1987) found "...some evidence of issue effects since large residuals for a particular issue show some tendency to persist from one quote sheet to the next." They did not pursue the matter further.

This will allow us to see if fitted-price errors are related to contemporaneous factors and whether these relations are method-specific and/or constant through time.

We find that for intermediate and long maturities there is clear evidence of unspecified, estimation method-independent factors which have been omitted from the bond pricing equation which nonetheless impact the pricing of bonds (henceforth, "omitted pricing factors" for the sake of brevity).

Our focus in this paper is not on which estimation method "wins the horse race" but on how one should judge the race. We wish to establish a methodology to be used in the future to test whether more complex pricing relations fully capture the systematic behavior of bond prices. The advantage of the tests developed here over direct tests of specific hypothesized pricing factors is that these methodologies test for unspecified effects. The two methodologies, tests of comparative ability to fit bond prices and tests for potential misspecification, provide a firm basis for evaluating future empirical research directed at modeling and testing term structure hypotheses.

No effort is made to resolve the issues of tax-clientele, tax-timing, liquidity effects, etc. While these effects (and unknown others) may exist, we cannot begin to study more complex hypotheses until we understand the empirical behavior of simpler models and the methods used to test them.

1.2 Background

The underlying theory is straightforward enough. The Fundamental Theorem of Asset Pricing [see Dybvig and Ross (1989), for example] implies that in a world of certain cash flows, c_m, and frictionless markets, absence of arbitrage is equivalent to the existence of a linear pricing rule, $\delta_m > 0 \ \ \forall m$, such that

$$P = \sum_{m=1}^{M} c_m \delta_m. \tag{1}$$

If markets are incomplete, there exist multiple sets of δ_m which satisfy this equation. In the term structure literature, δ_m, called the "discount function," is usually transformed into a discount rate curve by $r(m) \equiv -\log(\delta_m)/m$. Dermody and Prisman (1988) have argued that, because prices are quoted with a spread, multiple discount functions satisfying $P^{Bid} \leq \sum_{m=1}^{M} c_m \delta_m \leq P^{Ask}$ will be the rule.

Unfortunately, we need only look at the quotations of Treasury STRIPS to observe apparent violations of the Law of One Price,[5] implying that the frictionless markets assumption underlying equation (1) is not supported by the facts. Attempts to fit discount functions to sets of government bond prices find that no discount function, δ_m, exists to exactly price all bonds, even when bid-ask spreads are accounted for. The STRIPS quotes provide a sufficient counter-example. Other studies have used linear programming to find apparent arbitrage opportunities between individual bonds and portfolios of other bonds.[6]

Rather than accept this and other evidence as proof that arbitrage opportunities exist, the term structure estimation literature has sought to explain these phenomena in terms of frictions such as taxes, short sale constraints, and liquidity premia.

Estimating a term structure requires three decisions:

1. A pricing function relating bond prices, P_j, to the discount rate function, $r(m)$, via promised coupon and principal payments, c_{jm}, occurring at $m = 1, \ldots, M_j$ (the maturity of the bond) and perhaps other factors such as tax rates.

2. A functional form to be used to approximate the discount rate function, $r(m)$, or discount function, $\delta(m)$.

3. An econometric method for estimating the parameters of the term structure function.

All these decisions affect, more or less, the estimates obtained for the term structure. Tests of term structure hypotheses are, therefore, necessarily joint tests of the hypothesis in question (e.g., forward rates predict future spot rates), the pricing equation assumed, and the estimation method used (e.g., cubic splines estimated using ordinary least squares). For conciseness, in this paper "estimation method" will refer to a combination of an approximating function and the associated parameter estimation technique.

[5]It is common to find principal and coupon STRIPS maturing at the same time, both representing presumably identical default-free cash flows, quoted with non-overlapping bid-ask spreads. Evidence suggests that these disparities are real (not quotation errors). This phenomena is studied in Daves and Ehrhardt (1992). Similar examples involving disparities between short-term notes and bills have been studied by Kamara (1990) and Amihud and Mendelson (1991).

[6]Examples include Schaefer (1982), Ronn (1987), and Cornell and Shapiro (1989).

1.2.1 Pricing Function

The simplest pricing function, appropriate to a world without taxes, embedded options, or other frictions, is just the present value of the promised cash flows:

$$P_j = \sum_{m=1}^{M_j} c_{jm} e^{-r(m)m}. \tag{2}$$

Unfortunately, real markets (the source of our data) are not frictionless, and equation (2) does not hold exactly for any reasonable $r(m)$.[7] In practice we use, not an exact pricing relation such as equation (2), but an inexact relation such as

$$P_j = f\left[c_{jm}, r(m)\right] + \epsilon_j,$$

where $f(\cdot)$ captures all that we know, or wish to assume, about how bonds are priced, and $r(m)$ is fitted to minimize some function of the ϵ_j, which in turn should be random. Predictable ϵ_j's suggest that there is additional available information that could be included in $f(\cdot)$.

The definition of $f(\cdot)$ is the subject of ongoing studies of various hypothesized "pricing factors" which argue that the nominal, before-tax, promised cash flows over the life of the bond are not the relevant or sole consideration in pricing bonds. For example, the frequently studied tax-clientele effect has the effect of converting the before-tax cash flows, c_{jm}, to clientele-specific after-tax cash flows, $c_{jm}^{\tau_i,\tau_g}$, where $\{\tau_i, \tau_g\}$ are the ordinary income and capital gains tax rates. The after-tax cash flows, $c_{jm}^{\tau_i,\tau_g}$, replace c_{jm} in equation (2), which then becomes the basis for term structure estimation. The resulting term structure is the after-tax term structure for the clientele $\{\tau_i, \tau_g\}$ and is a function of those parameters.

More sophisticated approaches first screen the sample of available bonds, discarding those that would not be rationally held by the tax-clientele defined by $\{\tau_i, \tau_g\}$. Other potential

[7] It is usually argued that $r(m)$ should be a "smooth" function to avoid implicit discontinuities in the preference functions of market participants.

factors which might affect prices, and therefore the pricing equation, are tax-timing options and liquidity premia.[8]

1.2.2 Approximating Function

After deciding on the appropriate pricing function, the next step is to decide on the functional form to be used to approximate the discount rate function, $r(m)$, or the discount function, $\delta(m)$. It is not possible to estimate the value of the term structure at each possible horizon as the number of cash flow points will usually exceed the number of available bonds. In this case, the term structure would be under-identified. The usual practice is to select an approximating function and then to estimate the parameters of this function. Examples of approximating functions include polynomials [Chambers, Carleton, and Waldman (1984)], cubic splines [McCulloch (1975), Litzenberger and Rolfo (1984), and Fisher, Nychka and Zervos (1995)], step functions [Ronn (1987), Coleman and Fisher (1987)], piecewise linear [Fama and Bliss (1987)], and exponential forms [Nelson and Siegel (1987)].

1.2.3 Estimation Method

Lastly, the method for estimating the parameters of the approximating function must be selected. Methods used in the past include weighted least squares [McCulloch (1975)], maximum likelihood [Litzenberger and Rolfo (1984)], linear programming [Ronn (1987)], and iterative extraction [Fama and Bliss (1987)]. Related decisions include error weighting functions and how to handle the bid-ask spread (usually by collapsing the bid and ask quotes into a single price by taking their mean).

[8]See McCulloch (1975), Schaefer (1982), Litzenberger and Rolfo (1984), and Jordan (1984) for discussions and tests of tax-effects and tax-clientele. See Constantinides and Ingersoll (1984) for the theory of tax-timing options. See Kamara (1990), Amihud and Mendelson (1991), and Beim (1992) for tests of liquidity effects.

2 Methodology

In this paper, the available issues each month are sorted by maturity and then alternately divided into estimation and hold-out subsamples.[9] Because of the construction of the subsamples, the number and mix of issues (and hence weights) are not substantially different in the two subsamples. Term structure parameters are estimated each month using the estimation subsample from that month's available observations. The estimated term structure is then used to compute fitted prices and errors for bonds in the hold-out subsample for that month. The process is repeated for each of the 312 months of data using each of the several estimation methods tested. The results are summarized in a time-series of monthly statistics which form the basis for our comparative performance tests.

Since the assignment of subsamples to estimation and hold-out roles is arbitrary, the tests were repeated with the assignments reversed. The few cases which are not robust to reversing the subsamples are noted in the discussion of the empirical results. Stratified-random and spread-based partitions were also tried.[10] The results under these alternative subsampling schemes were similar to those reported in this paper.

2.1 Assumptions

In this paper we assume that riskfree bond prices are determined by equation (2). The assumptions underlying this pricing equation represent a gross simplification of the actual market for treasury securities. They also deliberately ignore the evidence for pricing factors such as tax-timing options, tax-clientele, liquidity premia and data quality. Although these issues have been examined individually, modeling their combined effects is extremely difficult, if not impossible. Introducing these factors into the pricing equation is beyond the scope of this paper and unrelated to the objective of testing estimation methods in the face of data noise and unspecified omitted factors in the pricing function.

[9] The longest issue in the hold-out subsample is discarded if its maturity exceeds that of the longest issue in the estimation subsample. This avoids using extrapolated term structures in the tests. Extrapolated splines are exceedingly poor at pricing longer maturity bonds; other functional forms, less so.

[10] These consisted of ordering the bonds by maturity, breaking them into groups of ten and then dividing each of these groups into five estimation subsample issues and five hold-out issues, either randomly or on the basis of spreads (with the five lowest spread issues being assigned to the estimation subsample).

We further assume that:

- The bid and asked quotes define a range of equally likely "true" prices for the bond.
- Quoted prices are observed with error; these may include random noise in the quotes. These errors are assumed to be cross-sectionally independent.

2.2 Error Definition and Weighting

Previous work has generally used some form of regression to estimate term structure parameters. The effects of this decision have been twofold: first, to necessitate the reduction of the quoted bid and asked prices to a single price (usually the mean) for econometric expediency, and second, to implicitly use a squared fitted-price error loss function, perhaps weighted in some way.

In this paper we distinguish between the econometrically necessary and the economically meaningful and recognize that the criteria for judging an estimated term structure may differ from that which is implicitly used to estimate it. Although it is desirable to use the economic criteria for estimation, it is not always feasible to do so.

The use of a single mean price is an *ad hoc* simplification of the data. In testing estimated term structures in this paper, we define a fitted-price error to be non-zero only if the fitted price lies outside the bid-ask spread as follows:

$$\epsilon = \begin{cases} P^A - \hat{P} & \text{if } \hat{P} > P^A \\ P^B - \hat{P} & \text{if } \hat{P} < P^B \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where P^B, P^A, and $\hat{P}$ are the bid and asked quotes and fitted prices, respectively. One of the estimation methods used in this paper (the Extended Nelson-Siegel, see below) permits utilizing this error definition in the estimation procedure as well.

Dealers (who are assumed to be risk-neutral) can, theoretically, take advantage of either over- or under-priced securities. We therefore use an absolute fitted-price error criterion. There is usually no particular economic rationale for utilizing a squared error loss function other than econometric convenience.

In combining fitted-price error across maturities in a single month, we weight the errors by the inverse of the duration of the issue. This weighting best captures the observed heteroscedasticity of fitted-price errors and the theoretical relation between prices and interest rate levels.[11] A second performance statistic we examine is the "hit rate" or percentage of issues with zero fitted-price errors. Our goodness-of-fit statistic, the monthly duration-weighted mean absolute error (WMAE), which includes both zero-error and "mispriced" bonds, combines the hit rate with the magnitude of the mispricing of the issues not correctly priced.

2.3 Data

The data consist of month-end price quotes for treasury issues for the period January 1970 through December 1995. Data are taken from the CRSP Government Bonds files. The following types of bonds were eliminated from the sample:

- Those with option features:
  - callable and flower bonds
- Those with special liquidity problems:
  - notes and bonds under one year to maturity
  - bills under one month to maturity.

As Figure 1 demonstrates, after eliminating the above issues, bonds with maturities beyond five years are sometimes sparse. All tests were repeated using two data sets: one consisting only of issues with up to five years to maturity (denoted the Short Data Set) and the other with all eligible issues (denoted the Long Data Set).

[11] In tests of various weighting schemes (not reported herein), it was found that duration- and maturity-weighting during estimation produced the highest proportion of "correctly priced" out-of-sample bonds. Equal weighting and spread weighting were noticeably less successful. Macaulay duration provides a theoretical relation between price changes and changes in yields to maturity, which in turn are averages of the discount yields we are estimating. Tests of duration models [e.g., Nelson and Schaefer (1983)] have found that simple (Macaulay) duration captures most of the interest rate risk and thus supports using duration weighting rather than maturity weighting, though the practical differences are minimal.

2.4 Models Tested

Five methods were used to estimate the term structure using the same pricing equation. Three estimation methods, the Unsmoothed Fama-Bliss, and the McCulloch and Fisher-Nychka-Zervos cubic splines, were taken from the literature. The remaining two estimation methods were developed for this research. These five estimation methods represent a broad variety of approximating functions and estimation methods. The use of broadly differing estimation methods is useful for distinguishing method-specific results from those due to the common pricing equations and data sets.

- **Unsmoothed Fama-Bliss**

  The Unsmoothed Fama-Bliss [see Fama-Bliss (1987)] method is an iterative method of forward rate extraction.[12] The discount rate function is extended each step by computing the forward rate necessary to price successively longer maturity bonds given the discount rate function fitted to the previously included issues. A series of filters is employed to throw out suspicious quotes. The resulting discount rate function exactly prices the included bonds. The mean price is used. Since fitted-price errors of bonds remaining after the filtering step are all zero, weighting is irrelevant. The resulting discount rate function is piecewise linear (jagged) with the number of parameters equal to the number of included issues.

- **McCulloch Cubic Spline**

  The McCulloch (1975) method uses a cubic spline to approximate the discount (not discount rate) function.[13] The spline is estimated using ordinary least squares. The mean price is used as the dependent variable and the fitted-price errors are weighted by the inverse of the spread. The number of parameters is moderate; however, the functional form leads to discount rates that tend to positive or negative infinity when extrapolated.

[12] Fama and Bliss (1987) appear to have been the first to publish an implementation of this method. Several earlier authors have proposed its use, and the method is known on Wall Street as the "bootstrap" for obvious reasons. The use of filters to limit the adverse effects of suspicious quotes was original. In this paper the filter is applied only to the estimation subsample.

[13] The author is indebted to Professor McCulloch for providing the programs to implement this estimation method. See McCulloch (1975) for the details of this method.

- **Fisher-Nychka-Zervos Cubic Spline**

  Fisher, et al. (1995) fit a cubic spline so as to minimize the following function:

  $$\min_{h(t)} \left[\sum_{i=1}^{N} (P_i - \hat{P}_i)^2 + \lambda \int_0^{T_{max}} h''(t)^2 dt \right],$$

  where $h(t)$ is the function used to compute the fitted bond prices, $\hat{P}_i$, and may be the discount, discount rate, or forward rate function. Following the recommendation of Fisher, et al., we estimate the forward rate function.[14] The function

  $$\lambda \int_0^{T_{max}} h''(t)^2 dt$$

  is a (un)smoothness penalty. The parameter λ controls the trade-off between smoothness and goodness of fit and is itself selected as part of the estimation process using cross-validation in a theoretically optimal manner. An additional "tuning parameter" enters into the cross-validation and also effects the trade-off between goodness of fit and smoothness. This parameter is preset to the value used by Fisher, et al. (1995), that is 2.0.

- **Extended Nelson-Siegel**

  The Extended Nelson-Siegel method introduces a new estimation method to fit a modified version of the approximating function developed by Nelson and Siegel (1987). The Extended Nelson-Siegel method brings together several desirable characteristics:

[14] The author is indebted to Mark Fisher for providing the Mathematica code for his method. The algorithm was translated into C++ by Dan Waggoner. This proved to be considerably more efficient both in terms of computational speed and freedom from convergence problems. In practice, the Fisher, et al. method is applied subjectively by its originators. The tuning parameter is re-adjusted and the estimation sample culled until the results seem reasonable. This unreported, *ad hoc* procedure does not constitute an algorithm that other researchers can replicate. We, therefore, test a straightforward application of their method using reported parameter values and compare the results to other similarly "off-the-shelf" methods.

accounting for bid-ask spreads, fitting the discount rate function directly to bond prices, and using the following asymptotically flat[15] approximating function:[16]

$$r(m) = \beta_0 + \beta_1 \left[\frac{1 - e^{-m/\tau_1}}{m/\tau_1} \right] + \beta_2 \left[\frac{1 - e^{-m/\tau_2}}{m/\tau_2} - e^{-m/\tau_2} \right]. \tag{4}$$

The parameters, $\boldsymbol{\phi} \equiv [\beta_0, \beta_1, \beta_2, \tau_1, \tau_2]$, are then estimated using the following non-linear, constrained optimization, estimation procedure:

$$\min_{\boldsymbol{\phi}} \sum_{i=1}^{N_t} (w_i \epsilon_i)^2,$$

where

$$\epsilon_i = \begin{cases} P_i^A - \hat{P}_i & \text{if } \hat{P}_i > P_i^A \\ P_i^B - \hat{P}_i & \text{if } \hat{P}_i < P_i^B \\ 0 & \text{otherwise} \end{cases}$$

and the weights, w_i, are defined in terms of Macaulay duration, d_i, measured in days

$$w_i = \frac{1/d_i}{\sum_{j=1}^{N_t} 1/d_j},$$

subject to

$$0 \leq r(m_{\text{min}})$$

$$0 \leq r(\infty)$$

---

[15]Livingston and Jain (1982) and Siegel and Nelson (1988) demonstrate that this property is appropriate if forward rates are finite.

[16]Nelson and Siegel (1987) state that

> "...if the instantaneous forward rate at maturity m is given by the solution to a second-order differential equation with real and unequal roots, we would have
>
> $$f(m) = \beta_0 + \beta_1 \cdot \exp(-m/\tau_1) + \beta_2 \cdot \exp(-m/\tau_2)$$
>
> where τ_1 and τ_2 are time constants associated with the equation ..."

Integrating this forward rate function produces the discount rate function used here. Nelson and Siegel found that for their sample of Treasury bills this equation is over-parametrized and therefore set $\tau_1 = \tau_2$. With the longer maturities used in this study, over-parametrization is not a problem and tests of the 4-parameter versus 5-parameter versions of the approximating function found that the 5-parameter version used here produces better results based on criteria similar to those developed in this paper.

and

$$\exp\left[-r(m_k)\, m_k\right] \geq \exp\left[-r(m_{k+1})\, m_{k+1}\right] \quad \forall m_k < m_{\max}.$$

The constraints ensure that the discount function is non-increasing (non-negative forward rates) and that the short and long ends of the discount rate function are positive. Only the long-rate constraint proved binding, and then only rarely.

- **Smoothed Fama-Bliss**

  The Smoothed Fama-Bliss method attempts to "smooth out" the unsmoothed Fama-Bliss discount rates by fitting an approximating function through them. The function used is the same as that used in the Extended Nelson-Siegel method, except here it is fit to the previously estimated discount rates. Fitted-discount rate errors are equally weighted and parameters are estimated using constrained non-linear optimization. While the functional form employed has five parameters, it is important to note that this method is a two-step process (first estimate unsmoothed discount rates, then smooth them) so the effective parametrization is ambiguous.

It is worth noting that the only commonality between the Extended Nelson-Siegel and the Smoothed Fama-Bliss is the approximating function. The methods used to fit the function differ radically; in one case, the function parameters are estimated directly from bond prices, in the other from discount rates produced by iterative extraction method.

3 Empirical Results

The empirical results begin with a descriptive overview of the in- and out-of-sample performance measures across estimation methods for the two data sets (Short and Long) and for various subsets and subperiods. A formal test of model comparisons is conducted using the Friedman Rank Comparison Test [Friedman (1937)]. These cross-method comparisons are followed by time-series tests of the fitted-price errors, where the results from the several estimation methods are used to determine whether the time-series patterns are method-specific or indicative of misspecification of the assumed pricing function.

3.1 In-Sample Results

We begin by examining in Table 1 the ability of the five estimation methods to fit bond prices and (parenthetically) the impact of increasing parametrization on the ability of estimated term structures to explain the variation in quoted prices.[17] Two variants of the Extended Nelson-Siegel function are added so as to include methods with even fewer parameters than our five candidates. They are a flat discount rate function and a simple monotonic function, asymptotically flat and either upward or downward sloping.

From Table 1 we can see that for Treasury bills (up to one year maturity) the mean absolute fitted-price errors (MAEs) are economically small for four of the five methods used in this paper. The magnitude of errors generally increases with maturity and above five years become serious regardless of the method employed, excepting perhaps the Unsmoothed Fama-Bliss. This suggests that the "present value of before-tax promised cash flows" pricing equation does reasonably well out to perhaps five years maturity. Differences across estimation methods appear to be small. However, none of the "smooth" methods is particularly successful at estimating term structures which can price long notes and bonds, at least not in conjunction with the pricing equation used here.

The Fisher, et al. cubic spline does poorly at fitting short maturities. This problem and the stability problem noted below were confirmed using the original Mathematica code provided by Mark Fisher. Lowering the tuning parameter does improve the in-sample fit. However, the more serious problem lies in the smoothness penalty itself, rather than the details of the implementation. The smoothness penalty is applied uniformly across the term structure.[18] In fact, greater flexibility is appropriate at shorter maturities where there is more true curvature in the term structure as casual inspection of Treasury bill yields shows. The result is that the Fisher, et al. method sometimes systematically under-prices all short

[17] *Ceteris paribus*, more parameters will usually produce better in-sample fit. When combining different functional forms however, this need not necessarily hold. Imagine fitting a 2-parameter exponential function (such as the monotonic function in Table 1 with $\beta_0 = 0$) to the discount price function, which is approximately exponential. Then imagine trying to fit a 3-parameter polynomial. We would expect the polynomial to do worse, despite the larger number of parameters.

[18] A possible improvement to the Fisher, et al. method would be to make the penalty maturity-dependent, e.g., $\int_0^{T_{max}} \lambda_t h''(t)^2 dt$, where λ_t is increasing in t.

maturity issues in a given sample. The other four methods all permit greater flexibility at the short end of the term structure.

As one would expect, the results also show that as the number of parameters increases, the different estimation methods produce fitted prices that are better able to explain the observed prices as shown by decreasing MAEs. This is true for all maturity subintervals. However, improvements obtained through additional parameters quickly decline. The most highly parametrized model, the Unsmoothed Fama-Bliss, produces only a marginal improvement in MAE over that obtained in moving from one to three parameters. Only for the longer maturities do the more highly parametrized two cubic splines and Unsmoothed Fama-Bliss methods show a non-trivial improvement over the more parsimonious (and asymptotically flat) Smoothed Fama-Bliss and Extended Nelson-Siegel methods. But these long maturities are where we expect to see the greatest effects of omitted pricing factors and quotation errors. Therefore, these "improvements" should be viewed with caution as they may be the result of over-fitting the long end of the term structure to the relatively sparse set of long bonds.

3.2 Out-of-Sample Results

To provide a basis for comparison of the out-of-sample results using our proposed performance measures, we first examine in-sample performance. The top panel of Table 2 presents the in-sample duration-weighted mean of the absolute fitted-price errors (WMAE) and hit rates. The Unsmoothed Fama-Bliss method clearly dominates. Differences among the Smoothed Fama-Bliss, Extended Nelson-Siegel, and McCulloch cubic spline methods are small, while the Fisher, et al. method does somewhat worse than the others.[19]

The picture changes when we turn to the out-of-sample issues presented in the middle panel of Table 2. Here the statistics are computed over the issues excluded from the subsample used to estimate the underlying term structures.

Since the objective functions minimized in estimating the term structures differ from the criteria we are using to evaluate them, it is not necessarily the case that the in-sample

[19] This latter result is due to a combination of the duration weighting of fitted-price errors together with the problem the Fisher, et al. method has with short-maturity (that is highly weighted) issues. In equally weighted comparisons, the Fisher, et al. method is on a par with the McCulloch cubic spline.

evaluation criteria statistics will be better than the out-of-sample statistics. However, the WMAE statistics are generally greater for the hold-out subsample than for the estimation subsample. The magnitude of the difference varies across methods. The Unsmoothed Fama-Bliss shows the largest increase in WMAE by far—to such a degree that its clear advantage in the in-sample comparisons all but disappears. The remaining methods show only slight increases in WMAE and occasionally decreases. This latter result (for the Smoothed Fama-Bliss) is sample specific. When the estimation and hold-out subsamples are reversed, all methods show increases in the WMAE statistic when moving from in-sample to out-of-sample. Again the Fisher, et al. estimation method does noticeably worse that the alternatives in WMAEs.

The Unsmoothed Fama-Bliss method, and thus the Smoothed Fama-Bliss method as well, effectively ignores bond quotes that may have large measurement errors. The other three methods do not benefit from this filter. Thus the term structures are estimated using slightly different subsamples. To test the effect this differential, the Extended Nelson-Siegel and two cubic splines were re-estimated using only those issues that survived the Fama-Bliss filtering process. The results are presented in the bottom panel of Table 2. There is little change in the out-of-sample WMAEs or hit rates for the Extended Nelson-Siegel and McCulloch cubic spline methods, indicating that they are robust to outliers in the data. The Fisher, et al. cubic spline does, however, show a marked out-of-sample performance improvement when the estimation subsample is first filtered. This suggests that the Fisher, et al. method is more subject to measurement error in the data than are the other methods.[20]

Figure 2 plots the time series of the monthly out-of-sample WMAE statistics for the several methods for the Long Data Set. The Short Data Set results are similar. Although variation between methods is observed, the general pattern is similar for all methods, although the scale for the Fisher, et al. plot is larger. This indicates that variations in the underlying data through time tend to dominate method-specific factors. For the 1970s, the WMAE statistics are generally much lower and show less variation than for the 1980s.

[20]Waggoner (1996) has investigated the robustness of these various estimation methods to small perturbations in prices. His results do not confirm the evidence in Table 2, suggesting that potential instability in the Fisher, et al. method is complex.

The McCulloch cubic spline shows a transitory increase in variation during the period 1973 through 1975 which is only faintly echoed by the Smoothed Fama-Bliss and Extended Nelson-Siegel. In the 1980s, there is a sharp rise in the level of the WMAE statistics, peaking in 1982, followed by a trough in 1984, a lesser peak in 1985, and a decline thereafter.

The increased levels and variance of the WMAEs in the 1980s is not due simply to the introduction of very long maturity bonds into the sample, for we see the same pattern of peaks and troughs in the Short Data Set plots where the composition of the sample is fairly uniform throughout the sample period. It is possible that the high levels of interest rates and interest rate volatility in the 1980s, and particularly in the early 1980s, made quotations of illiquid bonds less reliable. Another possibility is that the extremely high interest rates produced heterogeneous samples of deep discount and high premium bonds with the distributions shifting with levels of interest rates. Prices of long bonds have varied from under 60 to over 160 (par = 100). These extreme values would produce perceived price distortions under the pricing assumption, equation (2), if either tax-clientele or tax-timing were important.

Only the Unsmoothed Fama-Bliss shows substantial decreases in the hit rates when moving from the estimation subsample to the hold-out subsample. The Extended Nelson-Siegel has an *increase* in the hit rate, and this result persists when the hold-out and estimation subsamples are reversed as well as for other partitions of the data into subsamples. This latter result is strongest for short maturities (results not presented here), where the hit rates invariably increase, and weakest for the long maturities where the differences are small and the direction of change is unstable.

Differences among hit rates across estimation methods, like those among the WMAE statistics, have narrowed considerably from the in-sample results, although the Unsmoothed Fama-Bliss still has the highest hit rate regardless of the sample or maturity range used and the Extended Nelson-Siegel continues to have the lowest hit rate (or nearly so).

The time series of hit rates presented in Figure 3 again show similar patterns across methods. Differences between methods are less apparent. While the WMAEs peak in 1983 and decline thereafter, the hit rates decline throughout the 1980s and 1990s. This suggests an interaction of effects. The degree to which bonds are "correctly" priced relative to each other declines over time as indicated by the hit rates. However, the degree of mispricing for those bonds not correctly priced is highly variable in the 1980s. This apparent combination of effects suggests that modification of the pricing equation to account for one or another pricing factor will miss the interactions and not satisfactorily explain the observed prices.

For formal comparisons we use the Friedman non-parametric test applied to the monthly out-of-sample WMAE statistics and the original (unfiltered estimation subsamples).[21] The results of these tests are presented in Table 3. We can reject the hypothesis that there are no differences between the five methods. This result holds across both data sets for the entire sample period and in each subperiod.

For the Short Data Set, the Unsmoothed and Smoothed Fama-Bliss methods do better than the alternatives during most periods. The two are themselves never significantly far apart. Only in the 1979 to 1983 period does another method (the Extended Nelson-Siegel) do comparably well. The McCulloch and Fisher, et al. cubic splines always do worse than the two Fama-Bliss based methods, and the Fisher, et al. always does worse than the Extended Nelson-Siegel method.

The Long Data Set analysis is somewhat more mixed. In the 1983 to 1995 period, the Unsmoothed Fama-Bliss roundly bested the four alternatives and this result carries over to the overall 1970 to 1995 period. The McCulloch cubic spline bests the Smoothed Fama-Bliss in the later period and this also carries over to the overall period. Except in the 1983 to 1995 period, the Fisher, et al. method is again beaten by each of the alternatives.

Comparing these strong non-parametric results with the small average differences of the underlying out-of-sample WMAE statistics presented in Table 2, we see that the differences between method performance, while small, must nonetheless be fairly consistent through time.

3.3 Summary of Split Sample Results

The change in performance measures from in-sample to out-of-sample demonstrates the fallacy of relying on in-sample goodness-of-fit to evaluate term structure estimation methods.

[21]The Friedman test avoids strong distributional assumptions needed for parametric tests. The usual normality assumption is grossly violated by the WMAE statistics, as is independence across samples (methods).

Differences which were incontrovertible when examined in-sample become ambiguous when the out-of-sample results are viewed. Formal out-of-sample test results depend both on the sample used and subperiod and are, therefore, somewhat ambiguous.

These formal hypothesis tests suggest that the Unsmoothed Fama-Bliss method is best at fitting the underlying term structure and the Fisher, et al. method is worst. For maturities out to five years, smoothing the Fama-Bliss yields does not produce a significant degradation in the ability of the term structure to fit prices. Thus researchers who want a parsimonious representation of the term structure, for instance to study how levels, slope, and curvature change, may safely use the Smoothed Fama-Bliss, providing their analysis is restricted to short and intermediate maturities. The McCulloch and Extended Nelson-Siegel results are mixed (except against the Unsmoothed Fama-Bliss) and the Fisher, et al. method does poorly almost always.[22]

In general, either the Unsmoothed or Smoothed Fama-Bliss produce the best results, suggesting that filtering out "suspicious" issues before estimating the term structure is worthwhile, regardless of method. The consistent superiority of the Unsmoothed Fama-Bliss for long maturities suggests that either:

1. the approximating functions used are not sufficiently flexible to capture details of the long-maturity term structure, or

2. the high parametrization of the Unsmoothed Fama-Bliss term structure permits fitting to prices distorted by the presence of omitted pricing factors.

Since the result obtains out-of-sample, the latter explanation is plausible only if issues of like maturity have similar omitted pricing-factor effects. Because only very long maturity bonds are issued (20 years maturity or over) and their coupons adjust to prevailing long-term interest rates at time of issue, the sample of bonds with over 10 years to maturity will reflect the cyclical pattern of long-term interest rates. It is, therefore, not unreasonable to suppose that adjacent maturity long issues might have similar tax-related pricing effects. Certainly,

[22]Lowering the tuning parameter improves the in- and out-of-sample fit somewhat, but the Fisher, et al. method is still dominated by the two Fama-Bliss based methods. The problem of instability appears to be a result of the cross-validation proceedure and remains even for low values of the tuning parameter (0.5).

only the very longest maturity will be "on the run," with all others being illiquid to some degree.

If the Unsmoothed Fama-Bliss is over-fitting, the apparent goodness-of-fit will be an illusion and the estimated term structure will be distorted by the effects of the omitted factors.

3.4 Analysis of Fitted-Price Errors

The out-of-sample summary statistics presented above provide a basis for comparing term structure estimation methods. From these results we concluded that there are important differences between methods. We also note evidence of common variations over time in the ability of the several methods to fit the observed prices. This suggests the presence of data-specific factors which impact the pricing of bonds.

Under our maintained assumption that bond prices are determined solely by the present value of the before-tax cash flows, we would expect any fitted-price errors to be random and uncorrelated with other factors. We show this is definitely not the case by demonstrating that the fitted-price errors for specific bonds are correlated through time, correlated across estimation methods, and related to contemporaneous, observable factors.

The partition scheme used to split samples cannot guarantee that consecutive observations of any bond will both be in-sample or both out-of-sample, and it makes no sense to mix in- and out-of-sample fitted-price errors. To test the randomness of the fitted-price errors, we therefore use the full-sample estimated term structures to compute fitted prices. Using the five very different estimation methods, we look for patterns in the fitted-price errors. If the patterns are not method-specific, it suggests the problem is misspecification of the pricing equation which is common to all five methods—that is, this would be evidence of non-method-specific, unspecified, persistent, omitted factors in the pricing of bonds.

3.4.1 Transition Matrices

If the PV pricing equation is correctly specified and pricing errors are random noise, there should be no relation between the fitted-price error for a given bond in one period and the fitted-price error for that same bond in the next period. Since term structures for each month are estimated independently, there will be no interaction between successive observations as a result of errors in the estimated coefficients. Autocorrelated errors may be due to misspecification of the functional form or to the estimation method.[23]

We classify the fitted-price errors into positive, zero, and negative. If the underlying errors are random, the transformed variables will be also. Our test is to examine the transition matrices of consecutive categorized errors. Under the null hypothesis that the categorized errors are random, the classification of an error at time t should have no bearing on the observed classification at time $t + 1$. The chances of a positive error being followed by another positive error should be identical to the unconditional chance of a positive error. If the rows of the 3×3 transition matrices are not equal to the unconditional probabilities of positive, zero, and negative errors, it is evidence of non-randomness in the time series of fitted-price errors.

Notes and bonds differ in significant ways from bills. The notes and bonds market is less liquid and therefore liquidity effects may be more pronounced. The longer maturities give greater scope to potential tax-clientele and tax-timing effects. In addition, the number of notes and bond issues is small compared to the number of bills. Were they to be analyzed together, the results would be driven by the bills and the behavior of the notes and bonds would be obscured. We, therefore, separate the analysis of bills from that of notes and bonds. Table 4 presents the transition matrices for notes and bonds for all years and maturities (over one year). The results for each decade are similar to the overall results.

For the notes and bonds in Table 4, there is strong evidence that the classification of fitted-price errors persists through time. If a bond has a positive fitted-price error in one period, it is likely to also have a positive fitted-price error in the subsequent period; similarly, zero fitted-price errors tend to be followed by zeros and negatives by negatives. These results are clearly statistically significant.[24] The pattern of persistence in fitted-price errors for notes

[23]Of course, pricing errors might actually be correlated. However, this would suggest market inefficiency and, given the simplified nature of the assumed pricing equation, we are reluctant to draw such a conclusion.

[24]Given the large number of observations, any inequality in the rows will be significant at the usual levels.

and bonds does not depend on the estimation method used, even though the fraction of issues having positive, zero, or negative errors does.

Table 5 presents the same results for bills. Bills have a far lower fraction of issues "mispriced" than do notes and bonds regardless of method, excepting the Fisher, et al. method which shows a marked tendency to under-price bills. The Unsmoothed Fama-Bliss method shows little persistence in the fitted-price errors. The other methods show some persistence, though less than for notes and bonds.

For notes and bonds we conclude that there are method-independent persistent factors present in the prices which are not captured by the pricing equation. For bills we conclude that there is weaker evidence of persistence. The Fisher, et al. method shows a unique distribution of unconditional and conditional fitted-price errors for short, but not long, maturities that suggests a method-specific problem.

3.4.2 Coincidence of Errors

Having established the persistence of fitted-price errors generated by any given method, at least for notes and bonds, it remains to test whether the fitted-price errors are related *across* methods. We do this by once again classifying errors as positive, zero, and negative. We then examine the probability that an observation with a positive fitted-price error produced by one method will also have a positive fitted-price error under the other methods, etc.

If there is not a high degree of coincidence, we must conclude that the observed persistence is due to each method being misspecified in a different way, perhaps through shortcomings of its approximating function. If there is a high degree of coincidence in the contemporaneous classification of error across methods, it is evidence that errors arise from data-specific rather than method-specific factors. It is unlikely that all methods and functional forms will be misspecified in the same way since their only common elements are the pricing equation and the data.

For the sake of brevity we choose to use a single benchmark rather than present the 12 possible pairs of coincidence matrices. We restrict our analysis to the notes and bonds from the Long Data Set analyzed in the previous section, where the evidence of persistence was strongest. The comparisons will also be affected by the differences in the unconditional probabilities of positive, zero, and negative errors across methods. The Unsmoothed Fama-Bliss method is an unsatisfactory benchmark because its unconditional probabilities differ markedly from the others. We arbitrarily choose the McCulloch cubic spline method as the benchmark and present the coincidence results for errors between this and the four alternate methods in Table 6.

It is not surprising that the Unsmoothed Fama-Bliss method tends to produce zero fitted-price errors regardless of the classification of the errors produced by the McCulloch cubic spline. However, note that actual reversals of sign are comparatively rare: 20.0 percent of the McCulloch cubic spline positive-error issues also have positive errors under the Unsmoothed Fama-Bliss method, but only 1.5 percent have negative errors.

For the Smoothed Fama-Bliss and Extended Nelson-Siegel methods there is a strong coincidence of errors with those observed for the McCulloch cubic spline. Like the McCulloch cubic spline, these two have reasonably even unconditional distributions of errors (Table 2). The low frequency of fitted-price errors of opposite signs shows that the Fisher, et al. cubic spline method is most similar to the McCulloch cubic spline.

When combined with the results of the previous section, we conclude that fitted-price errors are largely method-independent and persistent. It is, therefore, very unlikely that the fitted-price errors are purely random.

3.4.3 Fitted-Price Error Regressions

We next examine the relation between the observed non-zero fitted-price errors and various characteristics of the bonds which might be related to the potential omitted pricing factors. We use the out-of-sample fitted-price errors from the Long Data Set.[25] Once again we distinguish between bills and notes and bonds. The explanatory variables were selected from those one might expect to be important under the various pricing hypotheses in the literature.

[25] In-sample fitted-price errors are biased towards zero, and we have no need to include all observations as we did in Section 3.4.1.

Due to the large number of observations, significance levels need to be treated with caution. Since we already know that there is a high degree of commonality among fitted-price errors produced by the five methods, we will ignore mixed results in the regressions and focus on those factors which enter strongly and with the same sign in all five regressions. For instance, in Table 7, while Term-to-Maturity is significant in the Extended Nelson-Siegel regression, it is much less so in the McCulloch cubic spline. It might, therefore, be imprudent to draw conclusions as to its relevance or lack of relevance in explaining non-zero fitted-price errors.

The "notes and bonds" regression results are presented in Table 7. The regressions are all highly significant, with p-values of less than 0.001. This does not reflect a high degree of explanatory power, but rather the large number of degrees of freedom. The R-squareds indicate that, overall, the explanatory variables explain only about 15 to 30 percent of the observed variation in fitted-price errors. On the other hand, we do not expect the explanatory power to be high. Under the various hypothesized pricing effects, the explanatory variables will enter into the pricing equation in complex, perhaps non-linear, ways. These regressions can only discover linear approximations to these relations.

In the bills regression, Table 8, the variables Term-to-Maturity (highly correlated with Discount) and Premium (always zero for bills) have been omitted. Since the tax-clientele and tax-timing options have little scope for impacting short-term prices, and the Treasury bills market is very liquid, we do not expect, and do not observe, any strong results in the bills regressions. The R-squareds are all less than 8 percent and the significance levels much lower than we observed in the notes and bonds regressions and inconsistent across methods. Only the time-trend is uniformly significant and is unrelated to the usual hypothesized omitted price factors. Because of the large number of degrees of freedom, the regressions are nonetheless statistically significant with p-values of less than 0.001.

The results of these regressions confirm the presence of omitted factors in the prices of notes and bonds previous established in the discussions of "persistence" (Section 3.4.1) and "coincidence" (Section 3.4.2) of the non-zero fitted-price errors. The regressors used here explain only a small fraction of the variation in fitted-price errors. Whether this is due to the variables incorrectly entering in a linear manner when the true relation is non-linear, or due to the omission of other variables, is unclear. Given the persistence of the fitted-price errors, it seems unlikely that the unexplained variance is purely noise. Lastly, there is little evidence of significant omitted factors in the prices of Treasury bills.

4 Conclusion

4.1 Summary

Five distinct methods of estimating the term structure of interest rates from prices of bills and coupon-bearing notes and bonds are compared using a series of parametric and non-parametric tests. A battery of residual analysis techniques are applied to detect evidence of general and method-specific misspecification.

Because the estimation methods tested differ greatly in their degree of parsimony, there is a potential problem of over-fitting the data for the more highly parametrized functional forms, particularly the Unsmoothed Fama-Bliss. The approach to the overfitting problem used here is to divide the sample of available bonds into estimation and hold-out subsamples, to use the estimation subsample to estimate the term structure, and to use the hold-out subsample to perform tests. This paper demonstrates that out-of-sample test results can differ markedly from in-sample results.

Two criteria for evaluating and comparing fitted term structures are used. The first is the duration-weighted mean of the absolute fitted-price errors. The second criterion for evaluating term structures is the “hit rate,” which is intuitive and has the advantage of not depending on a weighting scheme. These two criteria generally are in broad agreement when comparing estimation methods.

Fitted-price errors are measured relative to the bid-ask spread. The Extended Nelson-Siegel method introduces a general method for estimating term structure approximating functions using this error definition, thus avoiding the constraints inherent in regression-based approaches.

The results clearly demonstrate that over-fitting the data occurs with one of the methods—the Unsmoothed Fama-Bliss. Nonetheless, the Unsmoothed Fama-Bliss method is better than the others for fitting long-maturity term structures. This conclusion is tempered by the much higher parametrization of the Unsmoothed Fama-Bliss term structure. The parsimonious Smoothed Fama-Bliss and Extended Nelson-Siegel functional forms and the less parsimonious McCulloch cubic spline performed comparably to each other. Only at the longest maturities is there evidence that the flexibility of the cubic spline is needed. For short-to-intermediate maturities, differences among the four methods (excluding Fisher, et al.) are slight.

The Fisher, et al. estimation method is a dubious choice for estimating term structures. Both in-sample and out-of-sample, it performs poorly. It has systematic problems handling short maturities and is susceptible to measurement errors in the data. Correct measurement of the short end of the term structure is critical for computing short-horizon forward rates and forward prices. For instance, in the Heath-Jarrow-Morton model, the price of a bond next month is defined in terms of the forward price measured today for one month hence as well as the evolution of the state variables over that period. Computing the forward price requires a measure of the one month rate (in this case). The Fisher, et al. method is unsuitable for such an application. Similarly, the analysis of market expectations of future short-term interest rates relies on the term structure of forward rates. At short horizons the Fisher, et al. method is likely to give misleading results. While modifications to the penalty function may overcome some of these problems, in its current form the Fisher, et al. estimation method cannot be recommended.

This paper distinguishes between the pricing equation used to relate the fitted term structure to fitted prices and the estimation method used to arrive at the fitted term structure—focusing on the latter. This paper deliberately abstracts from the questions of tax and liquidity effects by assuming a "present value of before-tax promised cash flows" pricing equation throughout. The question of omitted pricing factors is addressed indirectly by examining common characteristics of the fitted-price errors across methods using three complementary tests: persistence, coincidence, and regression. This is the first study to use multiple methods to establish the method-independent characteristics of price errors in term structure fitting. It is also the first to search for unspecified omitted pricing factors, thus avoiding the need to identify and model specific effects.

We find that there is only weak evidence of non-random behavior in the Treasury bill errors. However, we do find convincing evidence of unspecified omitted pricing factors in the prices of notes and bonds with over one year to maturity. This evidence of misspecification is common across estimation methods as demonstrated by the as one would expect coincidence tests, persistant through time and related to contemporaneous observable factors. This suggests that misspecification arises form the underlying the pricing equation and not in the method used to estimate the term structures.

4.2 Implications for Future Research

While using out-of-sample tests is obvious, it is rarely done. This paper demonstrates that such methodological laxness may well lead to errors that are material. Researchers should report both in-sample and out-of-sample statistics to highlight possible over-fitting.

The estimated term structure depends on the estimation method chosen. This is a minor problem for short maturities, but may be significant for longer ones. One approach to this ambiguity is to use several estimates of the term structure to conduct studies to ensure that the results obtained are not estimation-method specific. For instance, Bliss and Ronn (1995) estimated implied volatilities of options embedded in callable Treasury bonds using Treasury STRIPS, the Unsmoothed Fama-Bliss and Extended Nelson-Siegel term structures. Differences in that case were immaterial.

The over-fitting evident in the Unsmoothed Fama-Bliss has implications for the application of interest rate options models, such as Heath-Jarrow-Morton, that are built on the estimated term structure, or those that are calibrated to the estimated term structure, such as Hull and White. If the term structure being used is "too flexible," the models will almost certainly be incorporating unwanted measurement error or idiosyncratic, bond-specific factors of no relevance to pricing securities in other markets. It remains to be seen whether it would be better to use more parsimonious estimates of the term structure for these applications.

Researchers testing for specific pricing factors (e.g., tax-clientele) should be cognizant of possible additional unspecified factors. To this end, papers testing for specific effects should include not only out-of-sample hypothesis tests, but should also test the remaining fitted-price errors for unspecified omitted pricing factors. The three residual analysis tests introduced here provide one set of such diagnostic tools.

In the end, however, term structure estimation is an art. The trade-off of fit against parsimony and judgment of what differences are material will always be subjective and depend on the problem at hand.

References

Amihud, Y. and H. Mendelson. "Liquidity, Maturity, and the Yields on U.S. Treasury Securities," *Journal of Finance* 46 (September 1991), 1411–1426.

Amin, K. and A. Morton. "Implied Volatility Functions in Arbitrage Free Term Structure Models," *Journal of Financial Economics* 35 (1994), 193–234.

Beim, D. O. "Estimating Bond Liquidity," Working Paper FB 92-08, Graduate School of Business, Columbia University, April 1992.

Black, F., E. Derman, and W. Toy. "A One-Factor Model of Interest Rates and Its Application to Treasury Bond Options," *Financial Analysts Journal* 46, (January/February 1990), 33–39.

Bliss, R. and E. Ronn. "The Implied Volatility of U. S. Interest Rates: Evidence from Callable U. S. Treasuries," Working Paper 95-12, Federal Reserve Bank of Atlanta, November 1995.

Chambers, D. R., W. T. Carleton, and D. W. Waldman. "A New Approach to Estimation of the Term Structure of Interest Rates," *Journal of Financial and Quantitative Analysis* 19 (September 1984), 233–252.

Coleman, T. and L. Fisher. "Estimating Forward Interest Rates and Yield Curves from Government Bond Prices: Methodology and Selected Results," Working Paper, Rutgers University, June 1987.

Constantinides, G. M. and J. E. Ingersoll, Jr. "Optimal Bond Trading with Personal Taxes," *Journal of Financial Economics* 13 (September 1984), 299–335.

Cornell, B. and A. C. Shapiro. "The Mispricing of U.S. Treasury Bonds: A Case Study," *Review of Financial Studies* 2 (1989), 297–310.

Cox, J. C., J. E. Ingersoll, Jr., and S. Ross, "A Theory of the Term Structure of Interest Rates," *Econometrica* 53 (March 1985), 385–407.

Daves, P. R. and M. C. Ehrhardt. "Liquidity, Reconstruction, and the Value of U.S. Treasury Strips," Working Paper, University of Tennessee, July 1992.

Dermody, J. C. and E. Z. Prisman. "Term Structure Multiplicity and Clientele in Markets with Transaction Costs and Taxes," *Journal of Finance* 43 (September 1988), 893–911.

Dybvig, P. H. and S. A. Ross. "Arbitrage" in *The New Palgrave: Finance*, ed. J. Eatwell, M. Milgate and P. Newman, W.W. Norton, New York, 1989.

Fama, E. F. and R. R. Bliss. "The Information in Long-Maturity Forward Rates," *American Economic Review* 77 (September 1987), 680–692.

Fisher, M., D. Nychka, and D. Zervos. "Fitting the Term Structure of Interest Rates with Smoothing Splines," Working Paper 95-1, Finance and Economics Discussion Series, Federal Reserve Board, January 1995.

Friedman, M. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association* 32 (December 1937), 675–701.

Heath, D., R. Jarrow, and A. Morton. "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation," *Journal of Financial and Quantitative Analysis* 25:4 (December 1990), 419–440.

Heath, D., R. Jarrow, and A. Morton. "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation," *Econometrica* 60 (January 1992), 77–105.

Ho, T. and S. Lee. "Term Structure Movements and the Pricing of Interest Rate Contingent Claims," *Journal of Finance* 41:5 (December 1986), 1011–1030.

Hull, J. and A. White. "Pricing Interest-Rate-Derivative Securities," *Review of Financial Studies* 3:4 (1993), 573–592.

Jordan, J. V. "Tax Effects in Term Structure Estimation," *Journal of Finance* 34 (June 1984), 393–406.

Kamara, A. "Liquidity and Short-Term Treasury Yields," Working Paper, Graduate School of Business Administration, University of Washington, September 1990.

Litzenberger, R. H. and J. Rolfo. "An International Study of Tax Effects on Government Bonds," *Journal of Finance* 39 (March 1984), 1–22.

Livingston, M. and S. Jain. "Flattening of Bond Maturity Curves for Long Maturities," *Journal of Finance* 37 (March 1982), 157–167.

McCulloch, J. H. "The Tax-Adjusted Yield Curve," *Journal of Finance* 30 (June 1975), 811–830.

Nelson, J. and S. Schaefer. "The Dynamics of the Term Structure and Alternative Portfolio Immunization Strategies," in *Innovations in Bond Portfolio Management*, ed. G. G. Kaufman, G. O. Bierwag and A. Toevs, JAI Press, Greenwich, Connecticut, 1983.

Nelson, C. R. and A. F. Siegel. "Parsimonious Modeling of Yield Curves," *Journal of Business* 60 (October 1987), 473–489.

Ronn, E. I. "A New Linear Programming Approach to Bond Portfolio Management," *Journal of Financial and Quantitative Analysis* 22 (December 1987), 439–466.

Schaefer, S. M. "Tax-Induced Clientele Effects in the Market for British Government Securities," *Journal of Financial Economics* 10 (July 1982), 121–159.

Siegel, A. F. and C. R. Nelson. "Long-Term Behavior of Yield Curves," *Journal of Financial and Quantitative Analysis* 23 (March 1988), 105–110.

Vasicek, O. "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5 (1977), 177–188.

Waggoner, D. "The Robustness of Yield and Forward Rate Curve Extraction Methods," Working Paper, Federal Reserve Bank of Atlanta, September 1996.

Table 1: Comparison of the Ability of Methodologies of Varying Complexity to Fit the Data

Estimation Method	# Parms.	Up to 1 Yr	1 to 3 Yrs	3 to 5 Yrs	5 to 10 Yrs	Above 10 Yrs	All Maturities
Flat	1	0.309	0.645	0.844	2.814	9.604	2.033
Monotonic	3	0.025	0.097	0.192	0.409	0.835	0.236
Extended Nelson-Siegel	5	0.013	0.066	0.144	0.338	0.646	0.181
Smoothed Fama-Bliss	5	0.011	0.066	0.146	0.344	0.696	0.188
McCulloch Cubic Spline	7 - 14	0.009	0.056	0.122	0.291	0.251	0.118
Fisher, et al. Cubic Spline	2 - 33	0.037	0.056	0.117	0.235	0.139	0.101
Unsmoothed Fama-Bliss	42 - 163	0.002	0.032	0.074	0.156	0.072	0.057
Number of Observations		9,794	11,634	6,641	6,553	4,734	39,356

- Reported statistics are the "in-sample" Mean Absolute Fitted-Price Errors where non-zero errors are defined relative to the bid-ask quotes as in equation (3).
- Results are based on estimates using all eligible bonds in the Long Data Set.
- Observations from different months were stacked together in each regression. "Number of Observations" is the sum over all 312 months of data.
- "# Parms." refers to the number of parameters in the term structure estimated in each of 312 months. For the McCulloch Cubic Spline and the Unsmoothed Fama-Bliss, this depends on the number of bonds available each month (after filtering in the latter case). For the Fisher, et al. Cubic Spline, the "effective" number of parameters is reported.
- The average numbers of parameters are: McCulloch–11.1; Fisher, et al.–11.2; Unsmoothed Fama-Bliss–98.7.
- The "Flat" and "Monotonic" term structures were estimated using restricted variants of the Extended Nelson-Siegel functional form, equation (2):

$$\text{Flat} \qquad r_m = \beta_0 \qquad\qquad \text{Monotonic} \qquad r_m = \beta_0 + \beta_1 \left[\frac{1 - e^{-m/\tau_1}}{m/\tau_1} \right].$$

- Due to the large number of observations, all parameter estimates are significantly different from zero.

Table 2: Comparison of In- and Out-of-Sample Mean WMAE Statistics and Hit Rates
(All Years)

Data Set Used	Unsmoothed Fama-Bliss	Smoothed Fama-Bliss	McCulloch Cubic Spline	Extended Nelson-Siegel	Fisher, et al. Cubic Spline
	Estimation Method Used				
In-sample results					
Short	0.0080 (92.4)	0.0197 (46.0)	0.0206 (44.6)	0.0217 (36.9)	0.0427 (39.1)
Long	0.0130 (92.2)	0.0390 (34.1)	0.0330 (36.9)	0.0392 (29.7)	0.0637 (34.7)
Out-of-sample results					
Short	0.0201 (50.8)	0.0204 (44.8)	0.0234 (40.8)	0.0221 (38.4)	0.0448 (36.7)
Long	0.0309 (44.7)	0.0377 (33.4)	0.0350 (35.0)	0.0376 (30.5)	0.0704 (31.3)
Out-of-sample results using "Fama-Bliss filtered" estimation subsample					
Short	Same as above	Same as above	0.0233 (41.6)	0.0216 (40.7)	0.0318 (40.3)
Long	Same as above	Same as above	0.0347 (35.7)	0.0368 (31.8)	0.0517 (34.1)

- Underlying results are the duration-weighted mean of the absolute fitted-price errors in the estimation subsample for each of 312 monthly partitions of the alternate available issues into estimation and hold-out (test) subsamples.

- Fitted-price errors are defined outside the issues' bid-ask range.

- Numbers in parentheses give the hit rates or percentage of issues with fitted prices falling within the quoted spread.

- The "Fama-Bliss filtered" estimation subsample consists of those bonds that survive the Unsmoothed Fama-Bliss filtering process. McCulloch, Extended Nelson-Siegel, and Fisher, et al. were re-estimated using this filtered estimation subsample to show the effects of removing outliers on these methods. Unsmoothed and Smoothed Fama-Bliss results are marked "Same as above" because they are identical to the previous panel.

Table 3: Out-of-Sample Non-Parametric WMAE Comparisons

		Rank scores, S_i, for individual methodologies				
Data Set Used	Friedman Statistic	Unsmoothed Fama-Bliss	Smoothed Fama-Bliss	McCulloch Cubic Spline	Extended Nelson-Siegel	Fisher, et al. Cubic Spline
			All Years			
Short	387.7**	693	675	1014	951	1347
Long	383.2**	563	994	817	998	1308
			1970–1978			
Short	152.2**	203	242	345	378	452
Long	148.3**	217	270	311	339	483
			1979–1982			
Short	92.5**	114	99	173	111	223
Long	89.0**	96	122	145	126	231
			1983–1995			
Short	176.6**	376	334	496	462	672
Long	248.8**	250	602	361	533	594

* (**) indicates the Friedman statistic is significant at the 5% (1%) level.

- Underlying results are the duration-weighted mean of the absolute fitted-price errors in the hold-out subsample for each of 312 monthly partitions of the alternate available issues into estimation and hold-out (test) subsamples.
- Fitted-price errors are defined outside the issues' bid-ask range.
- The Friedman Statistic is used to test whether all methodologies for a given maturity-range are identical. The critical value for $\alpha = 0.05$ (0.01) is 9.5 (13.3) for all tests.
- Pairwise comparisons may be made on the basis of the differences in individual rank scores, $\mid S_i - S_j \mid$. The critical values are:

α	0.05	0.01
All Years	111	130
1970–1978	65	76
1979–1982	43	51
1983–1995	78	92

Table 4: Transition Matrices for Full Sample Fitted-Price Errors for Notes and Bonds: (All Years)

Estimation Method Used		Unconditional Frequency(ϵ_t)	Conditional Frequency($\epsilon_{i,t+1} \mid \epsilon_{it}$) $\epsilon_{t+1} > 0$	$\epsilon_{t+1} = 0$	$\epsilon_{t+1} < 0$
Unsmoothed Fama-Bliss	$\epsilon_t > 0$	9.6	65.6	32.4	2.0
	$\epsilon_t = 0$	79.5	4.0	92.9	3.1
	$\epsilon_t < 0$	10.9	1.6	23.8	74.6
Smoothed Fama-Bliss	$\epsilon_t > 0$	37.6	75.2	18.6	6.2
	$\epsilon_t = 0$	27.6	25.3	52.6	22.2
	$\epsilon_t < 0$	34.8	6.9	18.8	74.3
McCulloch Cubic Spline	$\epsilon_t > 0$	33.9	74.5	21.7	3.8
	$\epsilon_t = 0$	36.0	21.2	61.0	17.8
	$\epsilon_t < 0$	30.2	4.5	23.2	72.3
Extended Nelson-Siegel	$\epsilon_t > 0$	41.6	76.5	17.6	5.9
	$\epsilon_t = 0$	26.0	28.1	49.8	22.1
	$\epsilon_t < 0$	32.4	7.7	19.1	73.2
Fisher, et al. Cubic Spline	$\epsilon_t > 0$	36.1	74.6	22.2	3.1
	$\epsilon_t = 0$	38.4	20.5	65.4	14.1
	$\epsilon_t < 0$	25.5	4.9	22.9	72.2

- The underlying data are the fitted-price errors generated using the full Long Data Set to estimate the term structures and compute fitted prices.
- Fitted-price errors are defined outside of the bid-ask spread.
- Transition frequencies report the likelihood that the fitted-price error on a given bond at time $t + 1$ is positive (zero, negative) given that the error on the same bond was positive (zero, negative) at time t. For instance, using the Unsmoothed Fama-Bliss term structures, 9.6% of the fitted-price errors were positive. Of these cases, 65.6% of the issues also had positive fitted-price errors the next period; 32.4% had zero fitted-price errors next period; and only 2.0% had negative fitted-price errors next period.

Table 5: Transition Matrices for Full Sample Fitted-Price Errors for Bills Only: (All Years)

Estimation Method Used		Unconditional Frequency(ϵ_t)	Conditional Frequency($\epsilon_{i,t+1} \mid \epsilon_{it}$)		
			$\epsilon_{t+1} > 0$	$\epsilon_{t+1} = 0$	$\epsilon_{t+1} < 0$
Unsmoothed Fama-Bliss	$\epsilon_t > 0$	0.4	7.1	89.3	3.6
	$\epsilon_t = 0$	97.7	0.3	97.8	1.8
	$\epsilon_t < 0$	1.9	2.2	71.7	26.1
Smoothed Fama-Bliss	$\epsilon_t > 0$	23.7	44.9	40.3	14.8
	$\epsilon_t = 0$	52.1	15.7	71.7	12.6
	$\epsilon_t < 0$	24.2	14.9	30.0	55.1
McCulloch Cubic Spline	$\epsilon_t > 0$	22.6	50.5	35.5	14.0
	$\epsilon_t = 0$	53.4	15.6	69.3	15.1
	$\epsilon_t < 0$	24.1	12.8	36.1	51.1
Extended Nelson-Siegel	$\epsilon_t > 0$	30.7	49.7	35.7	14.6
	$\epsilon_t = 0$	42.8	21.6	61.0	17.4
	$\epsilon_t < 0$	26.5	19.1	28.7	52.1
Fisher, et al. Cubic Spline	$\epsilon_t > 0$	19.6	48.5	31.9	19.7
	$\epsilon_t = 0$	29.7	13.1	52.0	34.9
	$\epsilon_t < 0$	50.7	5.3	13.7	81.0

- See notes for Table 4.

Table 6: Coincidence Matrices for Full Sample Fitted-Price Errors for Notes and Bonds: (All Years)

McCulloch Cubic Spline	Alternate Frequency($\epsilon_t^A \mid \epsilon_t^{MC}$)		
	$\epsilon_t^A > 0$	$\epsilon_t^A = 0$	$\epsilon_t^A < 0$
		Unsmoothed Fama-Bliss	
$\epsilon_t^{MC} > 0$	20.0	78.5	1.5
$\epsilon_t^{MC} = 0$	6.0	90.1	3.9
$\epsilon_t^{MC} < 0$	2.2	68.0	29.9
		Smoothed Fama-Bliss	
$\epsilon_t^{MC} > 0$	67.1	20.1	12.9
$\epsilon_t^{MC} = 0$	31.0	44.0	25.0
$\epsilon_t^{MC} < 0$	12.6	16.4	71.0
		Extended Nelson-Siegel	
$\epsilon_t^{MC} > 0$	72.8	15.5	11.7
$\epsilon_t^{MC} = 0$	36.8	41.9	21.3
$\epsilon_t^{MC} < 0$	12.4	18.9	68.7
		Fisher, et al. Cubic Spline	
$\epsilon_t^{MC} > 0$	75.0	21.4	3.5
$\epsilon_t^{MC} = 0$	24.2	65.4	10.3
$\epsilon_t^{MC} < 0$	6.5	25.2	68.3

- The underlying data are the fitted-price errors generated using the full Long Data Set to estimate the term structures and compute fitted prices.
- Fitted-price errors are defined outside of the bid-ask spread.
- Coincidence frequencies report the likelihood that the fitted-price error on a given bond using an alternate methodology (ϵ_t^A) is positive (zero, negative) given that the error on the same bond was positive (zero, negative) when using the McCulloch Cubic Spline methodology (ϵ_t^{MC}). For instance, consider the cases where the McCulloch Cubic Spline produced positive fitted-price errors. For these same issues, the Unsmoothed Fama-Bliss produced positive errors 20.0% of the time, zero errors 78.5% of the time, and negative errors 1.5% of the time.

Table 7: Out-of-Sample Non-Zero Fitted-Price Error Regressions for Notes and Bonds: (All Years)

	Unsmoothed Fama-Bliss	Smoothed Fama-Bliss	McCulloch Cubic Spline	Extended Nelson-Siegel	Fisher, et al. Cubic Spline
Constant	−1.862	−1.936	−1.376	−1.717	−1.845
	(−15.23)	(−16.01)	(−12.83)	(−14.95)	(−13.91)
Term to Maturity	0.001	−0.007	0.002	0.016	0.000
	(1.03)	(−8.19)	(3.05)	(19.05)	(−0.45)
Age	0.035	0.034	0.029	0.038	0.039
	(15.00)	(14.59)	(14.45)	(17.02)	(15.49)
Premium	−0.015	−0.027	−0.018	−0.023	−0.014
	(−21.76)	(−40.67)	(−30.04)	(−36.54)	(−18.51)
Discount	0.014	0.019	0.013	0.013	0.007
	(10.26)	(14.52)	(11.19)	(10.07)	(5.14)
Time	0.185	0.210	0.131	0.165	0.178
	(12.37)	(14.22)	(10.00)	(11.79)	(10.96)
Amount Outstanding	0.005	−0.001	0.004	0.006	0.007
	(2.79)	(−0.66)	(2.70)	(3.58)	(4.09)
Bid-Ask Spread	0.119	0.110	0.151	0.104	0.129
	(24.84)	(22.63)	(35.44)	(21.70)	(24.33)
Valid cases:	9188	10565	9964	10758	9752
R-squared:	0.215	0.305	0.312	0.248	0.167
Std error of est:	0.508	0.537	0.461	0.520	0.560
F:	358.8	661.1	645.5	507.5	280.0

Age	≡	Years since dated date
Premium	≡	$\hat{P} - 100$ if $\hat{P} > (P^A + P^B)/2$, zero otherwise
Discount	≡	$100 - \hat{P}$ if $\hat{P} < (P^A + P^B)/2$, zero otherwise
Time	≡	quote date expressed as YY.FRAC

- Issues in the Long Data Set were first alternately divided into estimation and hold-out subsamples each month. Term structures estimated each month using the estimation subsample were then used to compute fitted prices for issues in the hold-out subsample.
- Fitted-price errors were defined relative to the spread.
- Estimation subsample contains bills, notes and bonds using the Long Data Set.
- Numbers in parentheses indicate coefficient t-statistics.

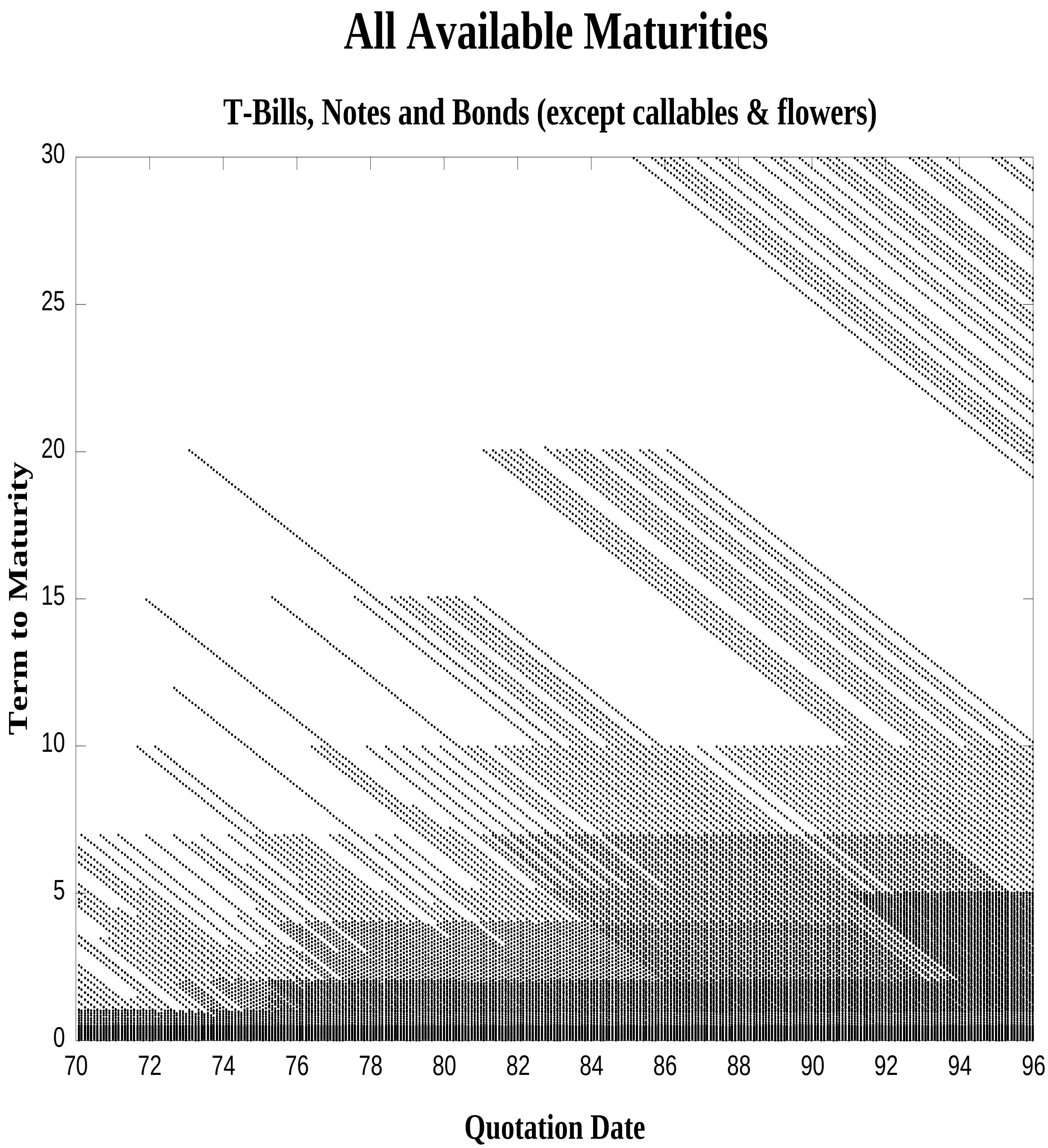
Table 8: Out-of-Sample Non-Zero Fitted-Price Error Regressions for Bills: (All Years)

	Unsmoothed Fama-Bliss	Smoothed Fama-Bliss	McCulloch Cubic Spline	Extended Nelson-Siegel	Fisher, et al. Cubic Spline
Constant	−0.090	−0.104	−0.032	−0.106	−0.124
	(−4.00)	(−5.77)	(−2.04)	(−6.89)	(−2.92)
Age	−0.008	−0.002	−0.009	−0.001	0.020
	(−1.77)	(−0.51)	(−2.78)	(−0.20)	(2.30)
Discount	−0.003	0.001	0.000	0.000	−0.004
	(−5.15)	(1.80)	(0.82)	(−0.67)	(−3.44)
Time	0.013	0.014	0.005	0.015	0.027
	(4.71)	(6.54)	(2.52)	(7.76)	(4.98)
Amount	−0.001	−0.001	0.000	−0.001	−0.003
Outstanding	(−2.85)	(−5.12)	(−1.22)	(−6.68)	(−6.25)
Bid-Ask	0.008	−0.023	−0.004	−0.016	−0.049
Spread	(1.33)	(−5.51)	(−1.06)	(−4.52)	(−5.00)
Valid cases:	1482	2297	2582	2654	3520
R-squared:	0.038	0.072	0.014	0.068	0.037
Std error of est:	0.039	0.040	0.036	0.037	0.116
F:	11.8	35.5	7.1	38.6	27.0

Age $\equiv$ Years since dated date
Discount $\equiv$ $100 - \hat{P}$ if $\hat{P} < (P^A + P^B)/2$, zero otherwise
Time $\equiv$ quote date expressed as YY.FRAC

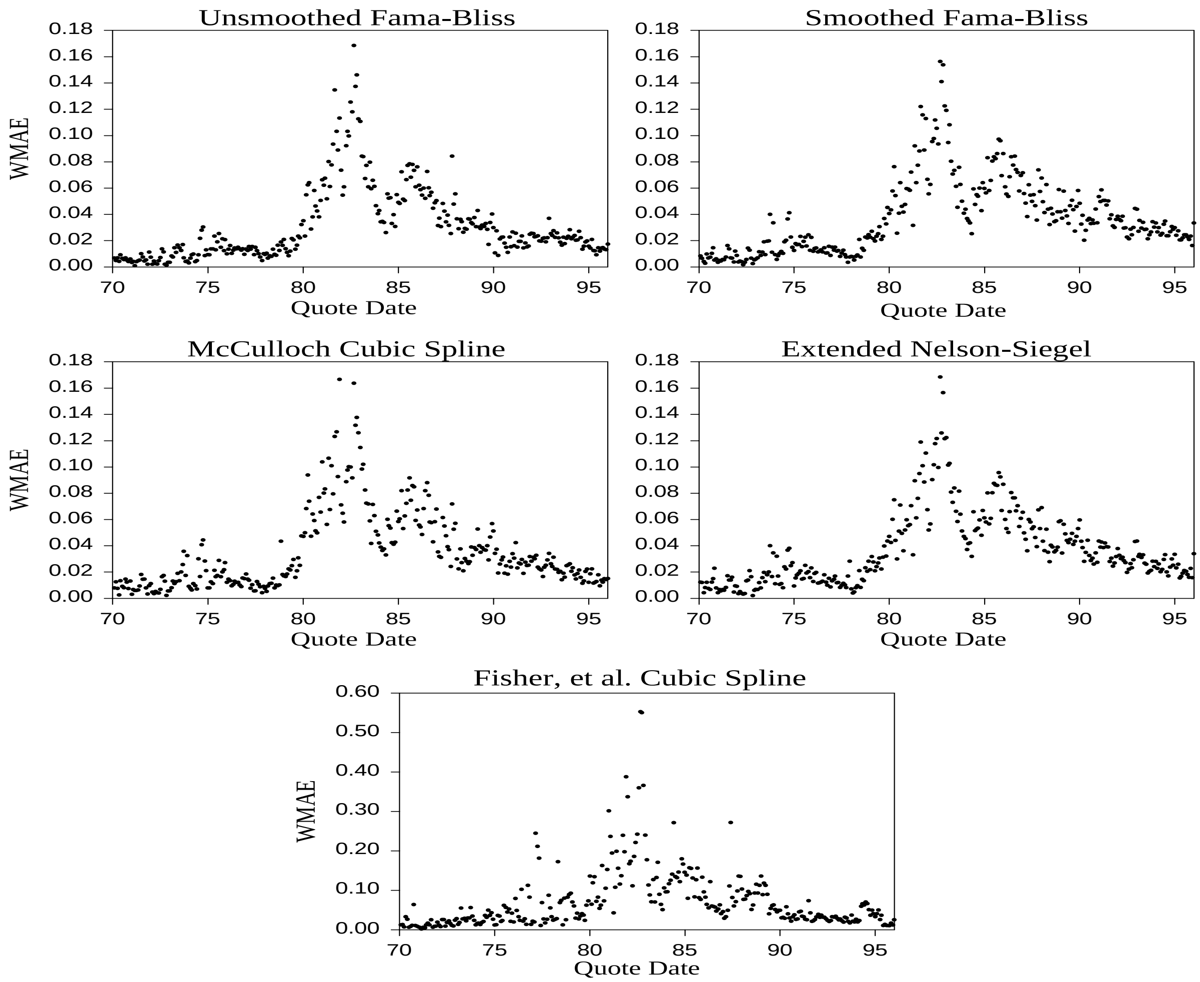
- Issues in the Long Data Set were first alternately divided into estimation and hold-out subsamples each month. Term structures estimated each month using the estimation subsample were then used to compute fitted prices for issues in the hold-out subsample.
- Fitted-price errors were defined relative to the spread.
- Estimation subsample contains bills, notes and bonds using the Long Data Set.
- Numbers in parentheses indicate coefficient t-statistics.

Figure 1: Maturity of Available Issues After Exclusions



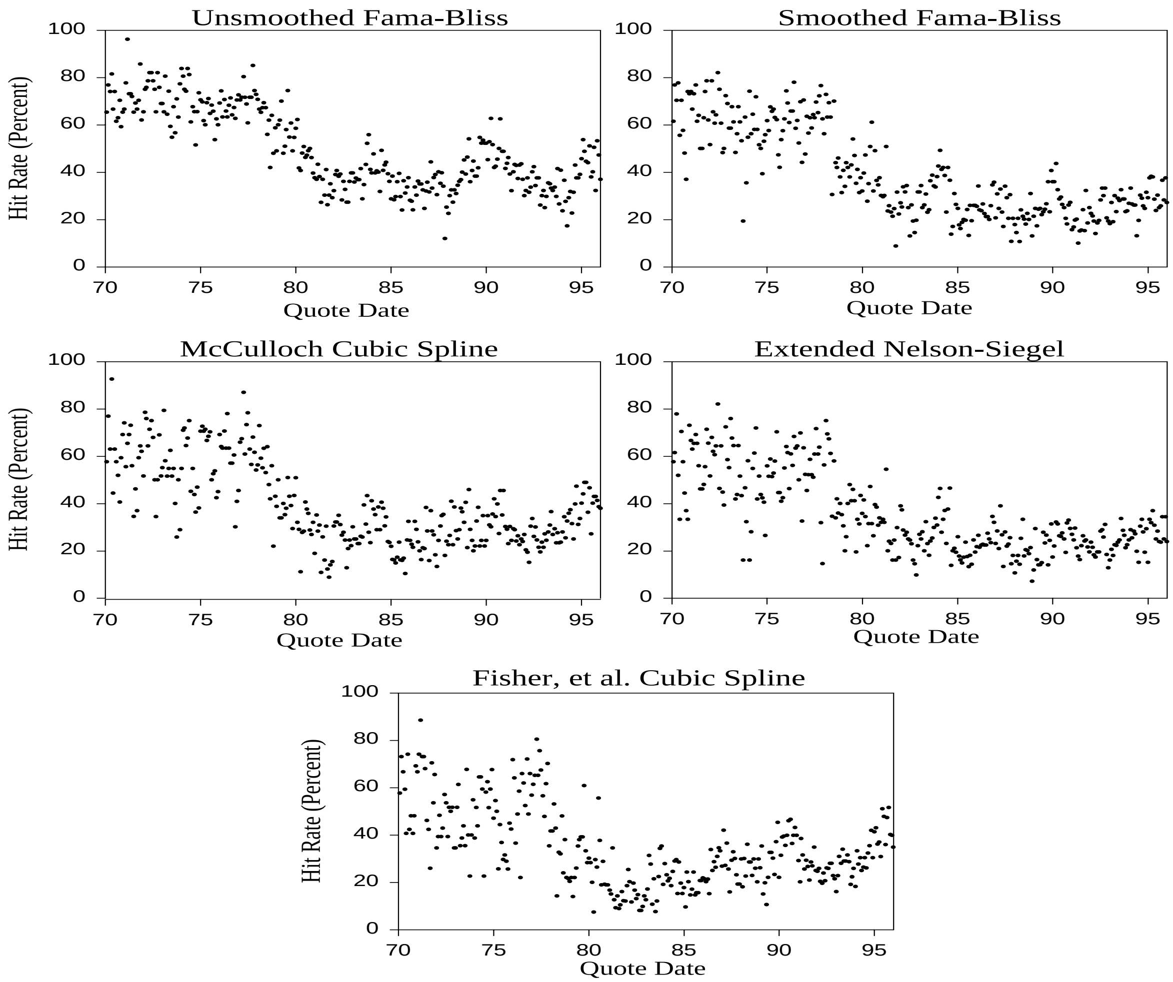
- Apparent changes in "shading" reflect increased density of individual issues.

Figure 2: Time Series of Long Data Set Out-of-Sample WMAE Statistics



- The underlying statistics are the out-of-sample, duration-weighted, mean-absolute fitted-price errors for each observation month.
- The Long Data Set was alternately divided into estimation and hold-out subsamples each month. The estimation subsample is used to estimate the term structure each month, which is then used to compute fitted prices for issues in the hold-out subsample.
- Fitted-price errors are defined relative to the bid-ask spread.

Figure 3: Time Series of Long Data Set Out-of-Sample Hit Rates



- The underlying statistics are percentage of out-of-sample bonds each month with fitted prices falling between the bid and asked quotes.
- The Long Data Set was alternately divided into estimation and hold-out subsamples each month. The estimation subsample is used to estimate the term structure each month, which is then used to compute fitted prices for issues in the hold-out subsample.
- Fitted-price errors are defined relative to the bid-ask spread.