Fordham University
DigitalResearch@Fordham

CRIF Working Paper series

Frank J. Petrilli Center for Research in International Finance

3-16-2000

Predictable Changes in Yields and Forward Rates

David Backus
New York University

Silverio Foresi
Goldman Sachs Asset Management

Abon Mozumdar
Virginia Tech

Liuren Wu
Fordham University

Follow this and additional works at: http://fordham.bepress.com/crif_working_papers

Part of the Finance and Financial Management Commons

Recommended Citation
Backus, David; Foresi, Silverio; Mozumdar, Abon; and Wu, Liuren, "Predictable Changes in Yields and Forward Rates" (2000). *CRIF Working Paper series.* Paper 21.
http://fordham.bepress.com/crif_working_papers/21

This Article is brought to you for free and open access by the Frank J. Petrilli Center for Research in International Finance at DigitalResearch@Fordham. It has been accepted for inclusion in CRIF Working Paper series by an authorized administrator of DigitalResearch@Fordham. For more information, please contact considine@fordham.edu.

Predictable Changes in Yields and Forward Rates*

David Backus,† Silverio Foresi,‡ Abon Mozumdar,§ and Liuren Wu¶

First draft: April 1, 1997
This version: March 16, 2000

Abstract

We make two contributions to the study of interest rates. The first is to characterize their dynamics in a new way. We estimate forecasting relations based on one-period changes in forward rates, which are more easily compared than earlier work on yields to the stationary theory of bond pricing. The second is to approximate these dynamics and other salient features of interest rates with an affine model. We show that models with "negative" factors come closer to accounting for the properties of interest rates, including their dynamics, than multifactor Cox-Ingersoll-Ross models.

JEL Classification Codes: E43, G12.

Keywords: forecasting; term premiums; expectations hypothesis; pricing kernels; affine models.

---

*We thank Robert Bliss for providing the computer programs and data used to estimate forward rates and yields. We also thank John Campbell, Qiang Dai, Darrell Duffie, Allan Gregory, Sandy Grossman, Robert Hodrick, Erzo Luttmer, Ken Singleton, and Stan Zin for helpful comments; and a referee for unusually constructive suggestions on several occasions. The views expressed in this paper are those of the authors and do not necessarily reflect those of Goldman, Sachs. The most recent version of this paper is available at http://www.stern.nyu.edu/~dbackus.

†Stern School of Business, New York University and NBER; dbackus@stern.nyu.edu.

‡Goldman Sachs Asset Management; Silverio.Foresi@gs.com.

§Pamplin College of Business, Virginia Tech; abon@vt.edu.

¶Graduate School of Business, Fordham University; liwu@mary.fordham.edu.

1 Introduction

Over the last 25 years and more, enormous progress has been made in both the empirical study of interest rates and the modern theory of bond pricing. The two lines of research remain, however, largely distinct. Empirical research continues to revolve around the expectations hypothesis: that term premiums on long forward rates are constant. Most studies conclude they are not. A large body of work to this effect has been surveyed repeatedly, most recently by Bekaert, Hodrick, and Marshall (1997a), Campbell (1995), and Evans and Lewis (1994). Although the evidence points away from the expectations hypothesis, these studies nevertheless provide useful summaries of interest rate dynamics. They illustrate, for example, how future interest rates can be predicted with (say) spreads between long and short rates.

The modern "arbitrage-free" theory of bond pricing continues to develop along lines laid down by Vasicek (1977) and Cox, Ingersoll, and Ross (1985). Modern theory, in its most general form, places few restrictions on the behavior of term premiums, and certainly carries no presumption in favor of the expectations hypothesis. We could ask, however, what kinds of models might account for observed patterns of predictability in interest rates. We consider a subset of the affine yield models characterized by Duffie and Kan (1996), which includes the Vasicek and Cox-Ingersoll-Ross models as special cases. Earlier work by Frachot and Lesne (1994) and Roberds and Whiteman (1999) studied the ability of one- and two-factor Cox-Ingersoll-Ross (1985) models to account for observed departures from the expectations hypothesis with term premiums that vary through time. We examine a broader class of models and ask whether they can account for both predictable variation in term premiums and other salient features of interest rates.

We take a series of steps that we think serve to integrate evidence with theory. The first is to cast the evidence in a new form based directly on the martingale property of forward rates under the expectations hypothesis. In this form, the largest differences from the expectations hypothesis are for maturities under two years. This feature of the data conforms with a broad class of stationary bond pricing models, in which term premiums on long forward rates are approximately constant, and thus holds out the possibility that evidence and theory might eventually be reconciled. The second is to illustrate the difficulty of accounting for this evidence with the one-factor Cox-Ingersoll-Ross model. The one-factor model cannot account simultaneously for the behavior of term premiums and the average upward slope of the yield curve. Although this result does not extend to multi-factor models, it highlights the source of tension between theory and data and suggests a possible resolution. The third is to propose and estimate a new model in the affine class in which the short rate depends negatively on one or more

"square-root" factors. We use this model and a two-factor extension to account for term premiums and other features of interest rates. We argue that they approximate the data substantially better than one- and two-factor Cox-Ingersoll-Ross models. The fourth is to compare our summary of the evidence (in which the largest deviations from the expectations hypothesis are at short maturities) to earlier work based on bond yields (in which the largest deviations are at long maturities). We suggest that the large numerical differences between these two approaches mask a broad similarity in their information content.

2 Notation and Data

In what follows, the continuously-compounded yield on an n-period bond at date t is denoted y_t^n and defined by

$$y_t^n = -n^{-1} \log b_t^n, \tag{1}$$

where b_t^n is the dollar price at date t of a claim to one dollar at $t+n$. One-period forward rates are defined by

$$f_t^n = \log(b_t^n / b_t^{n+1}), \tag{2}$$

so that yields are averages of forward rates:

$$y_t^n = n^{-1} \sum_{i=0}^{n-1} f_t^i. \tag{3}$$

The short rate is $r_t = y_t^1 = f_t^0$. In our data, a period is one month and interest rates are reported as annual percentages, meaning that they are multiplied by 1200.

In practice, yields and forward rates are estimated rather than observed. From prices of bonds for a variety of maturities, the "discount function" b_t^n (viewed as a function of n at each date t) is interpolated between missing maturities n and smoothed to reduce the impact of noise (nonsynchronous price quotes, bid/ask spreads, market anomalies, and so on). There is no generally-accepted best practice for doing this. We follow Bliss (1997) in using four methods: Smoothed Fama-Bliss, Unsmoothed Fama-Bliss, McCulloch Cubic Spline, and Extended Nelson-Siegel. Fama and Bliss's unsmoothed method extracts forward rates from prices of bonds of successive maturities using a relation analogous to (2) for coupon bonds. McCulloch approximates the discount function with a cubic spline. Nelson and Siegel's extended method approximates the yield curve, rather than the discount function, with a smooth function that gives the long end of the yield curve a horizontal asymptote. Fama and Bliss's smoothed method applies a similar approximation to yields implied by their unsmoothed forward rates. Bliss (1997) provides more

detailed descriptions of each of these methods. Among the four methods, Unsmoothed Fama-Bliss sticks out in not smoothing the raw data in some way.

We apply all four methods to bond prices collected by the Center for Research in Securities Prices (CRSP) at the University of Chicago using computer programs and data supplied by Bliss. Academic research (Elton and Green 1998, for example) and Wall Street folklore suggest that CRSP bond prices have more noise in them than those used in industry, but data of higher quality are not available over sufficiently long sample periods. Elton and Green, for example, study just 3 years of intraday data. Our sample runs from January 1970 to December 1995 (312 monthly observations).

3 Predictable Changes in Forward Rates

We use these four datasets to characterize the dynamics of interest rates in a form that highlights the connection between the evidence and stationary theories of bond pricing. We predict future forward rates with the regression

$$f_{t+1}^{n-1} - r_t \;=\; \text{constant} + c_n(f_t^n - r_t) + \text{residual} \tag{4}$$

for maturities n between 1 month and 10 years. As far as we know, this "forward rate regression" is new to the literature. We consider other approaches to interest rate prediction in Section 6.

We regard (4) as a linear forecasting relation, but its intellectual roots lie in the expectations hypothesis. The most common statement of the expectations hypothesis is that forward rates are expectations of future short rates,

$$f_t^n \;=\; E_t r_{t+n}, \tag{5}$$

which implies that forward rates are martingales,

$$f_t^n \;=\; E_t f_{t+1}^{n-1}. \tag{6}$$

See Roll (1970, Chapter 4) or Sargent (1987, Section X.7). Since (5) is easily contradicted by the data (average forward rates vary systematically across maturities), most people interpret the expectations hypothesis as including the possibility of a constant term premium p:

$$f_t^n \;=\; E_t r_{t+n} + p^n. \tag{5'}$$

The generic alternative is that term premiums vary through time:

$$f_t^n \;=\; E_t r_{t+n} + p_t^n. \tag{7}$$

Evidence against the expectations hypothesis, as defined by equation (5′), is thus evidence that term premiums vary over time. Given (6), or the weaker (5′), equation (4) has slope $c_n = 1$ for all maturities n. The role of the term premium is apparent from the complementary regression,

$$f_{t+1}^{n-1} - f_t^n \;=\; \text{constant} + (c_n - 1)(f_t^n - r_t) + \text{residual},$$

a linear transformation of (4). When term premiums are constant, $E_t f_{t+1}^{n-1} - f_t^n = p^{n-1} - p^n$, a constant, and $c_n - 1 = 0$. Values of c_n different from one thus indicate that term premiums vary through time. Whether or not the expectations hypothesis holds, nonzero values of c_n in (4) indicate that forward rate spreads contain information that can be used to forecast future forward rates.

In its most general form, the modern theory of bond pricing has few quantitative implications for the behavior of term premiums. However, it does have implications for the behavior of long forward rates. In a broad class of stationary models, the theoretical regression slope c_n approaches one as n approaches infinity. Related propositions are proved in different settings by Backus, Gregory, and Zin (1988, Proposition 2) and Dybvig, Ingersoll, and Ross (1996). More important to us: this feature is easily compared with the evidence.

We report estimates of equation (4) in Table 1 and Figure 1 for all four sets of forward rate data. (Figure 1 is based on a larger set of maturities, but is otherwise identical.) It should be no surprise that the data differ substantially from the expectations hypothesis. With the exception of the Unsmoothed Fama-Bliss data, estimated regression slopes are about one-half for $n = 1$, increase monotonically with maturity, and level off just below one. Figure 1 presents the evidence in particularly vivid form. There we see that the largest deviations from the expectations hypothesis come at short maturities: regression slopes are close to one for maturities of 24 months or longer. Although this feature of the data conforms reasonably well with stationary bond-pricing theory, it is not an exact reproduction of it. Estimated slope coefficients are about 0.96 at long maturities, not 1.0 as suggested by theory, with standard errors of 0.02 or smaller. The evidence suggests, then, that changes in forward rates at short maturities consist of predictable changes in both short rates and term premiums. It also suggests that the behavior of long forward rates is in approximate agreement with stationary theory.

Before turning to possible explanations, we consider two potential problems with the evidence. One is measurement error; see, for example, Bekaert, Hodrick, and Marshall (1997a) and Stambaugh (1988). The effect of such error on estimated regression slopes depends on its form. Error in the short rate would likely push slope estimates toward one, but error in long forward rates would probably push them toward zero. To make

this concrete, suppose our observations of forward rates $\hat{f}$ differ from "true" forward rates f in having measurement error η:

$$\hat{f}_t^i = f_t^i + \eta_t^i,$$

where η_t^i is independent of true forward rates, has variance σ_i^2, is uncorrelated with errors at different dates ($Corr(\eta_t^i, \eta_s^j) = 0$ for all dates $t \neq s$ and all maturities i,j), and has arbitrary correlation with contemporaneous errors ($Corr(\eta_t^i, \eta_t^j) = \rho_{ij}$). The population regression slope for true forward rates is

$$\begin{aligned} c_n &= \frac{Cov(f_{t+1}^{n-1} - r_t, f_t^n - r_t)}{Var(f_t^n - r_t)} \\ &= \frac{Cov(\hat{f}_{t+1}^{n-1} - \hat{r}_t, \hat{f}_t^n - \hat{r}_t) - [\sigma_0^2 - \rho_{0n}\sigma_0\sigma_n]}{Var(\hat{f}_t^n - \hat{r}_t) - [\sigma_0^2 + \sigma_n^2 - 2\rho_{0n}\sigma_0\sigma_n]}. \end{aligned} \tag{8}$$

Given estimates of error variances and correlations, we can estimate the impact on estimated regression slopes.

The difficulty is quantifying the error. One source of estimates is provided by McCulloch and Kwon (1993), who report standard errors for each estimated forward rate. The root-mean-square of these standard errors is therefore an estimate of the standard deviation of the measurement error. We report these standard deviations in Table 2 for a number of maturities for the period January 1970 to February 1991, the overlap between our sample and McColloch and Kwon's. Standard deviations vary between 9 and 21 basis points, depending on the maturity.

Another approach is suggested by Bekaert, Hodrick, and Marshall (1997a): standard deviations for differences between forward rates estimated by different methods, which are also reported in Table 2. Depending on the source, these numbers may either under- or over-estimate the magnitude of measurement error. If measurement error stems from the underlying data, and is therefore largely common across methods, we would expect the standard deviations computed this way to underestimate the measurement error. But if the error stems from the method, we might expect the standard deviations to overstate the error in one method alone. For the most part, the estimates point in the same direction as McCulloch and Kwon's: to measurement error with a standard deviation under 30 basis points for most maturities. The Unsmoothed Fama-Bliss estimates are a striking exception, with very large standard deviations for long maturities.

We use (8) to estimate the impact of measurement error on estimated regression slopes. For illustration, consider the effect on the first and last regression slopes, c_1 and c_{120}. For the Smoothed Fama-Bliss data $Cov(\hat{f}_{t+1}^{n-1} - \hat{r}_t, \hat{f}_t^n - \hat{r}_t) = 0.0985$ for $n = 1$ and 3.196 for $n = 120$. Similarly, $Var(\hat{f}_t^n - \hat{r}_t) = 0.2162$ for $n = 1$ and 3.317 for $n = 120$.

If we use the McCulloch and Kwon standard deviations from Table 2 and assume that errors are uncorrelated across maturities, then the error-adjusted slope estimates are

$$\begin{aligned} c_1 &= 0.462 \\ c_{120} &= 0.968. \end{aligned}$$

The impact is evidently small, particularly at long maturities. To get $c_{120} = 1$, for example, we need a standard deviation of 35 basis points for the error of f^{120}, a value more than double what we see in Table 2. Positive correlation between the errors typically reduces its overall effect. The effect of error is greater at the short end, largely because the relevant variances and covariances are smaller, but even here is not large enough to change one's interpretation of the evidence.

A second potential problem with the evidence is small sample bias. The methods used by Bekaert, Hodrick, and Marshall (1997a) imply that small sample bias tends to push estimates of c_n above one, however, not below. Suppose, as they do, that the short rate is AR(1):

$$r_{t+1} = \text{constant} + \varphi r_t + \varepsilon_{t+1},$$

where $\{\varepsilon_{t+1}\}$ is a sequence of iid innovations in the short rate. The counterpart of their Proposition 1 is

$$E(\hat{c}_n) = 1 - \left(\frac{\varphi^{n-1}}{1-\varphi^n}\right)[E(\hat{\varphi}) - \varphi],$$

where a "hat" again indicates an estimate of the underlying parameter. Since the small-sample bias in $\hat{\varphi}$ is approximately

$$E(\hat{\varphi}) - \varphi = -\frac{1+3\varphi}{T},$$

where T is the sample size, the bias in the regression slope under the expectations hypothesis is approximately

$$E(\hat{c}_n) - 1 = \left(\frac{\varphi^{n-1}}{1-\varphi^n}\right)\left(\frac{1+3\varphi}{T}\right). \tag{9}$$

Thus the estimated regression slope $\hat{c}_n$ is biased upward in small samples, the opposite direction of observed departures from the expectations hypothesis. Small sample bias does not, therefore, appear to be the source of apparent predictable variation in term premiums.

We conclude that while measurement error and small sample bias influence our estimated regression slopes, neither leads us to doubt the evidence that term premiums vary through time. Bekaert, Hodrick, and Marshall (1997b) suggest, in addition, that regime changes may increase the sampling variability of coefficient estimates, but nonetheless find the evidence against the expectations hypothesis persuasive.

4 Affine Models

The next step is to consider models that might account for the dynamics of forward rates for US government bonds. Affine models are a natural choice, because their linear structure makes it relatively simple to explore implications for linear forecasting relations like the forward rate regressions of equation (4). Multifactor Vasicek (1977) models can be ruled out immediately: they imply constant term premiums and therefore cannot account for the apparent correlation of forward rate changes with forward rate spreads.

We consider a subset of Duffie and Kan's (1996) affine yield models in which conditional variances, and hence term premiums, vary over time. Fisher and Gilles (1996), Frachot and Lesne (1994), and Roberds and Whiteman (1999) study similar models for the same reason. Bond prices in these models are based on a vector of state variables z following

$$z_{t+1} - z_t \;=\; (I - \Phi)(\theta - z_t) + \Sigma V(z_t)^{1/2}\varepsilon_{t+1}, \tag{10}$$

where $\{\varepsilon_t\} \sim \mathrm{NID}(0, I)$, Φ is a stable matrix with positive diagonal elements, Σ is a diagonal matrix with elements σ_i, and $V(z_t)$ is a diagonal matrix with elements $v_i(z_t) = z_{it}$. State prices are governed by a pricing kernel of the form

$$-\log m_{t+1} \;=\; \delta + \gamma^\top z_t + \lambda^\top V(z_t)^{1/2}\varepsilon_{t+1}. \tag{11}$$

These models are a subset of Duffie and Kan's affine-yield models in which we have restricted ourselves to nonconstant volatility functions v_i that depend only on the ith state variable. Translation into discrete time is now standard; see, for example, Backus (1993), Campbell, Lo, and MacKinlay (1996, Chapter 11), and Sun (1992). The most common examples of this class are versions of the Cox-Ingersoll-Ross model with one or more factors: $\delta = 0$, Φ is diagonal, and $\gamma_i = 1 + \lambda_i^2/2$. The choice of γ_i is a normalization that defines the short rate as $r_t = \sum_i z_{it}$.

As with the Cox-Ingersoll-Ross model, the parameters are restricted to values that guarantee nonnegative volatility functions v_i. Define the set D of admissible states as those values of z for which volatility is nonnegative:

$$D \;=\; \{z | v_i(z) \geq 0 \text{ for all } i\}.$$

In our class of models, D consists of the positive orthant. Duffie and Kan (1996, Section 4) give sufficient conditions for z to remain in D. For each i and all $z \in D$ satisfying $v_i(z) = 0$ (the boundary of positive volatility), the "drift" [the conditional mean of (10)] must be positive:

$$\kappa_{ii}\theta_i + \sum_{j \neq i} \kappa_{ij}(\theta_j - z_j) \;>\; 0,$$

where κ_{ij} is an element of $K = I - \Phi$. This is exact in continuous time, approximate in our discrete-time analog. This implies $\kappa_{ii} = 1 - \varphi_{ii} > 0$ for all i and $\kappa_{ij} = -\varphi_{ij} < 0$ for all $i \neq j$. Among other things, the condition rules out unit roots in the state variables z: these models are stationary.

With this structure, bond prices are log-linear functions of the state variables z,

$$-\log b_t^n \;=\; A_n + B_n^\top z_t, \tag{12}$$

for some choice of coefficients $\{A_n, B_n\}$, where A_n is a scalar and B_n is a vector with typical element B_{in}. Bond prices satisfy the pricing relation

$$b_t^{n+1} \;=\; E_t\left(m_{t+1} b_{t+1}^n\right), \tag{13}$$

starting with $b_t^0 = 1$ (a dollar today costs a dollar). Application of (13) generates the recursions

$$\begin{aligned} A_{n+1} &= A_n + \delta + B_n^\top (I - \Phi)\theta \\ B_{i,n+1} &= \gamma_i + \sum_j B_{jn}\varphi_{ji} - (\lambda_i + B_{in}\sigma_i)^2/2, \end{aligned}$$

starting with $A_0 = 0$ and $B_0 = 0$ (an implication of $b_t^0 = 1$).

One further condition is used below: that B_n converges to a constant denoted by B. This is not an implication of the affine class as presented, but it holds in the original Cox-Ingersoll-Ross model and in all the estimated examples of the next section.

Now consider the population values of the slopes of forward rate regressions. The unconditional variance of z is the solution to

$$\Gamma_0 \;=\; \Phi\Gamma_0\Phi^\top + \Sigma V(\theta)\Sigma^\top,$$

where $\Sigma V(\theta)\Sigma^\top$ is a diagonal matrix with positive elements $\sigma_i^2\theta_i$. The solution is

$$vec(\Gamma_0) \;=\; (I - \Phi\otimes\Phi)^{-1}\, vec[\Sigma V(\theta)\Sigma^\top], \tag{14}$$

where $vec(A)$ is the vector formed from the columns of the matrix A. Forward rate regression slopes in this setting are then

$$c_n \;=\; \frac{[B_1 + B_n - B_{n+1}]^\top \Gamma_0\, [B_1 - \Phi^\top(B_n - B_{n-1})]}{[B_1 + B_n - B_{n+1}]^\top \Gamma_0\, [B_1 + B_n - B_{n+1}]}, \tag{15}$$

which evidently may take on values other than one.

Although (15) is relatively opaque, its limiting behavior is not. If B_n converges, the population value of the regression slope approaches one:

$$\lim_{n\to\infty} c_n \;=\; \frac{B_1^\top \Gamma_0\, B_1}{B_1^\top \Gamma_0\, B_1} \;=\; 1. \tag{16}$$

This is a generalization of a well-known property of the Cox-Ingersoll-Ross model: The variance of forward rates falls to zero with maturity, so for long enough maturities we are effectively regressing $-r$ on itself.

5 Estimating Affine Models

The theoretical class of affine models contains examples whose statistical properties vary enormously. We ask whether relatively simple examples can account for the slopes of forward rate regressions and other salient properties of interest rates. Since theory and evidence are in broad agreement at long maturities (regression slopes are about one), we focus on short maturities, where the evidence of time-varying term premiums is the strongest. Models characterized by equations (10,11) exhibit stochastic volatility in interest rates and are therefore a logical place to look.

We estimate 5 affine models by GMM and report the results in Table 4. Each is estimated using 11 moment conditions based, respectively, on the standard deviation and autocorrelation of the short rate; the mean and standard deviation of yield spreads for maturities 12, 60, and 120 months; and forward rate regression slopes for maturities 1, 6, and 12 months. We impose a 12th condition exactly: we set the mean short rate in each model equal to the sample mean. As a result, our reported standard errors likely understate the sampling variability of estimated parameters. A more complete description of the moment conditions is supplied in Appendix A. We use the same weighting matrix for each model: the Newey-West covariance matrix implied by estimates of the three-factor Model E, which includes each of the other models as a special case and thus provides a common basis of comparison. The weighting matrix is approximately a fixed point for Model E: it both produces and is produced by the parameters reported in the table. The number of lags (4) is based on calculations suggested by Andrews (1991). If residuals from the moment conditions are AR(1), the optimal "lag truncation parameter" for Model E is 3.64, which we round up to 4. Further detail is provided in Appendix A.

One-Factor Models

Model A in Table 4 is the one-factor Cox-Ingersoll-Ross model, a special case of (10,11) in which z is a scalar, $\delta = 0$, and $\gamma = 1 + \lambda^2/2$. The J-statistic suggests a poor approximation to the data, something that should come as no surprise from earlier work by Chen and Scott (1993) and Pearson and Sun (1994). If Model A did in fact generate the data, its J-statistic has approximately a chi-square distribution in large samples. Although in finite samples the distribution can be substantially different (Tauchen 1986), the J for this model indicates that the model is a grossly inadequate summary of the behavior of interest rates. Note, in particular, that the estimated parameters imply forward rate regressions with slopes greater than 1.

The model's poor approximation of regression slopes and mean yields reflects a fundamental tension between these two features in this model. In the model, $A_1 = 0$, $A_2 = (1-\varphi)\theta$, $B_1 = 1$, and $B_2 = 1 + \varphi - \sigma(\lambda + \sigma/2)$. Therefore $r_t = z_t$, $f_t^1 = (1-\varphi)\theta + [\varphi - \sigma(\lambda + \sigma/2)]z_t$, $f_t^1 - r_t = (1-\varphi)\theta - [1 - \varphi + \sigma(\lambda + \sigma/2)]z_t$, and $r_{t+1} - r_t = (1-\varphi)\theta - (1-\varphi)z_t + \sigma z_t^{1/2}\varepsilon_{t+1}$. The population value of the slope of the forward rate regression (4) for $n = 1$ is therefore

$$c_1 \;=\; \frac{1-\varphi}{1-\varphi+\sigma(\lambda+\sigma/2)}. \tag{17}$$

Depending on the parameter values, this can take on values greater than, equal to, or less than one. For the regression slope to lie between zero and one, as we see in the data, we need $\sigma(\lambda + \sigma/2) > 0$. This implies, however, a downward-sloping mean forward rate curve. For example,

$$E(f_t^1 - r_t) \;=\; [1 - \sigma(\lambda + \sigma/2)]\;\theta - \theta \;=\; -\sigma(\lambda + \sigma/2)\;\theta,$$

which is negative when we choose parameters to reproduce the slope of the first forward rate regression. The parameters reported by Frachot and Lesne (1994) illustrate this point: they reproduce various regression slopes but imply downward-sloping average yield and forward rate curves. In short, the one-factor Cox-Ingersoll-Ross model cannot generate both an increasing mean forward rate curve and a regression slope between zero and one.

The source of this difficulty is the behavior of the term premium. Consider the behavior of the first forward rate spread, $f_t^1 - r_t$. The spread has two components, the expected change in the short rate and a term premium:

$$f_t^1 - r_t \;=\; E_t r_{t+1} - r_t + p_t^1;$$

see equation (7). If the two components move in the same direction, the implied regression slope is less than one, as we see in the data. But if we estimate the parameters

to generate an upward-sloping average yield curve, as in Model A, the term premium and the expected change in the short rate move in opposite directions. The first term premium is

$$p_t^1 \;=\; -\sigma(\lambda + \sigma/2)\; z_t$$

and the expected change in the short rate is

$$E_t r_{t+1} - r_t \;=\; (1-\varphi)\theta - (1-\varphi) z_t.$$

When the mean forward rate curve is increasing, the effect of z on the term premium is positive. But when z rises, the expected change in the short rate falls (the effect of mean reversion), so the regression slope cannot be between zero and one.

In a one-factor affine world, the regression slope requires a term premium that varies inversely with the short rate. A relatively simple way to accomplish this while retaining a positively sloped average yield curve is with what we term the "negative CIR model": $\gamma = -1 + \lambda^2/2$ and δ is a free parameter (needed to generate a positive mean short rate). Then $A_1 = \delta$, $A_2 = 2\delta - (1-\varphi)\theta$, $B_1 = -1$, $B_2 = -1 - \varphi + \sigma(\lambda - \sigma/2)$. Thus $r_t = \delta - z_t$ (hence the label, "negative CIR"), $f_t^1 = \delta - (1-\varphi)\theta + [-\varphi + \sigma(\lambda - \sigma/2)]z_t$, $f_t^1 - r_t = \delta - (1-\varphi)\theta + [1 - \varphi + \sigma(\lambda - \sigma/2)]z_t$, and $E_t r_{t+1} - r_t = (1-\varphi)(z_t - \theta)$. The first term premium is

$$p_t^1 \;=\; (f_t^1 - r_t) - (E_t r_{t+1} - r_t) \;=\; \sigma(\lambda - \sigma/2) z_t.$$

Its mean is positive (implying an upward-sloping mean forward rate curve) if $\sigma(\lambda - \sigma/2) > 0$. The first regression slope is

$$c_1 \;=\; \frac{1-\varphi}{1 - \varphi + \sigma(\lambda - \sigma/2)},$$

which is between zero and one under the same conditions. Thus the model seems capable of resolving the tension between the regression slopes and the average slope of the forward rate curve.

This possibility is born out in Table 4, where estimated parameter values for the negative CIR model are reported as Model B. The J-statistic suggests that the model is a dramatic improvement over Model A. Note that the model generates both regression slopes between zero and one and a positive average spread between 10-year and 1-month yields. For a variety of reasons, however, the model cannot be the last word on the subject. One we regard as relatively innocuous: since z takes on all positive values, the short rate is negative with positive probability. In our example, the probability is about 0.5% (see Appendix B). Like Dai and Singleton (2000), Duffie and Singleton (1997), and Pearson and Sun (1994) in similar contexts, we regard the possibility of

negative interest rates a small price to pay for the convenience of a linear model. A more compelling reason is the one-factor structure: evidence suggests that we need two factors, and possibly more, to reconcile the curvature of the mean yield curve (Gibbons and Ramaswamy 1993) with the dynamics of interest rates (Garbade 1986, Litterman and Scheinkman 1991, Stambaugh 1988). The large J-statistic in Table 4 verifies that similar difficulties apply to our data set.

Multifactor Models

We estimate three multifactor models: a two-factor Cox-Ingersoll-Ross model, a two-factor "mixed" model with one positive and one negative factor, and a three-factor model with one negative and two positive factors that includes the others as special cases. The motivation for models with negative factors is the same as the one-factor case.

The primary difficulty in computing estimates of multifactor affine models is in identifying the separate elements of θ. In the two-factor Cox-Ingersoll-Ross model, for example, the sum $\theta_1 + \theta_2$ is the mean short rate, which we set equal to the sample mean. Experience suggests, however, that the values of the components, θ_1 and θ_2, are less easily estimated. For Models C, D, and E, we follow Chen and Scott (1993) in computing these parameters separately by grid search. Model D illustrates the process: We set the mean short rate, $\theta_2 - \theta_1$, equal to the sample mean and choose the components by grid search. The grid consists of 101 equally-spaced points for which this restriction holds and $\theta_1, \theta_2 \geq 0$. At each point on the grid, we compute the other parameters by minimizing the GMM objective function (the J-statistic) in the usual way. We then choose the point associated with the smallest value of the GMM objective function. We refine the grid further by taking the interval defined by the eleven points surrounding the minimum, dividing it into 101 equally-spaced points, and again choosing the point that minimizes the objective function. This process is repeated until we reach a grid size of less than 10^{-5}. The estimates in Table 4 are based on the resulting value of θ. The objective function is relatively flat with respect to θ in the neighborhood of the estimates, suggesting that the data are not very informative about these parameters. Reported standard errors are conditional on the choice of θ.

The five models in Table 4 indicate the benefits of a negative factor. Model C is a two-factor CIR model like that studied by Roberds and Whiteman (1999). It fits substantially better than the one-factor CIR model (Model A), but worse than the one-factor negative CIR (Model B). As Roberds and Whiteman note, the parameters that reproduce regression slopes fail along other dimensions. The same is true of our estimates. We

could report parameter values that come substantially closer to the regression evidence, but they would approximate other features of interest rates less well. The estimates in Table 4 are a compromise. The mean yield curve is steeper than we see in the data and the autocorrelation of the short rate is higher.

Model D is a "mixed" model, with one positive and one negative factor: $\gamma_1 = -1 + \lambda_1^2/2$ and $\gamma_2 = 1+\lambda_2^2/2$, making the short rate $r_t = z_{2t} - z_{1t}$. Its J-statistic is significantly smaller than C's, suggesting a better overall approximation to the behavior of interest rates. Again, we could report estimates that approximate regression slopes more closely if we were willing to accept poorer performance along other dimensions. The probability of a negative short rate is again about 0.5% (see Appendix B). The clear implication of this model, as well as the negative CIR model B, is that a negative factor permits a better approximation to the data, on the whole, than traditional CIR models with comparable numbers of parameters. In this sense, the negative CIR and mixed models are useful additions to the literature.

Model E is a three-factor model with one negative and two positive factors: $\gamma_1 = -1 + \lambda_1^2/2$, $\gamma_2 = 1 + \lambda_2^2/2$, and $\gamma_3 = 1 + \lambda_3^2/2$, so the short rate is $r_t = z_{2t} + z_{3t} - z_{1t}$. Its primary role here is to provide a comprehensive "encompassing" model that we can use to assess the others. The large standard errors suggest that we are close to the limits of what this data and moment conditions can tell us.

Sampling Variability and Small Sample Bias

Model D provides a useful interpretation of the regression evidence. In Figure 2, we compare regression slopes implied by Model D to those we estimated directly and reported in Table 1. The properties of the data are represented by asterisks. The implications of the model are represented by three lines computed from 1000 random draws of the parameters. In each one, we draw parameter values from the asymptotic multivariate normal distribution for the parameters. The solid line is the median from these 1000 replications and the two dashed lines are the 5% and 95% quantiles. The message is that the model provides a passable approximation to the estimated regression slopes at short maturities. At long maturities the numerical differences between model and data are small, but the sampling variability is even smaller.

The discrepancy between long-maturity regression slopes in the model and the data is a robust feature of these models when their parameters are chosen to reproduce observed properties of interest rates. Even when we add moment conditions for long-maturity regression slopes, estimated models imply regression slopes that are closer to one than we see in the data. We can reduce the rate of convergence of the slopes to one by choosing

autoregressive parameters φ_{ii} closer to one, but this invariably raises the unconditional standard deviation of the short rate or yield spreads well beyond their sample values.

Model D also provides a new perspective on small sample bias. We estimate the bias in regression slopes by simulating the model. Using estimated parameter values, we generate 1000 samples of 312 observations each for the state variables z_1 and z_2, from which we calculate forward rates. For each sample, we use simulated forward rates to compute regression slopes. We estimate the small sample bias by the difference between the mean regression slope across the 1000 replications and the population regression slope given by equation (15). Figure 3 suggests that this bias (solid line) can be substantial, especially at short maturities. The methods proposed by Bekaert, Hodrick, and Marshall (1997a), summarized in our equation (9), give a similar answer (dashed line). Both suggest that small sample bias can be an important problem for samples of the size used in this paper. Curiously, the positive bias makes the differences between theory and evidence even more striking.

We conclude that affine models with negative factors appear capable of reconciling slopes of forward rate regressions with other properties of interest rates. The J-statistics suggest that our best efforts leave some features of the data unexplained, but nevertheless provide reasonable approximations to the dynamics of interest rates and slopes of forward rate regressions at short maturities.

6 Predictable Changes Revisited

We have focused our attention on forecasts of one-period changes in forward rates based on equation (4). Here we examine three other popular forecasting relations, compare them to each other, and consider their implications for our estimated affine models.

Other Forecasting Relations

Some of the most popular assessments of interest rate dynamics involve the "yield regression"

$$y_{t+1}^{n} - y_{t}^{n+1} = \text{constant} + d_n \left(\frac{y_t^{n+1} - r_t}{n} \right) + \text{residual}, \tag{18}$$

a relation that dates back at least to Roll (1970) and perhaps to Macauley (1938). Recent empirical studies include Bekaert, Hodrick, and Marshall (1997a), Campbell and Shiller (1991), and Evans and Lewis (1994) for the US and Bekaert, Hodrick, and Marshall (1997b) and Hardouvelis (1994) for other countries. The expectations hypothesis

[equation (5)] implies $d_n = 1$ for all maturities n, but the equation is otherwise quite different from the forward rate regression (4) studied earlier. Since yields are averages of forward rates [equation (3)], slopes for long maturities include information about short-maturity forward rates and term premiums. We will see shortly that there is no theoretical presumption that the regression slope is one, even at long maturities.

Two other common approaches to interest rate dynamics are based on multiperiod forecasts of the short rate. One example is

$$r_{t+n} - r_t = \text{constant} + e_n(f_t^n - r_t) + \text{residual}, \tag{19}$$

which we refer to as the "Fama regression." Estimates are reported by Fama (1984), Fama and Bliss (1987), and Mishkin (1988). Another example is

$$\sum_{i=1}^{n} \left(1 - \frac{i}{n+1}\right)(r_{t+i} - r_{t+i-1}) = \text{constant} + g_n(y_t^{n+1} - r_t) + \text{residual}, \tag{20}$$

which has been estimated by Campbell and Shiller (1991) with US data and Bekaert, Hodrick, and Marshall (1997b) with data for the US, the UK, and Germany. We refer to (20) as the "Campbell-Shiller regression," although this is only one of many regressions they estimated. If we rewrite (20) as

$$(n+1)^{-1}\sum_{i=1}^{n}(r_{t+i} - r_t) = \text{constant} + g_n(n+1)^{-1}\sum_{i=1}^{n}(f_t^i - r_t) + \text{residual},$$

the similarity to (19) is apparent. The expectations hypothesis implies regression slopes of one for both (19) and (20). Like the yield regression, they incorporate information about short-maturity term premiums even when n is large.

We report estimates of all three relations — equations (18), (19), and (20) — in Table 5 for the Smoothed Fama-Bliss data. With the exception of Unsmoothed Fama-Bliss, the other datasets produce broadly similar estimates (not reported). The yield regression uses data from the complete sample. The others are estimated over samples that leave room for future values in constructing the dependent variable; at long maturities, we are missing up to 10 years of data. Estimates of these three relations, like those of (4), exhibit pronounced differences from the expectations hypothesis at short maturities. At long maturities, slopes of yield regressions get progressively more negative. The standard errors are large, but the estimated slopes are significantly different from one by conventional statistical standards. The Fama and Campbell-Shiller regression estimates differ in two respects from our earlier forward rate regressions: the coefficients do not increase monotonically with maturity and the standard errors are substantially larger at long maturities. The former is the result, in our view, of cumulating term premiums over several maturities. With (say) $n = 3$, the regression incorporates the effects of term

premiums for maturities 1, 2, and 3. The latter is the consequence, primarily, of the variability and overlap in multiperiod forecast errors: we do not have many independent observations of (say) 120-month forecasts of the short rate. The increased sampling variability in these regressions means we cannot make any precise statements about the behavior of regression slopes at long maturities. The point estimates are different from one, but the standard errors are 0.25 or larger. In contrast, the standard errors for the forward rate regression were in the neighborhood of 0.012 (see Table 1). One advantage, then, of (4) over (19) and (20) is that we have more precise information about changes over one period than n. Hodrick (1992) makes a similar point in a different context.

Comparison

Despite the apparent differences among the four regressions, we think they capture similar information. This is an exact statement for $n = 1$, approximate for longer maturities.

Consider the relation between yield and forward rate regressions. For $n = 1$, the yield and forward rate regressions contain exactly the same information: $d_1 = 2c_1 - 1$. Thus a value of 0.4557 for c_1 corresponds to –0.0886 for d_1. These two numbers are equivalent ways of representing the same information. For other maturities, there is no exact correspondence. For $n = 2$, the yield regression can be rewritten as

$$\left[(f^0_{t+1} - r_t) + (f^1_{t+1} - r_t)\right] = \text{constant} + \left(\frac{d_2 + 2}{3}\right)\left[(f^1_t - r_t) + (f^2_t - r_t)\right] + \text{residual},$$

which contains the elements of forward rate regressions for $n = 1, 2$. However, d_2 cannot be computed from the first two forward rate slopes, c_1 and c_2, alone. We need to know, among other things, the covariance between $f^1_t - r_t$ and $f^2_t - r_t$. More simply: yield and forward rate regressions do not contain identical information at longer maturities.

We can make further progress if we impose additional structure on the problem. A framework we find appealing is to assume that forward rates are linear functions of a single, autoregressive state variable z:

$$f^n_t = \text{constant} + \alpha_n z_t \tag{21}$$

$$E_t z_{t+1} = \text{constant} + \varphi z_t, \tag{22}$$

with the normalization $\alpha_0 = 1$. This linear one-factor structure includes one-factor affine models as special cases in which the infinite parameter set $\{\varphi, \alpha_1, \alpha_2, \ldots\}$ depends on a small number of primitive parameters. We know that one-factor models are inadequate for some purposes, but we also know that the first factor typically accounts for 80% or more of the variance of yield changes (Garbade 1986, Litterman and Scheinkman 1991).

Our hope is that (21,22) will serve as useful approximations that help us clarify the relations between regressions.

With this structure in hand, we can compute the slopes of yield regressions implied by forward rate regressions, and vice versa. Forward rate and yield regression slopes are related to the α's by

$$\begin{aligned} c_n &= \frac{\varphi\alpha_{n-1} - 1}{\alpha_n - 1} \\ d_n &= \frac{(n+1)\varphi A_{n-1} - nA_n}{A_n - (n+1)}, \end{aligned}$$

where $A_n = \sum_{i=0}^{n} \alpha_i$. The inverse relations are

$$\begin{aligned} \alpha_n &= \frac{\varphi\alpha_{n-1} - 1}{c_n} + 1 \\ \alpha_n &= \frac{A_{n-1}[(n+1)\varphi - n - d_n] + (n+1)d_n}{d_n + n}. \end{aligned}$$

Thus we can compute the α's implied by (say) the forward rate regressions and use them to compute the slopes of yield regressions implied by them. Here and below, we use slopes estimated with Smoothed Fama-Bliss data, set $\varphi = 0.959$ (the autocorrelation of the short rate in Table 3), and interpolate between missing maturities with a cubic spline.

The results of this exercise are reported in Figure 4. Note that yield regression slopes derived from slopes of forward rate regressions (solid line) are remarkably similar to those estimated directly (asterisks). In this sense, we would say the two regressions appear to contain approximately the same information. (We turn to the reverse comparison shortly.) One striking by-product of this exercise is a new perspective on the long end of the yield curve: the large negative slopes of yield regressions for long maturities seem to correspond to the numerically small differences in forward rate slopes from one. To see this, suppose we set the slopes of forward rate regressions exactly equal to one for maturities of 24 months or more. With this change, the implied slopes of yield regressions level off (the dashed line in Figure 4). Apparently the increasingly negative slopes of yield regressions at long maturities are closely related to the small differences from one of slopes of forward rate regressions. Yield regressions simply report this information in a way that magnifies the numerical difference from the expectations hypothesis at long maturities.

We can do a similar analysis of the Fama and Campbell-Shiller regressions. Their slopes imply (given our framework)

$$\alpha_n = \frac{\varphi^n - 1}{e_n} + 1$$

$$\alpha_n = \frac{\sum_{i=1}^n \varphi^i - n}{g_n} + n + 1 - A_{n-1}.$$

From them, we can compute (say) the implied slopes of forward rate regressions. The results are pictured in Figure 5, where we see that the three sets of implied slopes are similar to those estimated directly. The Fama and Campbell-Shiller estimates differ the most, exhibiting less smoothness and approaching one more quickly than the others. Whether this reflects sampling variability or something more fundamental is difficult to say. More interesting to us is that the rapid convergence to one apparent in this translation is invisible in the originals.

Affine Interpretations

The next question is whether our estimated two-factor affine model can account for these additional forecasting relations. For the yield regression, the affine model of Section 4 implies regression slopes of

$$d_n = \frac{[B_1 - (n+1)^{-1}B_{n+1}]^\top \Gamma_0 \, [\frac{n}{n+1}B_{n+1} - \Phi^\top B_n]}{[B_1 - (n+1)^{-1}B_{n+1}]^\top \Gamma_0 \, [B_1 - (n+1)^{-1}B_{n+1}]}$$

Its limiting value (if it exists) is

$$\lim_{n\to\infty} d_n = \frac{B_1^\top \Gamma_0 \, (I - \Phi)^\top B}{B_1^\top \Gamma_0 \, B_1},$$

where B is the limit of B_n. Although the slope converges, it need not converge to one. For the Fama and Campbell-Shiller regressions, the model implies slopes of

$$\begin{aligned} e_n &= \frac{[B_1 + B_n - B_{n+1}]^\top \Gamma_0 \, (I - \Phi^n)^\top B_1}{[B_1 + B_n - B_{n+1}]^\top \Gamma_0 \, [B_1 + B_n - B_{n+1}]} \\ g_n &= \frac{[B_1 - (n+1)^{-1}B_{n+1}]^\top \Gamma_0 \, [\frac{n}{n+1}I - (n+1)^{-1}\sum_{i=1}^n \Phi^i]^\top B_1}{[B_1 - (n+1)^{-1}B_{n+1}]^\top \Gamma_0 \, [B_1 - (n+1)^{-1}B_{n+1}]} \end{aligned}$$

Both have limiting values of one, although we will see shortly that they converge less rapidly than slopes of forward rate regressions.

In Figure 6, we illustrate the implications for these other regressions of the mixed two-factor Model D. The bottom panel is devoted to the yield regression, and we see that the regression slopes in the model (the lines) are similar to those estimated directly (asterisks) at short maturities. The difference at long maturities is another reflection of the rapid convergence of regression slopes in the model, regardless of the regression used. The top panel is devoted to the Fama regression (the Campbell-Shiller regression is so similar in this case that it clouds the figure). In this case, the numerical differences between model and data are substantial.

7 Final Remarks

We have characterized the predictability of interest rates in a novel way, related our approach to others, and designed and estimated a model that approximates many of the features of interest rate data. Along the way, however, we ignored a number of issues that deserve comment. One is our approach to forecasting interest rates. If the object is to forecast, not to test the expectations hypothesis, then we might want to consider different linear or even nonlinear forecasting relations. The success of multifactor bond pricing models suggests, in particular, that multivariate regressions might improve prediction. However, in preliminary efforts along these lines the improvements have been modest. Another issue is the use of forecasting relations for active investment strategies. The forward rate regressions suggest that one might want to invest long (short) when the yield curve is steep (flat or inverted), since expected future changes in interest rates are generally smaller than the additional premium. We found, however, that the mean premium on long investments is so large that the regression-based strategy is to invest long virtually all the time. A third issue is the performance of our mixed model. Models with negative factors apparently approximate the data better than those that do not, but there is room for further exploration in the affine class and otherwise. Perhaps additional work along the lines of Dai and Singleton (2000), allowing correlation between factors, would pay dividends here, too.

A Moment Conditions for GMM Estimation

The estimates in Table 4 are based on 11 moment conditions. The idea is to choose parameter values to make properties of the model (functions of the parameters) close to analogous properties of the data. The relevant properties of the model include means, variances, autocorrelations, and regression coefficients based on linear combinations $x = a^\top z + b$ of the state vector z. For each such linear combination, a and b are computed from the parameters of the model using (12) and the definitions of yields and forward rates, (1) and (2). For the model, the mean of x is

$$M(x) \;=\; a^\top \theta + b,$$

a function of the parameters. Similarly, the variance and autocorrelation of x are

$$\begin{aligned} V(x) &= a^\top \Gamma_0 a \\ R(x) &= \frac{a^\top \Gamma_1 a}{a^\top \Gamma_0 a}, \end{aligned}$$

where Γ_j is the j covariance, computed from (14) and the recursion $\Gamma_{j+1} = \Phi\Gamma_j$ for $j \geq 0$. In several cases, we use variances of differences, which are better behaved than unconditional variances for highly persistent series. The variance of the first difference of x is

$$V(\Delta x) \;=\; a^\top (2\Gamma_0 - \Gamma_1 - \Gamma_1^\top) a,$$

where $\Delta x_t = x_t - x_{t-1}$. Finally, the regression coefficient c_n is defined by (15). It should be understood that all of these properties (M, V, R, and c) are functions of the parameters alone.

Given these functions of the parameters, the moment conditions used to compute the estimates in Table 4 are of five types:

$$\begin{aligned} h_t^1 &= (r_t - R(y^1) r_{t-1})(r_{t-1} - \bar{r}_{t-1}) \\ h_t^2 &= s_t^n - R(s^n) s_{t-1}^n - [1 - R(s^n)] M(s^n) && (n = 12, 60, 120) \\ h_t^3 &= V(\Delta r) - (\Delta r_t)^2 \\ h_t^4 &= V(\Delta s^n) - (\Delta s_t^n)^2 && (n = 12, 60, 120) \\ h_t^5 &= [(f_t^{n-1} - r_{t-1}) - c_n (f_{t-1}^n - r_{t-1})][(f_{t-1}^n - r_{t-1}) - (\bar{f}_{t-1}^n - \bar{r}_{t-1})] && (n = 1, 6, 12) \end{aligned}$$

where $s_t^n = y_t^n - y_t^1$ denotes a yield spread and $\bar{x} = T^{-1} \sum_{t=1}^T x_t$ is the sample mean of x. These conditions are based, respectively, on the autocorrelation of the short rate, mean yield spreads, the variance of short rate changes, variances of spread changes, and forward rate regression slopes. The second set of moments (h^2) bases mean spreads on

autoregressions, rather than directly on means. We find that this reduces the autocorrelation in the moment conditions and leads to more precise estimates.

The weighting matrix is constructed by the Newey-West method with de-meaned moment conditions (Bekaert and Urias 1996). We follow the approach of Andrews (1991) in setting the lag truncation parameter equal to 4. We estimated AR(1)'s for each moment condition, computed Andrews' $\hat{\alpha}(1)$ from his equation (6.4) using unit weights, and calculated the optimal bandwidth from equation (6.2). The lag truncation parameter is bandwidth minus one, or 3.64. Another approach, based on an estimated VAR(1) for the moment conditions [equation (6.8) with weighting matrix equal to the identity], implies a lag truncation parameter of 3.86. Both approaches suggest that 4 is a reasonable choice.

B Negative Short Rates

The models we study involve independent state variables z that follow square-root processes. In continuous time (and approximately in discrete time) such state variables are gamma distributed with parameters (a, b) defined by

$$\begin{aligned} ab &= E(z) = \theta \\ ab^2 &= Var(z) = \sigma^2/(1-\varphi^2). \end{aligned}$$

The characteristic function is

$$\phi(z,t) = (1-ibt)^{-a}.$$

For Model B (negative CIR), $r = \delta - z$ and

$$\Pr(r<0) = 1 - \Pr(z>\delta) = 1 - F(\delta, a, b),$$

where F is the distribution function for z. We compute this directly using the gamma distribution function in Matlab, gamcdf(x,a,b).

For Model D (mixed model), we approximate the probability of a negative short rate by inverse transforming the characteristic function. The short rate is $r = z_2 - z_1$. Its characteristic function is

$$\phi(r,t) = \phi(z_2,t)\phi(z_1,-t) = (1-ib_2t)^{-a_2}(1+ib_1t)^{-a_1}.$$

Its distribution function F can be computed by inversion:

$$F(r) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{[e^{itr}\phi(r,-t) + e^{-itr}\phi(r,t)]}{it}dt.$$

We do the integration numerically in Matlab using procedure trapz. Other linear combinations can be handled similarly: the scale factor b incorporates the coefficient.

References

Andrews, Donald, 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* 59, 817-858.

Backus, David, 1993, "Bond pricing in discrete time," manuscript, New York University, July.

Backus, David, Allan Gregory, and Stanley Zin, 1988, "Risk premiums in the term structure: Evidence from artificial economies," *Journal of Monetary Economics* 24, 371-399.

Bekaert, Geert, Robert Hodrick, and David Marshall, 1997a, "On biases in tests of the expectations hypothesis of the term structure of interest rates," *Journal of Financial Economics* 43, 29-77.

Bekaert, Geert, Robert Hodrick, and David Marshall, 1997b, "'Peso problem' explanations for term structure anomalies," NBER Working Paper No 6147, August.

Bekaert, Geert, and Michael Urias, 1996, "Diversification, integration and emerging market closed-end funds," *Journal of Finance* 51, 835-869.

Bliss, Robert, 1997, "Testing term structure estimation methods," *Advances in Futures and Options Research* 9, 191-231.

Campbell, John, 1995, "Some lessons from the yield curve," *Journal of Economic Perspectives* 9 (Summer), 129-152.

Campbell, John, Andrew Lo, and Craig MacKinlay, 1996, *The Econometrics of Financial Markets*, Princeton NJ: Princeton University Press.

Campbell, John, and Robert Shiller, 1991, "Yield spreads and interest rate movements: A bird's eye view," *Review of Economic Studies* 58, 495-514.

Chen, Ren-Raw and Louis Scott, 1993, "Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates," *Journal of Fixed Income* 3 (December), 14-31.

Cox, John, Jonathan Ingersoll, and Stephen Ross, 1985, "A theory of the term structure of interest rates," *Econometrica* 53, 385-407.

Dai, Qiang, and Kenneth Singleton, 2000, "Specification analysis of affine term structure models," *Journal of Finance*, forthcoming.

Duffie, Darrell, and Rui Kan, 1996, "A yield-factor model of interest rates," *Mathematical Finance* 6, 379-406.

Duffie, Darrell, and Kenneth Singleton, 1997, "An econometric model of the term structure of interest rate swap yields," *Journal of Finance* 52, 1287-1321.

Dybvig, Philip, Jonathan Ingersoll, and Stephen Ross, 1996, "Long forward and zero-coupon rates can never fall," *Journal of Business* 69, 1-25; early versions of this paper were in circulation by 1987.

Elton, E.J., and T.C. Green, 1998, "Tax and liquidity effects in pricing government bonds," *Journal of Finance* 53, 1533-1562.

Evans, Martin, and Karen Lewis, 1994, "Do stationary risk premia explain it all? Evidence from the term structure," *Journal of Monetary Economics* 33, 285-318.

Fama, Eugene, 1984, "The information in the term structure," *Journal of Financial Economics* 13, 509-528.

Fama, Eugene, and Robert Bliss, 1987, "The information in long-maturity forward rates," *American Economic Review* 77, 680-692.

Fisher, Mark, and Christian Gilles, 1996, "Term premia in an exponential-affine economy," manuscript, Board of Governors.

Frachot, Antoine, and Jean-Philippe Lesne, 1994, "Expectations hypothesis and stochastic volatilities," manuscript, Banque de France, March.

Garbade, Kenneth, 1986, "Modes of fluctuation in bond yields," Topics in Money and Securities Markets No. 20, Bankers Trust, June; reprinted in K. Garbade, ed., *Fixed Income Analytics*, Cambridge: MIT Press, 1996.

Gibbons, Michael, and Krishna Ramaswamy, 1993, "A test of the Cox, Ingersoll, Ross model of the term structure," *Review of Financial Studies* 6, 619-658.

Hardouvelis, Gikas, 1994, "The term structure spread and future changes in long and short rates in the G7 countries," *Journal of Monetary Economics* 33, 255-283.

Hodrick, Robert, 1992, "Dividend yields and expected stock returns: Alternative procedures for inference and measurement," *Review of Financial Studies* 5, 357-386.

Litterman, Robert, and Scheinkman, Jose, 1991, "Common factors affecting bond returns," *Journal of Fixed Income* 1, 54-61.

Macauley, Frederick, 1938, *Some Theoretical Problems Suggested by Movements in Interest Rates, Bond Yields, and Stock Prices in the United States Since 1856*, New York: NBER.

McCulloch, J. Huston, and Heon-Chul Kwon, 1993, "US term structure data, 1947-91," manuscript and computer diskettes, Ohio State University.

Mishkin, Frederick, 1988, "The information in the term structure: Some further results," *Journal of Applied Econometrics* 3, 307-314.

Pearson, Neil, and Tong-Sheng Sun, 1994, "Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model," *Journal of Finance* 54, 1279-1304.

Roberds, William, and Charles Whiteman, 1999, "Endogenous term premia and anomalies in the term structure of interest rates: Explaining the predictability smile," *Journal of Monetary Economics* 44, 555-580.

Roll, Richard, 1970, *The Behavior of Interest Rates*, New York: Basic Books.

Sargent, Thomas, 1987, *Macroeconomic Theory (Second Edition)*, Orlando: Academic Press.

Stambaugh, Robert, 1988, "The information in forward rates: Implications for models of the term structure," *Journal of Financial Economics* 21, 41-70.

Sun, Tong-sheng, 1992, "Real and nominal interest rates: A discrete-time model and its continuous-time limit," *Review of Financial Studies* 5, 581-611.

Tauchen, George, 1986, "Statistical properties of GMM estimators of structural parameters obtained from financial market data," *Journal of Business and Economic Statistics* 4, 397-416.

Vasicek, Oldrich, 1977, "An equilibrium characterization of the term structure," *Journal of Financial Economics* 5, 177-188.

Table 1
Forward Rate Regressions in Four Data Sets

Maturity n	Smoothed Fama-Bliss	Unsmoothed Fama-Bliss	McCulloch Cubic Spline	Extended Nelson-Siegel
1	0.4557	0.5241	0.4984	0.5184
	(0.1851)	(0.1063)	(0.1760)	(0.1054)
3	0.7308	0.5758	0.7176	0.7979
	(0.0916)	(0.0872)	(0.0875)	(0.0932)
6	0.7971	0.6211	0.7833	0.8280
	(0.0570)	(0.0626)	(0.0569)	(0.0579)
12	0.8913	0.7643	0.8916	0.9036
	(0.0393)	(0.0567)	(0.0373)	(0.0398)
36	0.9576	0.6758	0.9544	0.9587
	(0.0172)	(0.0609)	(0.0167)	(0.0178)
60	0.9635	0.5765	0.9565	0.9624
	(0.0124)	(0.0668)	(0.0122)	(0.0125)
120	0.9634	0.6910	0.9655	0.9643
	(0.0102)	(0.0697)	(0.0102)	(0.0101)

Entries are estimated slope parameters c_n from

$$f_{t+1}^{n-1} - r_t = \text{constant} + c_n(f_t^n - r_t) + \text{residual},$$

where f^n is the n-month-ahead continuously-compounded one-month forward rate and $r_t = f_t^0$ is the short rate, both expressed as annual percentages. Forward rates were computed with data and programs supplied by Robert Bliss and come in four versions: Smoothed Fama-Bliss, Unsmoothed Fama-Bliss, McCulloch Cubic Spline, and Extended Nelson-Siegel. The data are monthly. For most entries, dates t run from January 1970 to November 1995 (311 observations). The exceptions concern the McCulloch Cubic Spline data, for which the starting dates are August 1971 for $n = 84$ and November 1971 for $n = 120$. Numbers in parentheses are Newey-West standard errors (6 lags).

Table 2
Estimated Standard Deviations of Measurement Error

		Difference from Smoothed-Fama Bliss		
Maturity n	McCulloch and Kwon	Unsmoothed Fama-Bliss	McCulloch Cubic Spline	Extended Nelson-Siegel
		A. Forward Rates		
0	0.11	0.16	0.22	0.19
1	0.13	0.23	0.19	0.15
3	0.17	0.30	0.12	0.09
6	0.18	0.46	0.18	0.07
12	0.21	1.13	0.18	0.08
36	0.13	1.62	0.21	0.06
60	0.09	2.01	0.24	0.08
120	0.13	3.52	0.65	0.18
		B. Yields		
1	0.11	0.16	0.22	0.19
3	0.05	0.07	0.10	0.03
6	0.04	0.06	0.06	0.04
12	0.04	0.09	0.07	0.03
36	0.02	0.07	0.06	0.04
60	0.02	0.08	0.08	0.03
120	0.03	0.17	0.12	0.06

Entries are estimated standard deviations of measurement error in forward rates and yields of different maturities, measured as annual percentages (the number 0.11, for example, corresponds to 11 basis points). The column labelled McCulloch and Kwon is the root mean square (the square root of the mean of the squares) of the estimated standard deviations reported by McCulloch and Kwon (1993) over the period January 1970 to February 1992. The remaining three columns are standard deviations of differences in forward rates from Smoothed Fama-Bliss estimates. The sample period in most cases in January 1970 to December 1995. The exceptions concern the McCulloch Cubic Spline data, which start in August 1971 for $n = 84$ and November 1971 for $n = 120$.

Table 3
Properties of Forward Rates and Yields

Maturity	Mean	Std Deviation	Autocorrelation
	A. Forward Rates		
0	6.683	2.703	0.959
1	7.098	2.822	0.969
3	7.469	2.828	0.968
6	7.685	2.701	0.966
12	7.921	2.495	0.969
36	8.498	2.135	0.979
60	8.714	2.013	0.980
120	8.858	1.946	0.980
	B. Yields		
1	6.683	2.703	0.959
3	7.039	2.781	0.971
6	7.297	2.774	0.971
12	7.544	2.672	0.970
36	8.009	2.373	0.976
60	8.253	2.224	0.978
120	8.529	2.073	0.981

Entries are sample moments of continuously-compounded forward rates and yields constructed with data and programs supplied by Robert Bliss (Smoothed Fama-Bliss method). The data are monthly, January 1970 to December 1995 (312 observations). Mean is the sample mean, Std Deviation the sample standard deviation, and Autocorrelation the first autocorrelation.

Table 4
Estimates of Multifactor Affine Models

Parameter	A	B	C	D	E
γ_1	$1+\lambda_1^2/2$	$-1+\lambda_1^2/2$	$1+\lambda_1^2/2$	$-1+\lambda_1^2/2$	$-1+\lambda_1^2/2$
δ		0.00857			
θ_1	0.00557	0.00300	0.00258	0.00150	0.00155
θ_2			0.00299	0.00707	0.00442
θ_3					0.00270
σ_1	0.00474	0.00415	0.00092	0.00907	0.00906
	(0.00082)	(0.00188)	(0.00788)	(0.00188)	(0.00209)
σ_2			0.00054	0.00516	0.00011
			(0.00522)	(0.00071)	(7.6948)
σ_3					0.00860
					(0.01668)
λ_1	–0.01476	0.06710	–0.04359	0.03332	0.03436
	(0.00196)	(0.00542)	(0.00774)	(0.00778)	(0.01215)
λ_2			0.00269	–0.03390	–0.08465
			(0.00783)	(0.00857)	(1.9170)
λ_3					–0.00127
					(4.5233)
φ_{11}	0.9519	0.9577	0.9261	0.9954	0.9938
	(0.00523)	(0.00827)	(0.01044)	(0.00264)	(0.00366)
φ_{22}			0.9999	0.4750	0.1166
			(0.00032)	(0.1611)	(21.762)
φ_{33}					0.4937
					(13.858)
J-statistic	87.46	45.35	65.22	6.31	4.70
Deg of Fr	8	7	5	5	2
p-Value	10^{-16}	10^{-7}	10^{-13}	0.277	0.095
c_1	1.442	0.387	0.661	0.834	0.831
c_{12}	1.030	0.975	0.958	0.949	0.953
$E(y^{120}-y^1)$	2.174	2.038	2.737	1.799	1.780
$Auto(r)$	0.952	0.947	0.9997	0.986	0.981

Parameter values were estimated by GMM using Smoothed Fama-Bliss yields and forward rates. Numbers in parentheses are standard errors. The sample period runs from

January 1970 to November 1995 (312 observations). Models differ in the number of factors and in the choice of γ_1 ($\gamma_i = 1 + \lambda_i^2/2$ for $i = 2, 3$). Thus the models are

- Model A: one-factor CIR.
- Model B: one-factor negative CIR.
- Model C: two-factor CIR.
- Model D: two-factor "mixed" model with one positive factor and one negative factor.
- Model E: three-factor "encompassing" model with two positive factors and one negative factor.

The 11 moment conditions are based on: the autocorrelation of the short rate, the mean of the spreads between long yields and the short rate ($y^n - r$ for $n = 12, 60, 120$), the variance of the monthly changes of the short rate and the same spreads, as well as the slopes of forward regressions (c_n for $n = 1, 6, 12$). The same moment conditions and weighting matrix were used to estimate all five models. The weighting matrix is based on Model E and was computed by the Newey-West method using "de-meaned" moment conditions as suggested by Bekaert and Urias (1996). The lag truncation parameter (4) is based on an estimated AR(1) along lines described by Andrews (1991). Further detail on the moment conditions and weighting matrix is given in Appendix A. J is the Hansen's J-statistic, Deg of Fr is its degrees of freedom, and p-Value is based on the asymptotic chi-squared distribution of J. The other statistics are regression slopes for n equal to 1 and 12, the mean 10-year spread, and the first autocorrelation of the short rate. All statistics are conditional on the θ's, which were chosen by grid search to minimize the J-statistic subject to the restriction that the model's mean short rate equals the sample mean.

Table 5
Other Forecasting Relations

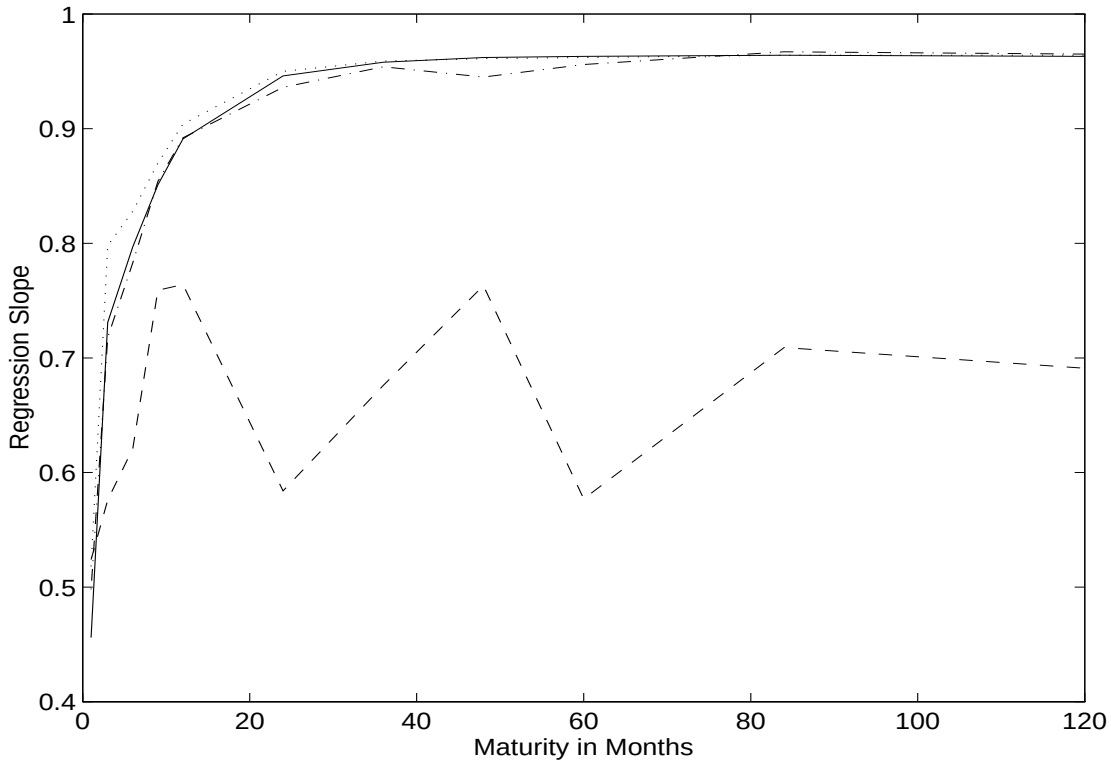
Maturity n	Yield Regressions	Fama Regressions	Campbell-Shiller Regressions
1	−0.0886	0.4557	0.4557
	(0.3702)	(0.1172)	(0.1172)
3	−0.4284	0.4108	0.4099
	(0.4808)	(0.1777)	(0.1833)
6	−0.8828	0.3624	0.3848
	(0.6398)	(0.1835)	(0.1529)
12	−1.4248	0.4590	0.4280
	(0.8249)	(0.2092)	(0.1763)
36	−1.9100	0.8431	0.6102
	(1.2954)	(0.2942)	(0.2635)
60	−2.4333	0.8513	0.7622
	(1.5190)	(0.2825)	(0.2352)
120	−4.1729	0.5988	0.7191
	(1.9847)	(0.2538)	(0.2582)

Entries are estimated slope parameters (d_n, e_n, g_n) from the regressions

$$\begin{aligned} y_{t+1}^n - y_t^{n+1} &= \text{constant} + d_n \left(\frac{y_t^{n+1} - r_t}{n} \right) + \text{residual} \\ r_{t+n} - r_t &= \text{constant} + e_n \left(f_t^n - r_t \right) + \text{residual} \\ \sum_{i=1}^{n} \left(1 - \frac{i}{n+1} \right) \left(r_{t+i} - r_{t+i-1} \right) &= \text{constant} + g_n \left(y_t^{n+1} - r_t \right) + \text{residual}, \end{aligned}$$

based on yields y^n and forward rates f^n computed with data and programs supplied by Robert Bliss (Smoothed Fama-Bliss method). We refer to them as yield regressions, Fama regressions, and Campbell-Shiller regressions, respectively. For the yield regression, dates t run from January 1970 to November 1995 (311 observations). For the other two regressions, dates t run from January 1970 to n months prior to December 1995 ($312 - n$ observations). Numbers in parentheses are Newey-West standard errors; for the yield regressions we use 6 lags, for the others we use n lags.

Figure 1
Slopes of Forward Rate Regressions in Four Data Sets

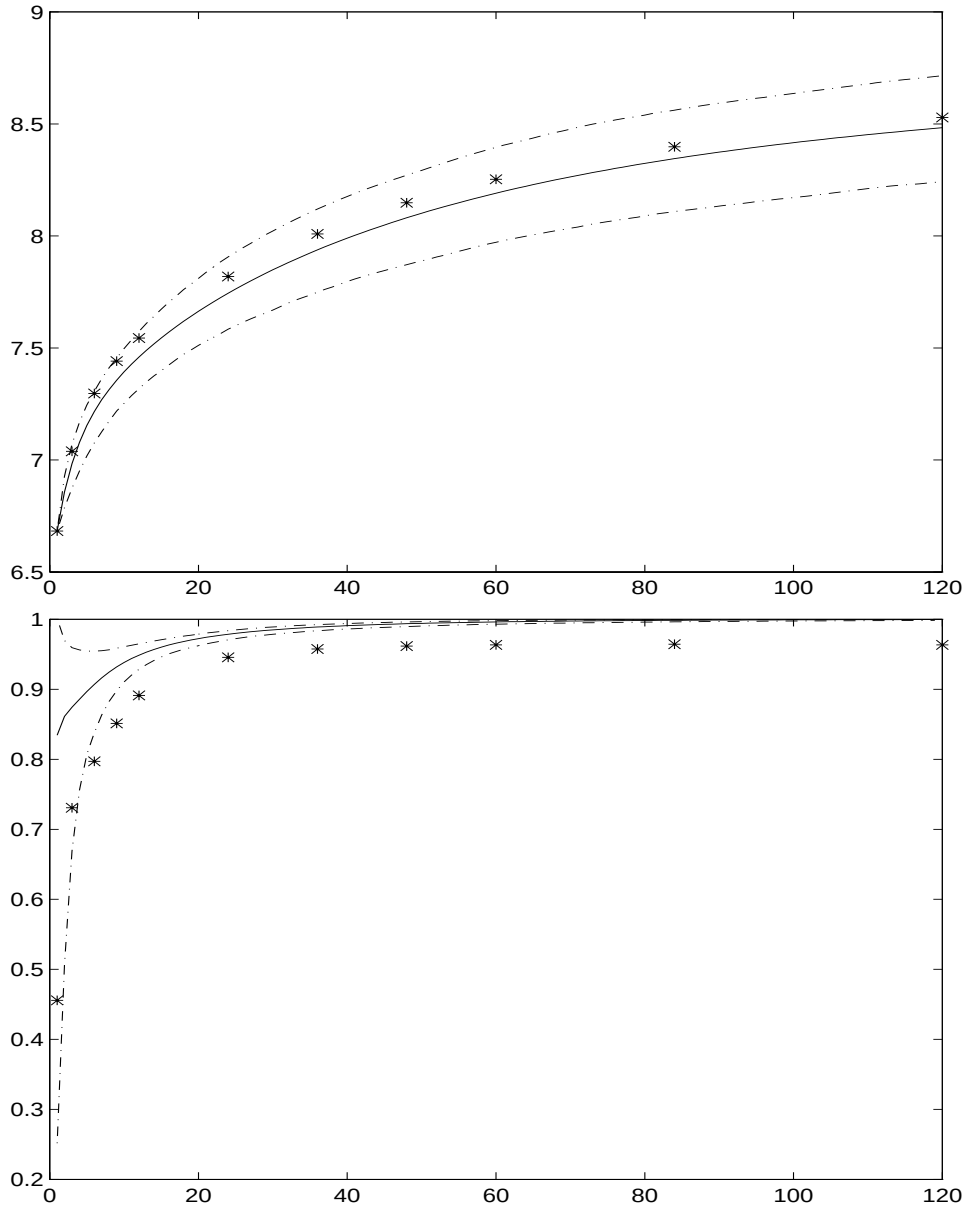


The figure summarizes in graphical form the estimates of slope coefficients c_n for forecasting relations based on forward rates,

$$f_{t+1}^{n-1} - r_t = \text{constant} + c_n(f_t^n - r_t) + \text{residual}.$$

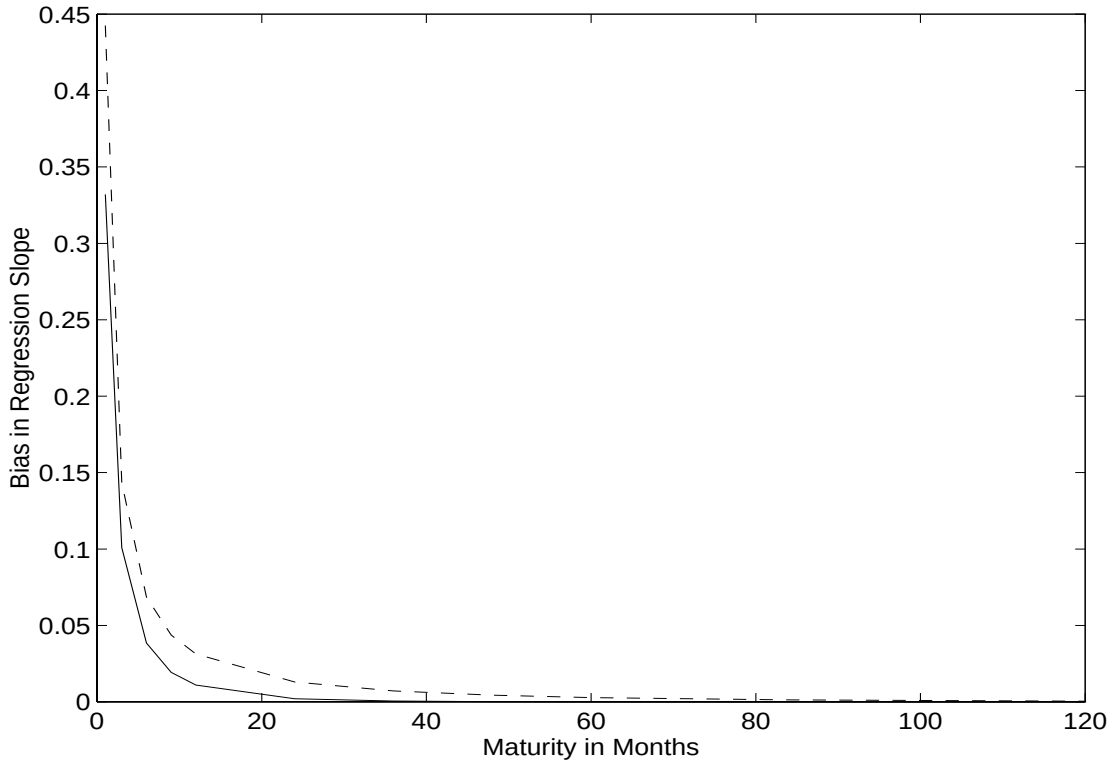
The same numbers are reported in Table 1 for a somewhat larger set of maturities. Lines represent slopes of forward rate regressions c_n for forward rates estimated by four different methods. The solid line is based on data estimated by the Smoothed Fama-Bliss method, the dashed line (the markedly different one) the Unsmoothed Fama-Bliss method, the dash-dotted line the McCulloch Cubic Spline method, and the dotted line the Extended Nelson-Siegel method.

Figure 2
Mean Yields and Regression Slopes in the "Mixed" Model



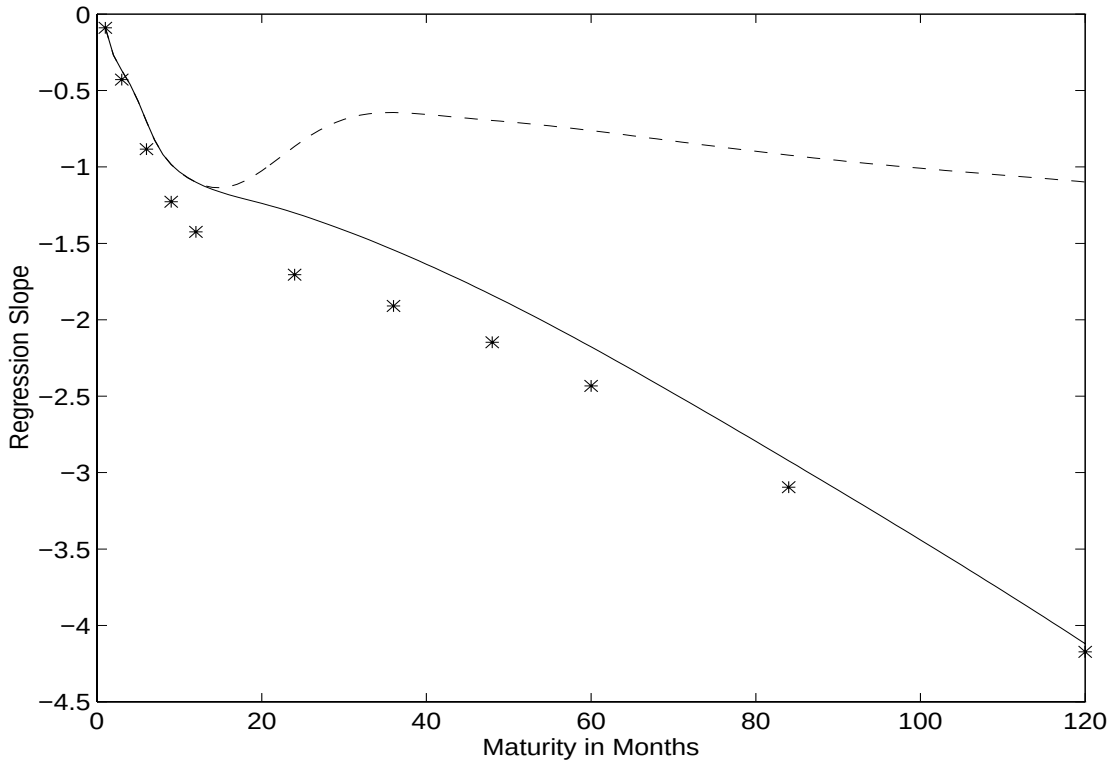
The two panels compare features of the data (asterisks, Smoothed Fama-Bliss data) with similar features of our estimated Model D (lines). The top panel is concerned with mean yields and illustrates the ability of the model to reproduce the slope and shape of the mean yield curve. The bottom panel is concerned with forward rate regression slopes. In each panel, the solid line represents the median from 1000 draws of the parameters from the asymptotic normal distribution of the parameters; dashed lines correspond to 5% and 95% quantiles, respectively.

Figure 3
Small Sample Bias in Regression Slopes



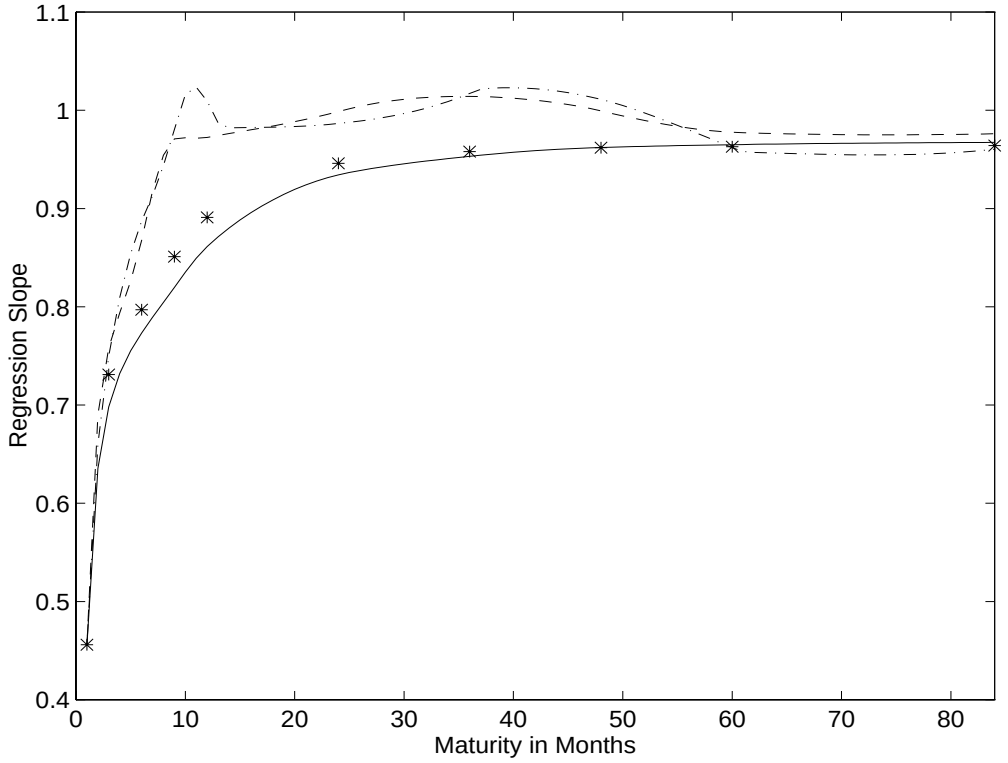
The lines illustrate small sample bias in slopes c_n of forward rate regressions. The solid line is the estimated bias for the mixed two-factor Model D: the difference between the mean from a Monte Carlo simulation and the true value implied by the model. The dashed line is a theoretical approximation to the bias, an application of equation (9) with $\varphi = 0.959$.

Figure 4
Yield Regressions Implied by Forward Rate Regressions



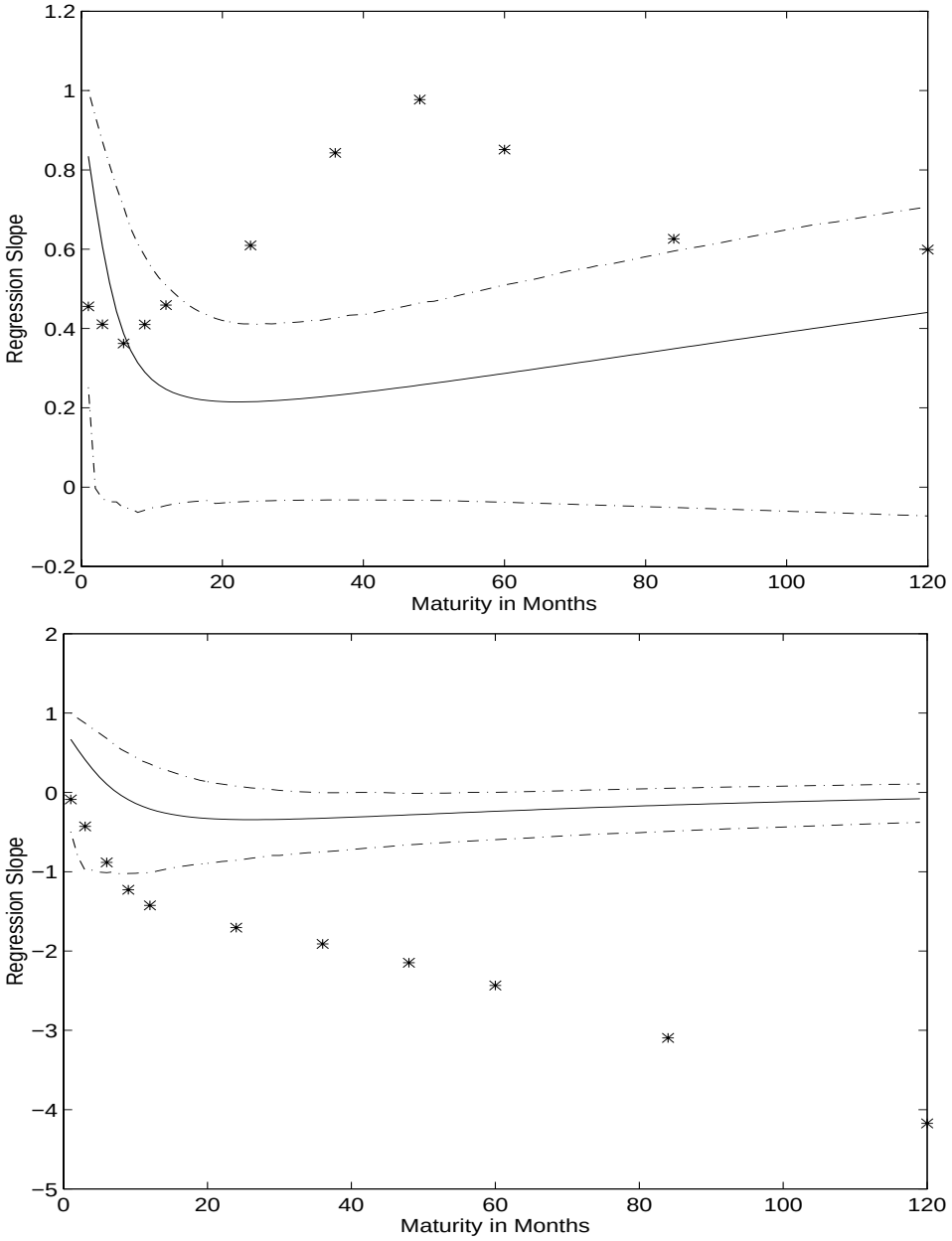
The figure compares estimated yield regression slopes (asterisks, Smoothed Fama-Bliss data) with those implied by forward rate regressions (lines) in a linear one-factor setting. The solid line represents slopes derived from forward rate regressions based on equations (21,22). The dashed line represents slopes derived from slopes of forward rate regressions that have been modified to equal one for maturities of 24 months or more. The point is that relatively small numerical differences between long forward rate regression slopes and one (Figures 1 and 2) correspond to much larger numerical differences in long yield regression slopes.

Figure 5
Forward Rate Regressions Implied by Other Regressions



The figure compares estimated forward rate regression slopes with those implied by other regressions in the linear one-factor setting of equations (21,22). Asterisks are direct estimates (Smoothed Fama-Bliss data). The solid line represents slopes derived from yield regressions, the dashed line represents slopes derived from Fama regressions, and the dash-dotted line represents slopes derived from Campbell-Shiller regressions. The point is that the slopes are broadly similar, suggesting that the various regressions contain similar information.

Figure 6
Other Regression Slopes in the "Mixed" Model



The two panels compare regression slopes implied by the mixed two-factor Model D with those estimated directly. The top panel is devoted to the Fama regression, equation (19), the bottom panel to the yield regression, equation (18). In each panel, the solid line represents the median from 1000 draws of the parameters from the asymptotic normal distribution of the parameters; dashed lines correspond to 5% and 95% quantiles, respectively.