THE JOURNAL OF FINANCE • VOL. LVII, NO. 3 • JUNE 2002

Empirical Analysis of the Yield Curve: The Information in the Data Viewed through the Window of Cox, Ingersoll, and Ross

CHRISTOPHER G. LAMOUREUX and H. DOUGLAS WITTE*

ABSTRACT

This paper uses recent advances in Bayesian estimation methods to exploit fully and efficiently the time-series and cross-sectional empirical restrictions of the Cox, Ingersoll, and Ross model of the term structure. We examine the extent to which the cross-sectional data (five different instruments) provide information about the model. We find that the time-series restrictions of the two-factor model are generally consistent with the data. However, the model's cross-sectional restrictions are not. We show that adding a third factor produces a significant statistical improvement, but causes the average time-series fit to the yields themselves to deteriorate.

THE SEMINAL TERM STRUCTURE MODEL of Cox, Ingersoll, and Ross (1985; CIR) has spawned numerous empirical investigations and theoretical generalizations. The model imposes restrictions on the cross-sectional and time-series properties of the yield curve. Although the model is cast in continuous time, the discrete-time transition densities of the state variables are known. This paper demonstrates how to estimate the model and integrate over the (latent) state variables using the appropriate transition density. In so doing, we allow all of the data to provide information about the state variables, and we take full advantage of the parametric implications of a continuous-time model for discretely sampled data. We approach the estimation problem as econometricians viewing data generated by a CIR world with noise. We implement a likelihood-based approach that does not concentrate the latent process into the likelihood. Under our approach, the conditional likelihood and the exact transition density of the latent state process—conditioned on the data—enable us to derive formal density functions of model parameters and (possibly complicated) functions of the parameters and state variables. By not identifying the state variables with specific yields in the sample, we are able to isolate the model's time-series implications from its cross-

* Lamoureux is from the University of Arizona, and Witte is from the University of Missouri. The authors thank John Cochrane, Qiang Dai, Rick Green, Dan Houser, René Stulz, and an anonymous referee as well as participants of seminars at the University of Arizona, University of California at San Diego, University of California at Riverside, University of Chicago, University of Illinois, University of Southern California, and University of Wisconsin, Duke University, Michigan State University, Northwestern University, Stanford University, and Washington University.

sectional implications. We can look at aspects of the term structure that are considered important and evaluate the performance of the model along these dimensions. For example, we find that (contrary to much published work) the two-factor CIR model does a very good job of fitting the time-series properties of short rates. It also does well at characterizing aspects of the time series of long rates. The biggest tension between the estimated CIR models and the data appears to be that long rates are too variable to be consistent with the model and the observed behavior of short rates. That is, it is the cross-sectional restrictions of the model (as opposed to its time-series properties) that stand at odds with the data.

As noted, we can construct the exact marginal density of any function of the model's parameters and state variables. This enables us to move model evaluation beyond a simple analysis of parameter values and portmanteau statistics, to a more general evaluation of how well the model conditioned on the data can characterize functions that traders and market analysts may be interested in. Such functions include bond prices and interest rate derivative prices, as well as functions of the parameters and state variables that govern the dynamics and cross-sectional properties of these prices. Professional traders are typically hesitant to use standard economic models of pricing because these models do not fit the yield curve exactly. Since the models start with primitives about endowments, preferences, and production functions, it is not surprising that they cannot fit any data exactly. Equilibrium/arbitrage models of the term structure such as CIR do not provide adequate degrees of freedom to afford an exact fit to a panel of yield data with more instruments than factors. However, the model is internally consistent and characterizes an economy in equilibrium with no arbitrage opportunities. If it were to characterize broad patterns in fixed income markets, it would yield fundamental insights into the sources and pricing of risk and show how derivatives should be priced.

While there have been many attempts to bring dynamic term structure models to data, little is known about the ability of these models to fit the data along dimensions that are important to traders. Most of the empirical literature compares models with different numbers of factors or alternative factor specifications using likelihood ratio or chi-squared portmanteau tests. These tests provide no indication as to the usefulness of the models in practice or the precision of the estimates when used to identify functions of the term structure model, such as interest rate derivatives. We provide an example of the extent to which the likelihood metric can be misleading. The three-factor model has a significantly higher marginalized likelihood value than the two-factor model and fits the cross-sectional functions of the yield curve better than the two-factor model. However, the two-factor model is better than the three-factor model in fitting the average yields in the sample.

In the two-factor model, we find that in levels the two factors are highly correlated with the short rate and the slope of the yield curve, respectively. However, when we look at log differences, the correlations between the fac-

tors and any observables are fairly low. We note that the model generates interest elasticities of volatility that are statistically indistinguishable from those in the data for several rates. We construct predictive densities for butterfly spread regressions and see that the coefficients on the shorter-term rate are smaller, and the coefficients on the longer-term rate are larger in the data than from the model. We also note that under the model, the first eigenvalue explains statistically more than its empirical counterpart in the data in both the two- and three-factor models. Finally, several recent papers have questioned the ability of dynamic term structure models, such as CIR, to explain the deviations in the data from the expectations hypothesis. We also look at this dimension of the model—in a sense combining the cross-sectional and time-series implications in a single metric. Here we see that the predictive density from the model actually departs more from the expectations hypothesis than does the data.

Dai and Singleton (2000a) generalize the class of affine term structure models and demonstrate that models with correlated factors are needed to fit swap spreads. They use Simulated Method of Moments to estimate the parameters of their models and compare the models. The findings of this paper complement Dai and Singleton. Our findings on the inconclusiveness of the number of factors complements studies that have documented that two independent factors are inadequate, but that three *independent* factors may overfit the data. More importantly, Dai and Singleton is primarily a model selection exercise, while we focus on the model's ability to explain and fit important properties of interest rates. *In addition*, we explicitly examine the amount of information that can be extracted from a panel of yield data using the CIR model for the pricing and hedging of derivatives.

A natural theoretical extension to term structure models would be to allow uncertainty as to the model parameters and factor levels on the part of market participants. The approach taken in this paper is to assume that the agents in the economy know the parameters and the current levels of the state variables. The econometrician knows neither and must infer them after observing the data. It might be more realistic to model the economic agents analogously to our econometrician. The methods introduced in this paper are a necessary first step in assessing the efficacy and viability of such an extension to the theory.[1]

The remainder of the paper is organized as follows. Section I reviews the problem of estimating term structure models. Section II lays out the model, the likelihood function, and the conditional densities used in the Gibbs sampler. Section III explains how we implement the Gibbs sampler. Section IV introduces the data used in the analysis. Section V contains the results. Section VI concludes the paper.

[1] Parameter and state variable uncertainty play an important role in modeling asset pricing for both equities as well as derivatives (Veronesi (1999) and David and Veronesi (1999), respectively.)

I. Literature Review

The substantive problems addressed by this paper are the estimation of the CIR model's parameters and latent factors and an analysis of CIR's ability to fit cross-sectional and time-series patterns in interest rate data. This model is a member of the class of dynamic term structure models (Dai and Singleton (2000a) provide a taxonomy of this class). Empirical analysis of this class of models has proceeded along a variety of different paths. One question is the role of errors. Wall Street practitioners often require that their specification fit the yield curve exactly. This is accomplished by adding degrees of freedom that also make the specification dynamically inconsistent. Another way to avoid adding errors to the model is application of string theory (Santa-Clara and Sornette (1999)), which uses an infinite dimensional state space (although with restrictions) to fit the yield curve exactly.

More traditional approaches specify a low dimensional state space and add errors to a term structure model so that the model is not expected to fit any yield curve exactly, but might explain broad patterns in both cross-sectional and dynamic aspects. Even this is not straightforward, since the most important term structure models specify non-Gaussian transition densities for the factors. This has the effect of rendering standard filtering methods inappropriate. To circumvent this problem and retain consistency, two approaches have been applied. In the first, it is assumed that J of the M bonds in the sample are observed without error. This has the effect of forcing the time-series restrictions of the model to hold and forcing the tension between the model and the data to the cross-sectional implications of the model. Duffie and Singleton (1997) offer an example of this approach. They find that a two-factor CIR model fits the term structure of swap rates. Pearson and Sun (1994) also use this approach in a study of monthly yield data on a variety of Treasury instruments over the period 1971 through 1986. They find that the model performs poorly: The sensitivity of parameter estimates to the different data sets is used as grounds for rejection. Also, they report that the model provides worse out-of-sample fit than a simple martingale specification for yields.

From an efficiency perspective, this procedure is unattractive because it assumes that the yields used as the factors are observed without error, but that the other yields in the sample are observed with error. In fact, all instruments are likely to contain some independent information about the state variable(s). This suggests that efficiency could be improved by using all of the information in the data to identify the state variables and test simultaneously. Furthermore, identifying a factor with a yield presupposes a very specific effect from adding additional factors. We show below that adding a third factor to the two-factor model improves the fit on the shortest and longest term yields in the sample, but provides a worse fit on the other three yields. If we were to identify a factor with a yield, it is less likely that we would observe this effect (if for no other reason that perforce, we would now fit three of the bonds exactly and have only two for overidentifying restric-

tions). Finally, Pearson and Sun's (1994) finding that the parameter estimates varied substantially as a function of the choice of instruments selected to identify the factors provides motivation for a more efficient identification scheme. Since such a choice is arbitrary, an approach that conditions on all of the data is indicated.

Estimation in this context may use a conditional likelihood function or generalized method of moments to ascertain whether the relationships among the data—both cross-sectionally and over time—are unusual given the model.[2] Similarly, one could use rank restrictions on the cross section of bonds, since under the model, the rank of this variance covariance matrix is the number of factors.[3] These approaches provide little insight into potential strengths and weaknesses of the model.

A second approach uses the Kalman filter to identify the state variable(s).[4] While this approach has the potential to be more efficient than the previous method, it suffers from a theoretical inconsistency. The state variables in CIR are not allowed to be negative. The Kalman filter uses a linear projection that assumes a normally distributed state variable to be optimal. However, the transition density of the state variable(s) in CIR is noncentral chi-squared. An additional critique of this method is that it does not provide insights into the strengths and shortcomings of the model. We might use a likelihood ratio test to reject a two-factor model against the alternative of a three-factor model, but how do we evaluate the three-factor model? It is not nested in a more general pricing framework. De Jong (2000) uses the Kalman filter to estimate an affine term structure model. He analyzes monthly yields on Treasury instruments over the period 1970 through 1991. He finds that a three-factor model—with correlated factors—improves the fit of the two-factor model, concluding that a three-factor model gives a nice fit to the data.

All of these methods add a separate model to CIR to allow econometric testing. This auxiliary model is simply a statement about the data. It says that under the null hypothesis, the data are equal to the fitted value from the CIR model plus a well-behaved error term. Recall that without this additional model, CIR would be solved via a just-identified system of equations and then rejected almost surely by failing to fit additional data.

As noted, our focus is on constructing predictive densities of functions of the term structure model. Decision makers can integrate over these densities in applications and use them to evaluate the fit of a particular model for a particular purpose. The error-augmented model states that any function of

[2] Gibbons and Ramaswamy (1993) is an example of the use of Generalized Method of Moments. Dai and Singleton (2000a) use Simulated Method of Moments.

[3] Stambaugh (1988) is an example of this approach. In addition, there is a large literature that focuses exclusively on the time-series restrictions of term structure models, such as CIR. It is common for this literature to examine the nonlinear dynamics of the short rate. Such research is an exploratory first step in evaluating a subset of the restrictions of a term structure model, since the cross-sectional restrictions of such models are more difficult to deal with econometrically than the time-series restrictions.

[4] Chen and Scott (1994) and DeJong (2000) are examples of this approach.

the term structure may be expressed as $f(\text{CIR}(\Theta) + \epsilon)$. Note the additive nature of the CIR model, which is a function of unknown parameters and state variables (collected in Θ), and the errors in the data, represented by ϵ, that have the cumulative distribution function $G(\epsilon)$. We will evaluate the exact predictive densities of $\int f(\text{CIR}(\Theta) + \epsilon)\, dG(\epsilon)$. Given the additive nature of the two models, this expectation is a way to isolate the fit of the CIR model, per se. We are also in a position to evaluate the cross-sectional versus the time-series properties of the model separately.[5]

Finally, many recent papers have looked at the time-series properties of a single interest rate over a long time horizon (e.g., Chan et al. (1992), Aït-Sahalia (1996), and Stanton (1997)). This literature documents important nonlinearities and, in general, more volatility of variance than would be expected under a CIR model.[6] We use a short horizon wherein Fed operating policy is essentially fixed, and we use a wide cross-section of interest rates on zero-coupon bills and strips.

II. The Model

CIR models an equilibrium, no-arbitrage economy, with a representative agent. From an econometric point of view, the model links the data on yields to one or more latent factors. The model posits that the time-series evolution of this latent factor is a mean-reverting, square-root process:

$$dz_j = \kappa_j(\theta_j - z_j)\,dt + \sigma_j\sqrt{z_j}\,d\omega_j, \tag{1}$$

where $j = 1,\ldots,J$ (the number of factors) and ω_j is a Wiener process.

CIR shows that the transition density for any z_j at time $t + s$ conditional on its realization at time t is given by (suppressing the j subscript)

$$f(z_{t+s}|z_t) = ce^{-u-\nu}\left(\frac{\nu}{u}\right)^{q/2} I_q(2(u\nu)^{1/2}), \tag{2}$$

where

$$c = \frac{2\kappa}{\sigma^2(1 - e^{-\kappa\cdot s})}, \qquad u = cz_t e^{-\kappa\cdot s}, \qquad \nu = cz_{t+s}, \qquad q = \frac{2\kappa\theta}{\sigma^2} - 1,$$

and I_q is a modified Bessel function of the first kind of order q.

Bond prices depend on the current value of the state variable, as well as its expected evolution, along with a risk premium, λ. Specifically, the price of a τ-year bond, at time t is

[5] The method of statistical model evaluation (essentially evaluating the viability of a model without comparing it to alternatives) by comparing functions of the data to a predictive (posterior) density is discussed in Meng (1994) and Gelman et al. (2000).

[6] Although the small sample properties of some of these estimators are questioned by Chapman and Pearson (2000).

$$P_{t,t+\tau} = \prod_{j=1}^{J} \Lambda_{j,t,\tau}\, e^{-\beta_{j,t,\tau} \cdot z_{j,t}}, \tag{3}$$

where

$$\Lambda_{j,t,\tau} = \left[\frac{2\gamma_j e^{[(\kappa_j+\lambda_j+\gamma_j)\tau]/2}}{(\kappa_j + \lambda_j + \gamma_j)(e^{\tau\gamma_j} - 1) + 2\gamma_j} \right]^{2\kappa_j\theta_j/\sigma_j^2},$$

$$\beta_{j,t,\tau} = \frac{2(e^{\tau\gamma_j} - 1)}{(\kappa_j + \lambda_j + \gamma_j)(e^{\tau\gamma_j} - 1) + 2\gamma_j}, \qquad \gamma_j = ((\kappa_j + \lambda_j)^2 + 2\sigma_j^2)^{1/2}.$$

For zero coupon bonds, the continuously compounded yield to maturity is

$$R_{t,t+\tau} = \frac{\sum_{1}^{J} (\beta_{j,t,\tau} \cdot z_j - \log \Lambda_{j,t,\tau})}{\tau}. \tag{4}$$

A. Likelihood Functions

Suppose that we have T observations on M bonds and J factors. If we assume that the data are measured with Gaussian errors, we may write the conditional likelihood function as

$$\mathcal{L}(\text{data}|z, \text{parameters}) \propto \prod_{t=1}^{T} |\Sigma|^{-1/2} e^{\epsilon_t' \Sigma^{-1} \epsilon_t} \tag{5}$$

$$\propto |\Sigma|^{-T/2} e^{\sum_{t=1}^{T} [r_t + A_t - B_t z_t]' \Sigma^{-1} [r_t + A_t - B_t z_t]}, \tag{6}$$

where ϵ is the M-vector of residuals, Σ is $E(\epsilon\epsilon')$, and r_t is an M-vector. A_t is an M-vector, with elements $a_{m,t} = (1/\tau_{m,t}) \sum_{j=1}^{J} \log \Lambda_{j,t,\tau_{m,t}}$. B_t is $M \times J$, with elements $b_{m,j,t} = (1/\tau_{m,t}) \beta_{j,t,\tau_{m,t}}$ and z_t is a J-vector. The reason for the t subscript on A and B is that there is no requirement that the bills and principal-only treasury strips (POs) in our sample have the same duration at each point in the sample. In fact, since long bonds are typically auctioned every six months, the durations of the three long-term POs in our sample vary somewhat over the period.

Unlike many models that have latent time-series processes, the parameters governing the evolution of the latent process enter directly into the *conditional* likelihood. Nevertheless, the change of measure (i.e., nonzero λ) is not identified in the conditional likelihood (note that only σ, $\kappa + \lambda$, and $\kappa \cdot \theta$ are identified there). The law of motion for the z is specified in the actual measure, however (equation (1) or (2)). Therefore, all parameters are identified in the *joint* likelihood:

$$\mathcal{L}(\text{data}, z \,|\, \text{parameters}) \propto \prod_{t=1}^{T} |\Sigma|^{-1/2} e^{\epsilon_t' \Sigma^{-1} \epsilon_t} \prod_{t=0}^{T} f(z_{t+1} | z_t). \tag{7}$$

The problem at this point is that the z_t are not observed. Obtaining the marginal distribution of the parameters under the likelihood is a daunting task because of the high dimensionality of the latent factor space. Since the number of bonds is larger than the number of factors, it is not necessary to impose factor orthogonality to identify the model, and we do not impose this orthogonality restriction in the estimation. When we examine the results, we will look at the predictive densities of the correlations between the factors. In this sense, orthogonality between the factors provides an overidentifying test of goodness-of-fit.

To use the likelihood to probabilistically describe parameters and other functions of interest, we have to integrate over the state space. Intuitively, these integrals will be obtained by treating (7) as the conditional density of random variable $z_{j,t}$—for example, taking the data, all parameters, and all other values of z as given. Proceeding with this approach through the entire state and parameter space generates a sequence of conditional densities. The average of these conditional densities forms the marginal predictive density. This type of Monte Carlo integration is referred to as the Gibbs sampler, and it has been shown to converge to the appropriate marginal distribution under mild regularity conditions (Geweke (1995a)).[7]

B. Conditional Densities for the Gibbs Sampler

As noted above, the Gibbs sampler works by taking sequential conditional draws from the parameter and state space.[8] We construct marginal predictive densities from these conditional draws. In the conditional densities described below, we treat the particular parameter or state variable as a random variable. When taking the relevant draw, we condition on the data and the most recent draws of the other parameters and state variables. A "draw" refers to the step when a single parameter or state variable is obtained by simulating a single variate from the full conditional distribution function for that variable. An "iteration" refers to a sequence of such draws over the entire state and parameter space. Thus, a single iteration from the two-factor model involves taking eight parameter draws, a draw from Σ, and $(T + 2) \times 2$ draws of z. The key to obtaining the full conditional density of a parameter or state variable is to view the joint likelihood (7) as a kernel (proportional to the density) for that variable.

[7] It may appear odd that we introduce the Gibbs sampler, which is simply a technique for doing the computations, prior to presenting the conditional densities. But it is because the Gibbs sampler obtains the marginal predictive densities by repeated sequential draws from the full conditional densities that our attention shifts to these conditional densities. Were we obtaining the predictive using direct analytics, we would not need the conditional densities.

[8] For an introductory treatment of the Gibbs sampler, see Casella and George (1992). For conditions under which the Gibbs sampler converges to the marginal predictive density, see Geweke (1995a).

B.1. Conditional Density of State Variable(s)

Since z is a first-order Markov process, the conditional density for z_t, given all other z variables, the parameters, and the data is (from the law of total probability)

$$f(z_t \mid \cdot) \propto f(r_t \mid z_t) f(z_t \mid z_{t-1}) f(z_{t+1} \mid z_t). \tag{8}$$

The first kernel on the right-hand side is the conditional likelihood function on date t, viewed as a function of the random variable z_t, with the data, all other parameters, and the values of the other state variable(s) at t (for the multifactor model), and the state variable(s) at all other t treated as fixed (equation (6)). The second and third kernels of $f(z_t)$ are noncentral chi-square densities (equation (2)). Note that z_t has support only on the positive real line.[9] We integrate over the starting and ending conditions, by drawing z_0 conditional on the current draw of z_1. We draw z_{T+1} conditional on z_T, using (2).

B.2. Residual Variance-Covariance Matrix

If we view Σ as the random variable in the joint likelihood (7), we have a standard inverse Wishart density [i.e., $\Sigma^{-1} \sim W(T, \sum_{t=1}^{T} \epsilon_t \epsilon_t')$]. Note that in dropping all unnecessary terms in (7) that do not involve Σ, we see that Σ is identified in the conditional likelihood (6).

For our case, Σ is a 5×5 matrix. Draws from this multivariate density are made by recognizing that the Choleski factorization may be obtained by taking univariate draws from a χ^2 density for the diagonal elements and a univariate normal density for the off-diagonal elements (see Muirhead (1982), p. 99).

B.3. Model Parameters

Identification of κ, λ, σ, and θ involves the following full conditional densities (presented here explicitly for κ, but synonymous for the other three parameters):

$$f(\kappa \mid \text{data, all other parameters}) \propto \mathcal{L}(\text{data} \mid \kappa, z, \cdot) f(z \mid \kappa, \cdot).$$

The density above is evaluated by viewing κ as the random variable and everything else as fixed. Also, in the case of λ, the last kernel is a constant (there is no information in the time series of z about λ) and is dropped for the purpose of obtaining the conditional density $f(\lambda \mid \text{data, all other param-}$

[9] This provides a contrast to the use of the Kalman filter where a linear projection is used to concentrate z_t into the likelihood function. There is no restriction that z be nonnegative, and the linear projection is optimal only for the case of normality, which is inappropriate here. Duffie and Singleton (1997, p. 1300) add a constant to the factor to ensure nonnegativity of the state variable. Since our state variables are drawn from the noncentral χ^2 density, they must be positive—obviating the need for such a constant.

eters). These conditional densities must be evaluated numerically, as discussed below. Therefore, as with z, we have to evaluate these in a univariate way: In the case of the two-factor model, we take a draw from κ_1 conditional on κ_2 (and the rest of the parameters and the state variables), and similarly for κ_2, and so on.

Note that unlike the other parameters in (1), σ is identified in the conditional likelihood (6). We could therefore evaluate this parameter, without including the effect of σ_j on the z_j time series in the conditional predictive density. We choose to include the z in learning about σ here to preserve symmetry with the other parameters—using all information from the time series and cross section of the data to identify the parameters.[10]

III. Implementing the Gibbs Sampler

With the exception of Σ, which is inverse Wishart, the implication of the joint likelihood for the distributions of the parameters and the state variables are nonstandard. We use a numerically intensive procedure to take draws from these conditional densities. The procedure requires evaluating the kernel (discussed in the previous section) at a series of points along a grid to identify an appropriate range. Next, we create an unequally spaced grid within this range, with more points centered around the maximum. The ordinate in this grid is normalized and linearly interpolated to construct a cumulative density function (cdf). An inverse cdf method is used to map a uniform [0, 1] draw onto this cdf. Simulations suggest that a local linear interpolation method is a robust way to approximate the cdf (as opposed to a spline method, which tends to degenerate in the tails). This method avoids the need to evaluate any integrals numerically. It takes roughly 25 seconds on a Sun Ultra 10 to obtain one complete iteration from the parameter space for the two-factor model. This procedure is called the "griddy Gibbs sampler," in Ritter and Tanner (1992).

A nice feature of the Gibbs sampler is that we can store the draws from each iteration and return to these later to evaluate functions of interest. The predictive densities of these functions of interest are the set of values obtained by evaluating the function at each iteration from the entire state variable and parameter space. This forms a Markov chain for any parameter or function of interest, and we can evaluate the numerical accuracy by examining the serial dependence within the Markov chain (e.g., by comparing the spectral density at 0 frequency of this chain to the variance of the iterations).

[10] In an earlier draft of this paper, the conditional densities of σ were not dependent on z—they were only obtained from the conditional likelihood. The qualitative results in that version were virtually identical to those in this version. The predictive densities on σ are tighter in this draft, but the other parameters are a bit more diffuse. The reason this parameterization does not have much of an effect is that the z after all are not observed, and the likelihood is itself conditioned on the z.

To eliminate dependence on starting conditions, we take a burn-in set of iterations before keeping any iterations to form the set of predictive densities. Standard practice is to use a burn-in of around 1,000 draws. We took 9,000 burn-in iterations in all cases, recognizing the high dimensionality of the state space and slow evolution of the state variables. For the analysis that follows, we have 12,500 iterations, post burn-in in all cases.

IV. Data

The data are hand collected from the *Wall Street Journal*. We have the ask yields-to-maturity for the 13-week and 26-week Treasury bills, along with approximate 5-, 15-, and 25-year Stripped Principal T-Bonds (POs). The durations of the Principal Strips vary throughout the sample. This poses no problem for the analysis, as the appropriate term to maturity for each instrument is always used in the conditional likelihood function. The data are at a weekly frequency, from each Wednesday covering the period June 14, 1989, through February 28, 1996. For the one year in the sample when Christmas and the following New Year fell on Wednesday, we use data from the preceding Tuesday. In describing the functions below, we refer to the mth shortest-term yield in our sample on date t as $r_{m,t}$.

V. Results

We present the results in three stages to highlight the insights that our estimation approach affords. First, we look at the goodness of fit of the one-, two-, and three-factor models. Next, we examine the relationship between the model and the data in more detail. Last, we identify the amount of information in the interest rate data to characterize (or to define *with precision* a density for the function of interest from the likelihood and data about) derivative prices and elasticities.

A. One-Factor Model

Figures 1 and 2 show the time series of the 90-day T-bill and 25-year PO yields, respectively, along with the 99th percentile bands of the predictive densities for these yields. As noted above, we use the entire sample to construct predictive densities on the model parameters, as well as the state variable(s). Within the context of a single iteration from this entire space in the Gibbs sampler, there is an implication for any interest rate on each date in the sample (which is simply a deterministic function of the parameters and state variables). There are 12,500 such iterations, which comprise the marginal predictive density on any function of the model, such as the interest rate in question. In the parlance of Bayesian econometrics, the posterior on an observable is usually referred to as a predictive density (Geweke (1994)). The figures overlay characteristics of the interest rate predictive density on

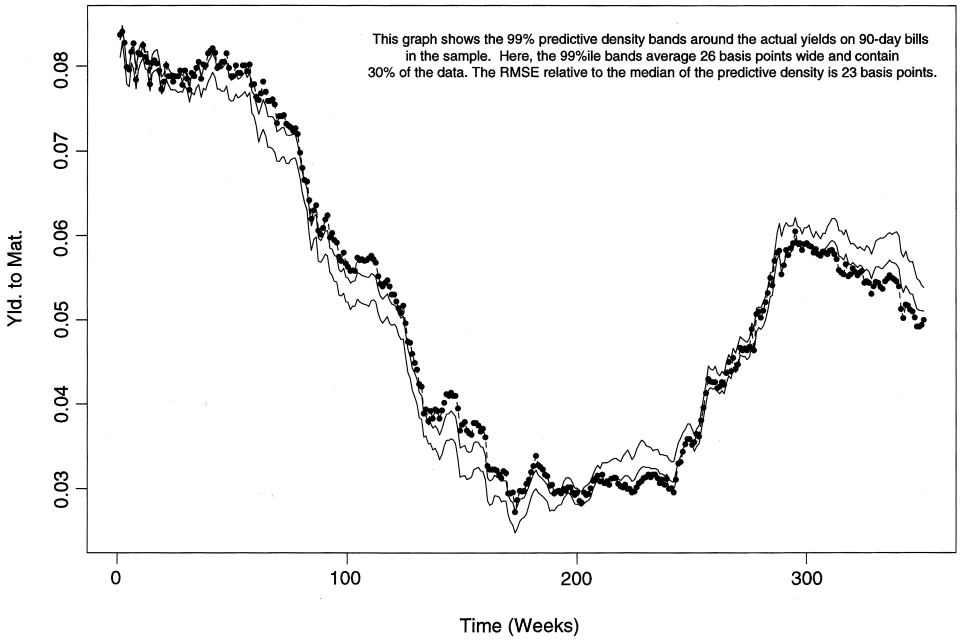

Figure 1. Time series of 90-day bill yield. Ninety-ninth percentile predictive density bands from the one-factor CIR model, overlaid the data, which are weekly observations of yields on 90-day U.S. Treasury bills.

each date in the sample on actual data. These predictive densities are marginal densities for the observables (e.g., yields), obtained by integrating over the predictive densities of the parameters and state variables. In comparing the predictive densities to the data, we have argued that the data are measured with error, so the predictive densities must integrate over the error space. This can be accomplished by evaluating the following integral at each Gibbs draw: $\int f(\text{model} + \epsilon)\, dG(\epsilon)$. When f is a linear function, this integral is tantamount to ignoring the errors. When it is nonlinear, we must evaluate the integral directly. In this manner, the functions of interest are themselves expectations from the joint model (CIR cum errors). These predictive densities are different from confidence intervals in that they characterize the probabilistic structure of the CIR model only (i.e., the expectation taken over the error space).

The one-factor model does a reasonably good job of fitting the short rates in sample, as demonstrated in Figure 1. The 99th percentile bands of the predictive density average 26 basis points wide. These bands contain 30 percent of the data. The sample root-mean-square error (RMSE) between the data and the median of the predictive density on each date is 23 basis points. However, the one-factor model is clearly inadequate to explain the long rates. The predictive density on the 25-year PO (Figure 2) is almost horizontal, reflecting the mean reversion of the model, and the dominant effect of the

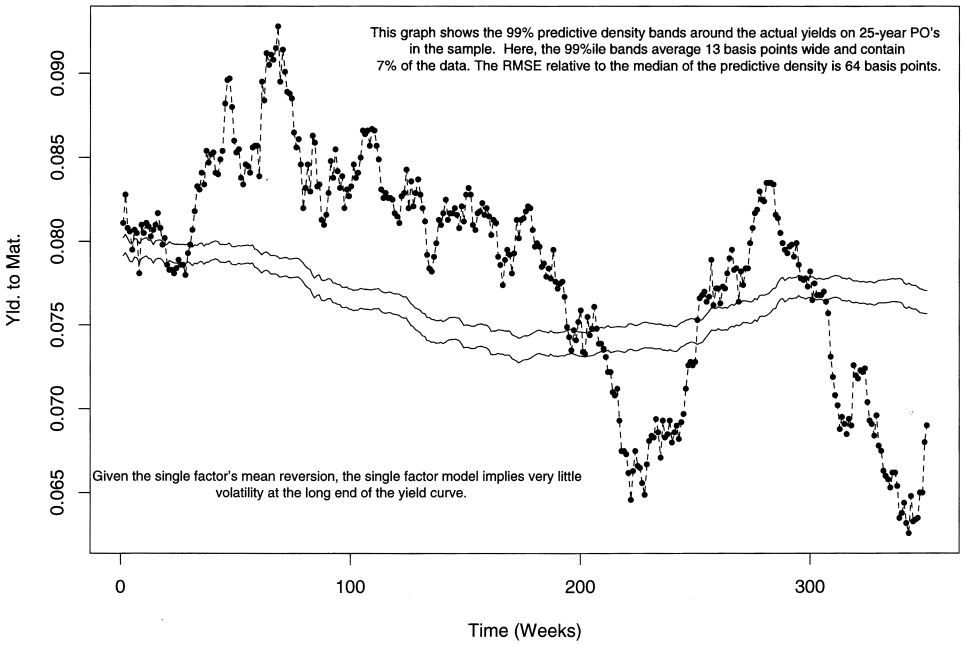

Figure 2. Time series of 25-year bond yield. Ninety-ninth percentile predictive density bands from the one-factor CIR model, overlaid the data, which are weekly observations of yields on 25-year principal only strips of U.S. Treasury bonds.

long-run mean of the state variable.[11] Figure 2 provides a context to evaluate the econometric treatment of the two separate models (and the role of the joint likelihood function). It is interesting that the joint likelihood attributes most of the inability of the CIR model to fit long yields to errors. Thus, there is adequate information in the short rate and cross-sectional restrictions of CIR to precisely identify the parameters and state variables: The failure of CIR to describe long rates did not create uncertainty about them. This highlights the importance of model evaluation using predictive densities—under the one-factor model we have little uncertainty as to where long rates are and how they behave, but the model is obviously a bad way to fit the data.

Table I shows properties of the predictive density of the error variance–covariance matrix from the one-factor model. Each iteration from the set of conditionals on the parameters includes a draw from Σ. The diagonal elements in Table I (in bold) show the distributional properties of the square-root of the diagonal elements of Σ. The off-diagonal elements of Table I show the distributional properties of the correlations implicit in Σ. As with all

[11] Several earlier empirical studies are more favorably disposed to the one-factor model, but it should be noted that such studies typically do not have any instruments with durations exceeding five years. By extending this to 25 years, we highlight an important problem with ignoring the long end of the yield curve.

Table I

Predictive Density of Variance–Covariance Matrix One-, Two-, and Three-Factor Models

Diagonal (bold) entries are square roots of the diagonal entry in the variance–covariance matrix of the error terms. Off-diagonal entries are correlations between residuals. The residuals are the result of adding a model of independent, identically distributed, Gaussian errors in the data to the CIR model to generate a likelihood function. The predictive density is the likelihood function—conditioned on the data—viewed as a density function for Σ and marginalized over all other unknown parameters and state variables.

	One-Factor					Two-Factor					Three-Factor				
Element	5%ile	10%ile	mean	90%ile	95%ile	5%ile	10%ile	Mean	90%ile	95%ile	5%ile	10%ile	Mean	90%ile	95%ile
1,1	**0.0019**	**0.0020**	**0.0024**	**0.0029**	**0.0030**	**0.0007**	**0.0008**	**0.0011**	**0.0014**	**0.0014**	**0.0006**	**0.0007**	**0.0010**	**0.0014**	**0.0015**
1,2	0.6789	0.7243	0.7992	0.8642	0.8744	−0.0275	0.0955	0.4454	0.7445	0.7807	0.5400	0.6316	0.8396	0.9676	0.9743
1,3	0.7884	0.8179	0.8771	0.9250	0.9347	0.1380	0.2630	0.5746	0.8366	0.8641	0.4519	0.5569	0.7961	0.9480	0.9574
1,4	0.8446	0.8726	0.9226	0.9592	0.9645	0.0051	0.1083	0.4291	0.7126	0.7506	0.5378	0.6228	0.8193	0.9455	0.9546
1,5	0.8521	0.8773	0.9208	0.9529	0.9581	−0.3694	−0.2675	0.0048	0.2643	0.3222	0.3714	0.4763	0.7361	0.9107	0.9264
2,2	**0.0024**	**0.0026**	**0.0030**	**0.0035**	**0.0036**	**0.0015**	**0.0016**	**0.0020**	**0.0025**	**0.0026**	**0.0021**	**0.0022**	**0.0027**	**0.0032**	**0.0033**
2,3	0.9480	0.9570	0.9724	0.9853	0.9876	0.9117	0.9223	0.9468	0.9682	0.9715	0.9427	0.9485	0.9639	0.9772	0.9799
2,4	0.8976	0.9110	0.9410	0.9684	0.9733	0.7957	0.8216	0.8859	0.9360	0.7456	0.9295	0.9388	0.9574	0.9734	0.9768
2,5	0.8578	0.8754	0.9124	0.9457	0.9522	0.1243	0.1800	0.3665	0.5421	0.5822	0.8764	0.8909	0.9229	0.9513	0.9567
3,3	**0.0072**	**0.0073**	**0.0078**	**0.0082**	**0.0083**	**0.0030**	**0.0031**	**0.0035**	**0.0040**	**0.0041**	**0.0029**	**0.0030**	**0.0035**	**0.0040**	**0.0041**
3,4	0.9672	0.9687	0.9732	0.9775	0.9785	0.8620	0.8818	0.9279	0.9631	0.9678	0.9607	0.9656	0.9781	0.9887	0.9904
3,5	0.9340	0.9367	0.9452	0.9533	0.9553	0.1863	0.2342	0.3916	0.5473	0.5853	0.9566	0.9627	0.9743	0.9840	0.9859
4,4	**0.0077**	**0.0078**	**0.0082**	**0.0086**	**0.0088**	**0.0014**	**0.0014**	**0.0018**	**0.0021**	**0.0022**	**0.0018**	**0.0019**	**0.0023**	**0.0027**	**0.0029**
4,5	0.9826	0.9834	0.9857	0.9878	0.9883	0.4824	0.5186	0.6309	0.7289	0.7477	0.9309	0.9394	0.9574	0.9724	0.9749
5,5	**0.0061**	**0.0062**	**0.0065**	**0.0068**	**0.0069**	**0.0012**	**0.0013**	**0.0015**	**0.0017**	**0.0018**	**0.0015**	**0.0016**	**0.0019**	**0.0022**	**0.0023**

parameters, the density function is formed by treating each of the Gibbs iterations as a draw from the marginal posterior. Note that the average standard deviation of the error on a 90-day bill yield is 24 basis points. The maximum average error size occurs at the 15-year PO yield (82 basis points). The noise terms for all five yields are highly intercorrelated. On average, the correlation between the unexplained yield on a 15-year bond and a 25-year bond is 98.6 percent.

The main result here is the predilection of the one-factor model to fit short rates. The means (from the predictive densities) of the standard deviation of the three bond yields are more than twice as large as those on the two bills. There is too much volatility in long rates to be consistent with short rates and a one-factor CIR model. The model's prediction of very stable long rates is strongly at odds with the data. This result suggests that a one-factor CIR model cannot characterize bond prices in the United States.

B. Two-Factor Model

As noted above, the two-factor model used here is the special case in CIR where the factors are orthogonal (see Dai and Singleton (2000a) for a discussion of these identifying restrictions). Nevertheless, the parameters (and state variables) are identified without imposing orthogonality on the state variables. We assume that the data contain unmodeled noise that is multivariate normally distributed and independent over time, as in the one-factor case. Table I also contains the predictive density of the "noise" standard deviation/correlation matrix from the two-factor model. Note that the diagonal entries for the long bonds are roughly one-third as large as in the one-factor case. Also, the off-diagonal entries are materially smaller than in the one-factor case. This evidence suggests that the errors from the two-factor model behave more like their theoretical counterparts than in the one-factor case. The errors are largest for the five-year PO strip yields. The mean standard deviation of 35 basis points for this instrument's yield errors seems high relative to the bid-ask spread (expected to be less than 15 basis points), but is still less than half the size of the one-factor model's.

Figures 3 and 4 show the ability of the two-factor model to fit the short and long end of the yield curve, respectively. Clearly, the two-factor model improves on the one-factor model's fit of the long bond yields. For the six-month bill rates, this two-factor model generates 90th (99th) percentile bands of the predictive density that are 18 (26) basis points wide on average, and these model bands contain 33 percent (43 percent) of the data. The RMSE of the 180-day bill rates relative to the median of the predictive density is 19 basis points. Figure 4 shows that the two-factor model clearly has enough flexibility to fit the overall pattern of long rates in the sample, which is in contrast to the one-factor model. Of the 351 yields in the sample, 163 do not fall within the 95 percent bands of the predictive density. In these cases, 84 lie below the 2.5 percentile bound of the predictive density, with an average distance of 7.5 basis points, and 79 lie above the 97.5 percentile predictive

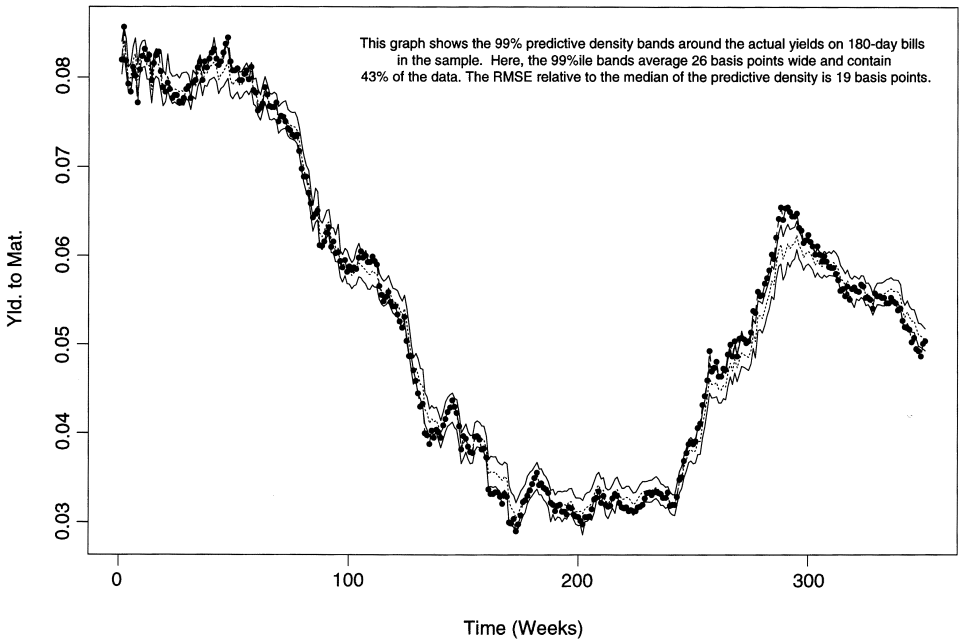

Figure 3. Time series of six-month bill yield. Ninety-ninth percentile predictive density bands from the two-factor CIR model, overlaid the data, which are weekly observations of yields on 180-day U.S. Treasury bills.

density bound, with an average distance of 6.4 basis points. The average width of the 95 percent (99 percent) bands is 22 (26) basis points. Over the sample, the standard deviation of the width of the 95 percent bands is three basis points—suggesting that the precision of estimates changes significantly over time. Finally, it is clear from both figures that the model errors are serially dependent, which violates an assumption of the model—that all time series dependencies are due to the factors. However, the efficacy of the model depends on the user's needs. For the yields on the 25-year PO strip, the 99 percent predictive density bands average 26 basis points wide, contain 62 percent of the data, and the data that does not lie within the predictive density bands averages 6.96 basis points away from the band.

Table II reports the predictive densities of the parameters and steady-state variance of the factors from the two-factor model. Note that the standard deviation of the predictive density of the sum of θ_1 and θ_2 is smaller than that of θ_2. This reflects the negative correlation between these two parameters (of -74 percent). We also see that the second factor is much less mean-reverting than the first ($\kappa_2 < \kappa_1$). The risk premia of both factors is of the standard sign-higher risk premia associated with longer horizons ($\lambda_1, \lambda_2 < 0$). The risk premium on the first factor is much larger than that on the second factor.

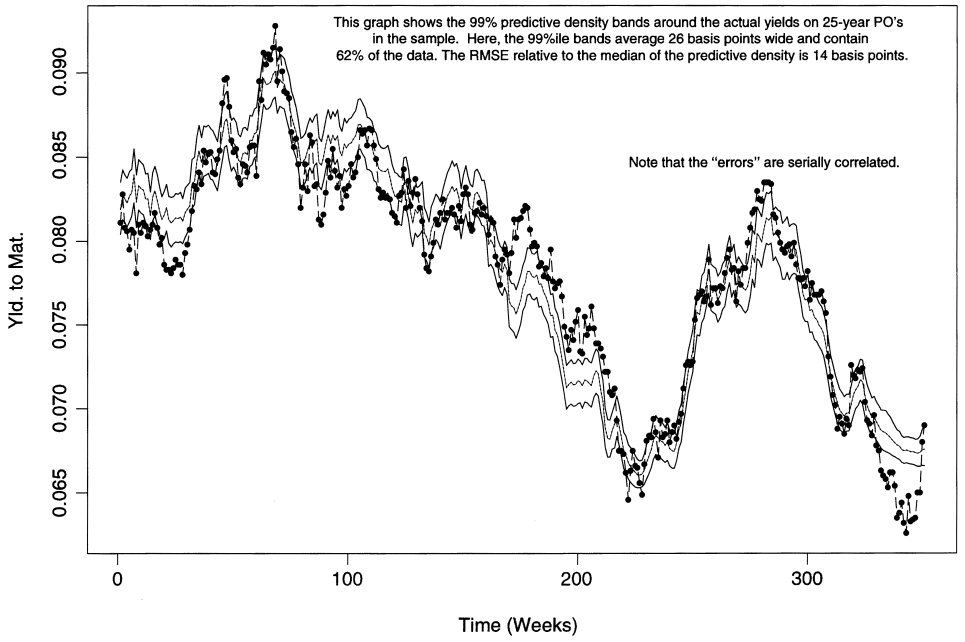

Figure 4. Time series of 25-year bond yield. Ninety-ninth percentile predictive density bands from the two-factor CIR model, overlaid the data, which are weekly observations of yields on 25-year principal only strips of U.S. Treasury bonds.

C. Three-Factor Model

The two-factor model produced a much better fit of the data than the one-factor model. It is therefore natural to evaluate the three-factor model. Figure 5 shows the fit of the three-factor model to the 25-year PO yield. Table I also contains the same information for the three-factor model as the other models. A standard way to evaluate the improved fit of the three-factor model is a portmanteau test based on the likelihood or its GMM counterpart. Consider the conditional likelihood (equation (6)). By averaging this conditional likelihood over the Gibbs draws, we construct the predictive density of the marginalized conditional likelihood. For the three-factor model, the median value of this log-likelihood is 10,303.0. For the two-factor model, the median is 9,809.7. The interquartile range on this function for the three-factor (two-factor) model is [10,259.9,10.351.9] ([9,770.9,9,850.7]). From these predictive densities, we can construct a standard portmanteau test statistic. A likelihood ratio test would reject the two-factor model in favor of the three-factor model at infinitesimal p-values.

However, if we are interested in a root-mean-square-error criterion, relative to the median of the predictive density, and in weighting all five yields equally, then the two-factor model actually is preferable to the three-factor

Table II

Predictive Density of Hyperparameters and Functions of Interest: Two-Factor Model

This table highlights features of the predictive densities of the eight hyperparameters and other functions of interest from the two-factor CIR model, that is, the shape of the likelihood conditioned on the 351 weeks of yields on the five bonds in the sample. The parameters are as defined in the text (e.g., equations (1) and (3)). The variable $\sigma^2_{ss,j}$ is the steady-state variance of factor j. The variable $\text{corr}(x_1, x_2)$ is the correlation between a parameter of Factor 1 and the same parameter of Factor 2. The closer this correlation is to one in absolute value, the less well identified are the separate factors.

Parameter	5%ile	25%ile	Median	Mean	75%ile	95%ile	Std. Dev.	$\text{corr}(x_1, x_2)$
θ_1	0.038	0.039	0.039	0.039	0.040	0.041	0.001	−0.740
θ_2	0.017	0.021	0.030	0.029	0.036	0.046	0.009	
$\theta_1 + \theta_2$	0.0578	0.0609	0.0684	0.0687	0.0751	0.0841	0.0083	NA
σ_1	0.038	0.040	0.042	0.042	0.044	0.046	0.003	−0.166
σ_2	0.080	0.082	0.083	0.083	0.085	0.087	0.002	
$\sigma_1 + \sigma_2$	0.1201	0.1231	0.1251	0.1251	0.1271	0.1300	0.0030	NA
κ_1	0.742	0.777	0.799	0.800	0.825	0.854	0.034	−0.775
κ_2	0.015	0.020	0.025	0.028	0.036	0.046	0.010	
$\kappa_1 + \kappa_2$	0.7856	0.8085	0.8249	0.8278	0.8469	0.8746	0.0275	NA
λ_1	−0.294	−0.263	−0.251	−0.249	−0.234	−0.208	0.024	−0.757
λ_2	−0.087	−0.078	−0.069	−0.072	−0.065	−0.062	0.008	
$\lambda_1 + \lambda_2$	−0.3598	−0.3300	−0.3205	−0.3211	−0.3124	−0.2882	0.0183	NA
$\sigma^2_{ss,1}\cdot 1000$	0.034	0.038	0.044	0.043	0.048	0.054	0.006	−0.466
$\sigma^2_{ss_2}\cdot 1000$	1.345	1.933	4.111	4.583	6.333	10.568	2.921	
$(\sigma^2_{ss,1} + \sigma^2_{ss,2})\cdot 1000$	1.394	1.980	4.152	4.626	6.378	10.610	2.918	NA

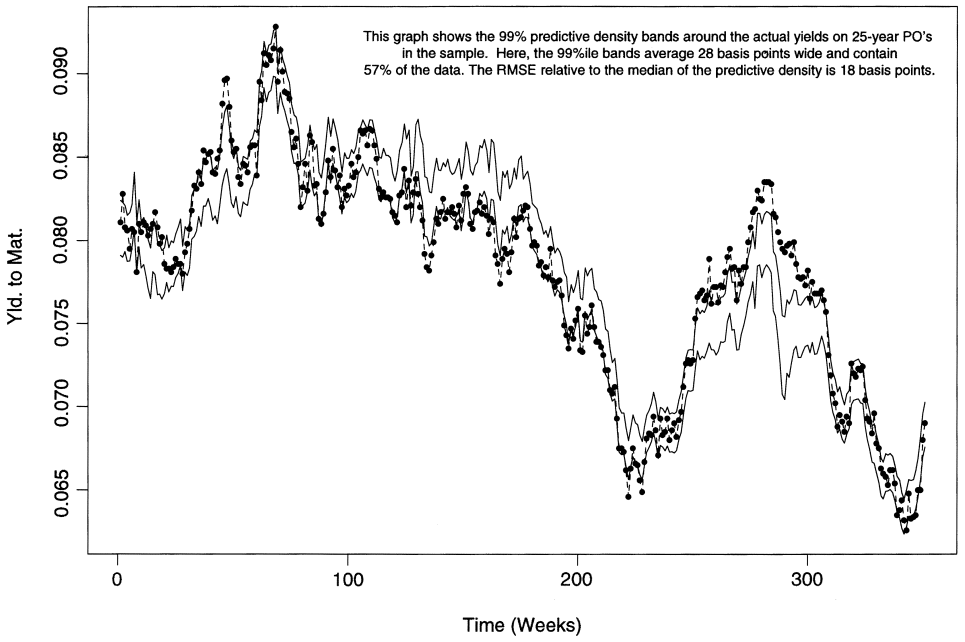

Figure 5. Time series of 25-year bond yield. Ninety-ninth percentile predictive density bands from the three-actor CIR model, overlaid the data, which are weekly observations of yields on 25-year principal only strips of U.S. Treasury bonds.

model, as seen in Table III.[12] Panel C of this table provides the goodness-of-fit analysis for the three-factor model. We see that the three-factor model fits the 90-day bill yield almost perfectly. The 95th percentile band of the predictive density contains 96 percent of the data, with a band width of 27.1 basis points. The RMSE of the three-factor model's predictive density median relative to the data is 7.4 basis points. This compares to 9.1 basis points for the two-factor model. The only other instrument that three-factor model fits better is the five-year PO. In this case, the RMSE of the three-factor model's predictive density median relative to the data is 33.2 basis points, as compared to 34.4 basis points for the two-factor model.

For the other three yields though, the two-factor model is superior to the three-factor model, in terms of goodness of fit: The average improvement for these yields is 5.4 basis points. In both cases, the hardest data to fit are the yields on the five-year PO. The predictive densities are much wider in the case of the three-factor model—the average width is 29.2 basis points compared to 23.8 basis points under the two-factor model. Averaging over all five yields (equally), we see that the two-factor model has a RMSE of 18.5 basis points, whereas the RMSE of the three-factor model is 21.2 basis points.

[12] This table includes results for a proper prior that will be discussed in the Appendix and can be ignored at this point in the text.

Table III

Overall In-Sample Fit of Two-Factor versus Three-Factor CIR Models: Diffuse Priors

Here we compare the actual yields with the predictive densities of those yields from both the two- and three-factor models. A diffuse prior means that the predictive densities are simply the likelihood function viewed as a density for the particular function of the parameters and state variables. The proper prior is used for control purposes, and is discussed in the Appendix. RMSE rel. to med. is the root-mean-square error of the model for the 351 yields in the sample on each instrument. The bands referred to in this table are 95th percentile bands from the posterior distribution of the interest rate (constructed by integrating over the parameter and state variable space). All entries in this table are in basis points.

Instrument (Yield)	RMSE Rel. to Med. (Bas. Pts.)	% Not in 95%ile Bands	Avg Dev from Band	Avg Band Width	Std. Dev. of Band Width
Panel A: Two-Factor—Diffuse Prior					
90-day bill	9.1	23.6	3.3	22.0	5.3
180-day bill	19.2	62.4	10.2	21.9	5.2
5-year bond	34.4	77.8	22.3	26.3	4.8
15-year bond	16.1	43.0	8.2	26.8	4.4
25-year bond	13.8	46.4	7.0	22.1	3.1
Eql wtd avg	18.5	50.6	10.2	23.8	3.14
Panel B: Two-Factor—Proper Prior					
90-day bill	7.9	15.1	3.2	21.6	3.0
180-day bill	12.6	36.2	7.1	21.5	2.9
5-year bond	25.5	64.7	16.4	25.2	4.0
15-year bond	13.9	36.2	7.6	26.0	3.8
25-year bond	19.8	59.5	12.6	21.1	2.4
Eql wtd avg	15.9	42.3	9.4	23.1	3.24
Panel C: Three-Factor					
90-day bill	7.4	4.0	1.4	27.1	5.5
180-day bill	25.7	59.0	14.9	27.6	5.4
5-year bond	33.2	57.5	20.2	36.4	4.9
15-year bond	21.5	44.7	12.0	31.3	4.9
25-year bond	18.1	51.8	9.7	23.3	3.4
Eql WTD AVG	21.2	43.4	14.2	29.2	5.12

The fact that the predictive densities are wider in the three-factor model than in the two-factor model is not surprising. The state space is larger and so the joint likelihood function is flatter. The worse fit relative to the two-factor model is surprising. Most previous analyses have concluded that the three-factor model is an improvement relative to the two-factor model (see, e.g., Dai and Singleton (2000a) and de Jong (2000)). Table IV reports the predictive densities as in Table II, for the three-factor model. Note that the values for κ for the two factors from the two-factor model (shown in Table II) are centered near 0.8 and 0.03. The corresponding values from the three-factor model are centered near 1.1, 0.16, and 0.1. The third factor, which is

Table IV

Predictive Densities of Hyperparameters and Functions of Interest: Three-Factor Model

This table reports posterior distribution characteristics of hyperparameters and functions of interest from the three-factor CIR model. This table highlights features of the predictive densities of the 12 hyperparameters and other functions of interest from the three-factor CIR model (i.e., the likelihood function conditioned on the 351 weeks of yields on the five bonds in the sample). The parameters are as defined in the text (e.g., equations (1) and (3)). The variable $\sigma^2_{ss,j}$ is the steady-state variance of factor j. The variable $\text{corr}(x_j, x_k)$ is the correlation between a parameter of factor j and the same parameter of factor k. The closer these correlations are to one in absolute value, the less well identified are the separate factors.

Parameter	5%ile	25%ile	Median	Mean	75%ile	95%ile	Std. Dev.	$\text{corr}(x_1, x_2)$ $\text{corr}(x_2, x_3)$ $\text{corr}(x_1, x_3)$
θ_1	0.034	0.034	0.034	0.034	0.034	0.034	0.0005	−0.188
θ_2	0.0007	0.0009	0.0010	0.0010	0.0012	0.0014	0.000002	−0.304
θ_3	0.013	0.013	0.014	0.014	0.015	0.016	0.0009	0.350
$\theta_1 + \theta_2 + \theta_3$	0.0480	0.0485	0.0490	0.0493	0.0499	0.0517	0.0011	NA
σ_1	0.036	0.038	0.039	0.039	0.040	0.042	0.0020	0.125
σ_2	0.067	0.068	0.069	0.069	0.070	0.073	0.0016	0.269
σ_3	0.164	0.168	0.173	0.173	0.179	0.185	0.0067	0.223
$\sigma_1 + \sigma_2 + \sigma_3$	0.2704	0.2750	0.2815	0.2820	0.2887	0.2952	0.0080	NA
κ_1	1.035	1.083	1.116	1.125	1.157	1.259	0.063	−0.426
κ_2	0.147	0.153	0.157	0.157	0.160	0.164	0.005	−0.239
κ_3	0.089	0.096	0.106	0.104	0.112	0.119	0.010	0.223
$\kappa_1 + \kappa_2 + \kappa_3$	1.2922	1.3426	1.3783	1.3855	1.4181	1.5182	0.0640	NA
λ_1	−0.293	−0.279	−0.269	−0.270	−0.260	−0.251	0.013	−0.216
λ_2	−0.252	−0.244	−0.240	−0.240	−0.235	−0.231	0.006	0.544
λ_3	−0.244	−0.226	−0.216	−0.219	−0.209	−0.203	0.013	−0.276
$\lambda_1 + \lambda_2 + \lambda_3$	−0.767	−0.739	−0.726	−0.729	−0.715	−0.706	0.0186	NA
$\sigma^2_{ss,1}\cdot 1000$	0.019	0.021	0.023	0.024	0.026	0.029	0.003	0.066
$\sigma^2_{ss,2}\cdot 1000$	0.010	0.014	0.016	0.016	0.019	0.024	0.004	0.044
$\sigma^2_{ss,3}\cdot 1000$	1.517	1.707	1.945	2.043	2.325	2.864	0.413	0.286
$(\sigma^2_{ss,1} + \sigma^2_{ss,2} + \sigma^2_{ss,3})\cdot 1000$	1.5555	1.7439	1.9890	2.1070	2.3670	2.9027	0.4149	NA

the least mean-reverting, also has a relatively high standard deviation. Its σ of 17 percent is nearly twice as large as the more volatile factor in the two-factor model. Further, its steady-state variance is more than twice as large as the more volatile factor in the two-factor model. Note also from Tables II and IV that in the three-factor model, all three factor risk premia (λ) are centered between -0.21 and -0.27. In the two-factor case, the values are more distinct from one another at -0.25 and -0.07. It appears that the likelihood from the three-factor model uses the third factor to bounce around quite randomly, which gives the almost perfect fit to the 90-day bill yield. The relatively large λ in absolute value and the lack of mean reversion presumably allow a better fit at the 25-year yield as well.

This section highlights that the likelihood criterion may not be a useful one. It seems clear, however, that the deterioration of fit of the three-factor model relative to the two-factor model is evidence of model misspecification. The likelihood is dramatically improved by adding a third factor, but even the average in-sample fit, relative to the predictive density median, is worse with the three-factor model. The remainder of the results section focuses on additional functions of these models to isolate where they do well and where they do poorly in fitting properties of the data.

D. More Detail

D.1. Factor Correlations

Using the draws from the factor realizations, we construct the predictive densities of the correlations between various functions of the factors from both the two- and three-factor models. In the two-factor model, there is weak negative unconditional correlation between the factors. The mean of the correlation between the two factors is -10.4 percent, and the interquartile range is [-12.1 percent, -8.9 percent]. The rates of change of the factors are more negatively correlated. The median of the predictive density of this correlation is -20.3 percent, and the interquartile range is [-23.1 percent, -17.5 percent].

There is weak negative correlation between the squared rates of change of the two factors. The correlation between the shock to the factors, after conditioning on their own lags, is negative. The median of the predictive density of this correlation is -37 percent, and the interquartile range is -40 percent to -35 percent. Similarly, in the three-factor model, all three of the pairwise correlations between these factor shocks are negative. However, when we look at the predictive densities of the correlations in the factor levels from the three-factor model, the correlation between factors one and two is negative (mean of -47 percent) and the correlations between factors one and three and two and three are positive.

As noted above, Dai and Singleton (2000a) find that the factors must be negatively correlated to best fit their data, which are yields on LIBOR and 2-year and 10-year fixed-for-variable rate swaps (sampled weekly over roughly

Table V

Correlations between the Interest Rates and Squared Change in Rates: Data and Predictive Densities from the Two-Factor CIR Model

Panel A shows the actual correlation between the yield in week t and the squared change in the rate from week $t - 1$ to week t. Also shown are the predictive densities of this function-of-interest, under the model. For the three bonds in the sample, the duration of the POs varies over the sample. In each case, the model evaluates a zero-coupon bond with the same duration as the PO in the sample. Panel B shows the actual correlation between the yield in week t and the squared change in the rate from week t to week $t + 1$, along with the predictive densities of this correlation.

Instrument	Actual	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std. Dev.
				Panel A				
90-day bill	0.158	0.102	0.145	0.165	0.184	0.221	0.164	0.030
180-Day Bill	0.192	0.098	0.140	0.161	0.180	0.218	0.160	0.030
5-year bond	0.008	0.085	0.128	0.152	0.175	0.220	0.152	0.034
15-year bond	−0.007	0.092	0.136	0.159	0.183	0.226	0.159	0.034
25-year bond	0.093	0.096	0.142	0.165	0.188	0.232	0.165	0.034
				Panel B				
90-day bill	0.170	0.116	0.159	0.179	0.198	0.235	0.178	0.030
180-day bill	0.203	0.111	0.155	0.175	0.195	0.232	0.174	0.030
5-year bond	−0.002	0.092	0.136	0.160	0.183	0.227	0.159	0.034
15-year bond	−0.026	0.098	0.142	0.165	0.187	0.229	0.165	0.034
25-year bond	0.057	0.105	0.149	0.172	0.194	0.236	0.172	0.034

our same sample period). Dai and Singleton also looked at the implied factors in Duffie and Singleton (1997) and found that they are strongly negatively correlated: Extant studies that have looked at the correlation between the two implied factors find it to be around −0.5.

D.2. Time Series Properties

As noted above, there is a large literature that examines the time-series properties of a single interest rate over a long horizon (this is typically a short-term interest rate). A common finding in this context is that the relationship between level and volatility of the rate is stronger than implied by the square-root specification of CIR. The coefficient on rates is often estimated to be between 1.5 and 2.0, instead of 0.5 as in CIR.[13] This relationship is addressed in Tables V and VI. Table V reports the correlations between the level of interest rates in week t with the squared change in the rate from

[13] Although, to be accurate, the implication of the continuous-time specification for discretely sampled data should be recognized, as in this study.

Table VI

Regressions of Log-Squared Price Change on Log-Lagged Rate: Data and Predictive Densities from the Two-Factor CIR Model

This table reports regression estimates from the data and the predictive densities of these estimates from the following regression:

$$\log (r_{t+1} - r_t)^2 = \alpha_0 + \alpha_1 \cdot 2 \log r_t .$$

Panel A shows the intercept from this regression, and Panel B shows the slope coefficient.

Instrument (r)	Actual	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std. Dev.
				Panel A: Intercept				
90-day bill	−9.47	−14.185	−12.993	−12.371	−11.770	−10.615	−12.377	0.917
180-day bill	−10.35	−14.112	−12.839	−12.170	−11.503	−10.257	−12.172	0.993
5-year bond	−11.12	−12.938	−10.145	−8.778	−7.423	−4.676	−8.789	2.081
15-year bond	−6.36	−12.308	−9.217	−7.600	−5.979	−2.718	−7.590	2.434
25-year bond	−14.78	−10.951	−6.631	−4.320	−1.971	2.641	−4.284	3.468
				Panel B: Slope				
90-day bill	0.575	0.150	0.349	0.452	0.555	0.749	0.452	0.154
180-day bill	0.872	0.171	0.387	0.501	0.616	0.828	0.501	0.169
5-year bond	1.36	0.416	0.930	1.185	1.436	1.953	1.184	0.387
15-year bond	0.040	0.546	1.142	1.455	1.764	2.395	1.455	0.469
25-year bond	1.935	0.903	1.745	2.198	2.657	3.563	2.206	0.680

week $t - 1$ to t in Panel A, and with the squared change in the rate from week t to $t + 1$ in Panel B. Even though the model is cast in continuous time, these predictive densities are exact in discrete time. For both correlations, we see that the 90-day bill yields fall within the interquartile range of the predictive density. This suggests that contrary to the studies of short rate dynamics, the two-factor CIR model captures the relationship between the level and the volatility of the short rate well. In the case of the 180-day bill yield, the data falls below the 97.5 percentile of the predictive density for that parameter under both lag structures. For the three longer term bonds in the sample, the value of the correlations estimated from the data lie below the 2.5 percentile of the predictive density for that parameter in both lag treatments. This is not surprising in light of the goodness-of-fit results above. Here we document that the volatility of long yields in the sample is not sensitive enough to the level of the yield to be consistent with the CIR model—but this is not true at the short end of the yield curve.

Table VI shows the results of regressing the log of the squared change in yield on two times the log level of yield at the beginning of the period. This can be viewed as a log-linear, discrete-time approximation to the diffusion, where the slope is an estimate of the interest rate elasticity of volatility. This is motivated by Chan et al. (1992). As always with this analysis, the implications of

the model for this regression are complicated, but the exact densities of the parameters are available via the Gibbs sampler. It is clear from the predictive density on the slope parameters that this parameter is not estimated precisely. The 95th percentile predictive density band for the slope coefficient in the case of the 25-year yield is [0.9, 3.6]. Here, the 90-day, 5-year, and 25-year yields all lie within the 95th percentile bounds of the predictive density on the slope. The slope in the case of the 180-day yield lies above the 97.5 percentile of the predictive density, whereas the slope in the case of the 15-year yield lies below the 2.5 percentile of the predictive density.

In this sample, the two-factor CIR model provides a reasonable characterization of the time-series properties of the short-term interest rate. The overall patterns of the longer rates are also characterized by the model. There is evidence that whether the model can adequately fit the interest elasticity of volatility depends on the framing of the question. While the correlations (in Table V) for the 5-year and 25-year yields lie below those from the model, the slopes (in Table VI) for both of these yields fall within the model's interquartile bands. However, in both cases, the 15-year yield volatility is less sensitive to the level in the data than under the model. We might speculate that the higher interest rate elasticity of volatility found in earlier studies is due to regime shifts over the long horizons commonly used in those studies. We use a short period that is characterized by a single Federal Reserve policy, but nevertheless experiences a variety of yield curve shapes and heterogeneous interest rates. If we had estimated the model on data starting in 1979, we would be asking a single model to fit both the high interest rate, high volatility regime of 1979–1983, as well as the lower rate, lower volatility regime of the 1990s. It is clear that measured interest rate elasticity of volatility would be much higher had we included the two regimes.[14]

Another time-series issue involves pinning down the factors, which may be achieved by evaluating the correlations between the factors and observed variables. There is almost unanimous agreement that the most important factor in the term structure is the level of rates, followed by the slope of the yield curve. There is some question of whether, in fact, the second factor is volatility, but identification is a problem, and as discussed in Dai and Singleton (2000a), CIR involves all factors exerting a separate influence on yield volatility. We examine the predictive density of the correlations between the factors and various observables for both the two-factor and three-factor models. Under both models, the first factor is highly correlated with the short

[14] In fact, we replicate the ad hoc GMM estimation of the discrete-time specification of Chan et al. (1992) on 90-day yields post 1979, obtained from the Federal Reserve Database, and then on the post-1988 period. We find an interest rate elasticity of volatility of 1.5 on the former and 0.46 on the latter. This is not a formal statement that there are regime shifts. But, from the perspective of a simple model that is not informed by macroeconomic factors, it seems likely that estimation over long intervals will be confounded by changing *policy* regimes. Since it is possible that the model that is not informed by the macroeconomy might appear to experience regime shifts, it is incumbent on term structure empirics to identify how much information can be obtained from cross-sectional data, from a relatively short time period.

rate: The mean correlation is 88 percent in both models, and the predictive densities are very tight. This factor is also very highly correlated with the negative of the difference between the 25-year PO yield and the 90-day bill yield in the two-factor model: The mean correlation is 98 percent, and this predictive density is also very tight. The second factor in the three-factor model is most highly correlated with the long rate (although the maximum correlation between this factor and all of our observables is under 75 percent).

Under the two-factor model, the correlations between both factors and various measures of interest rate variance are small in absolute value. There is weak negative correlation between the variance of the first factor and the variance of the 90-day bill rate and the variance of the five-year PO rate. There is weak positive correlation between this factor and the variance of the 25-year PO rate and the variance of the slope of the yield curve (as measured by the difference between the five-year PO rate and the 90-day bill rate). Thus, by looking at levels, we would infer that the first factor is proxying for (the negative of) the slope of the yield curve, and the second factor is proxying for the long rate. The third factor from the three-factor model is highly correlated with both the five-year PO yield and the variance of the yield curve slope.

Since the yields and the factors are highly persistent, it is possible that some of the correlation in levels is spurious. For this reason, we also examine the correlations between the rates of change of the factors and various observables for the two-factor model and the three-factor model. This analysis does indeed suggest that some of the correlations in levels may be spurious. For example, in the two-factor model, the correlation between the rate of change in the first factor and the rate of change in the negative of the slope is 48 percent (compared to 99 percent in levels). This is still the highest absolute value of the correlations, so the "identification" of the first factor is the same in rates of change as in levels. The effect of using rates instead of levels is similar for the second factor. Here the largest absolute correlation is on the rate of change in the yield on the 15-year PO, but the correlation is 38 percent. The precision of the correlation is much less when we look at rates of change than at levels. For example, in the two-factor model, the 95 percent range on the correlation between Factor 2 and the yield on the 15-year PO is [0.92, 0.96], whereas the 95 percent range on the correlation between the rates of change of Factor 2 and the yield on the 15-year PO is [0.19, 0.52]. These results are qualitatively the same in the three-factor model, where none of the correlations between the rates of change of the factors and observables is greater than 50 percent (the strongest such correlation being that between the second factor and the long rate).

We also examine the correlations between the rates of change in the factors and the rate of change in seasonally adjusted money supply (measured by M2). Since this weekly data is announced on Mondays, we align the weeks with both the lead and lagged weeks in our sample (Tuesday closing data). This analysis suggests that Factor 2 from the two-factor model and Factors 2 and 3 from the three-factor model are uncorrelated with money, whereas

Table VII

Overall In-Sample Fit of Cross-Sectional Properties Two-Factor versus Three-Factor CIR Models

Here we compare the functions of the actual yields related to the cross-section of the yield curve with the predictive densities of those functions from both the two- and three-factor models. The functions are defined as follows: Slope = $r_{5,t} - r_{1,t}$; Curve-1 = $r_{5,t} + r_{1,t} - 2 \cdot r_{3,t}$; Curve-2 = $r_{5,t} + r_{1,t} - 2 \cdot r_{4,t}$; Curve-3 = $r_{5,t} + r_{1,t} - \frac{2}{3} \cdot (r_{2,t} + r_{3,t} + r_{4,t})$; Hump = $\text{Max}_m(r_{m,t}) - r_{5,t}$. The first subscript indexes the term of the yield. 1 is for the 90-day yield; 2 is for the 180-day yield; 3 is for the 5-year yield; 4 is for the 15-year yield; and 5 is for the 25-year yield. RMSE rel. to med. is the root-mean-square error of the model for the 351 realizations of each function in the sample. The bands referred to in this table are 95%ile bands from the posterior distribution of the relevant function (constructed by integrating over the parameter and state variable space). All entries in this table are in basis points.

Function of Interest	RMSE Rel. to Med. (Bas. Pts.)	% Not in 95%ile Bands	Avg Dev from Band	Avg Band Width	Std. Dev. of Band Width
Panel A: Two-Factor Model					
Slope	17.3	61.2	9.5	21.7	3.5
Curve-1	59.1	87.8	43.8	23.0	4.2
Curve-2	22.6	55.0	13.8	25.7	3.6
Curve-3	36.9	86.0	27.4	13.0	0.9
Hump	7.2	96.6	3.2	1.8	1.5
$\int$ (Hump) $dG(\epsilon)$	5.0	85.6	2.3	1.7	1.4
Panel B: Three-Factor Model					
Slope	11.5	32.8	4.8	22.7	1.5
Curve-1	41.4	61.8	27.4	37.9	3.6
Curve-2	18.9	43.9	11.2	27.4	6.7
Curve-3	28.4	65.5	19.4	22.4	2.6
Hump	3.0	71.5	0.3	4.2	3.5
$\int$ (Hump) $dG(\epsilon)$	3.0	65.2	0.4	4.3	2.8

Factor 1 under both models has a small negative correlation with next week's money supply.

D.3. Cross-Sectional Properties

Bond traders may feel that certain bonds are mispriced relative to neighboring instruments and engage in barbell spreads or other positions. A term structure model may be evaluated not just by how it does in characterizing the time series of the individual bonds, but also by how well it characterizes the relationship across the various bonds.

Table VII summarizes the average cross-sectional properties across the sample period. The functions reported are slope ($r_{5,t} - r_{1,t}$), curvature-1 ($r_{1,t} + r_{5,t} - 2 \cdot r_{3,t}$), curvature-2 ($r_{1,t} + r_{5,t} - 2 \cdot r_{4,t}$), curvature-3 ($r_{1,t} + r_{5,t} - 2/3 \cdot (r_{2,t} + r_{3,t} + r_{4,t})$), and hump ($\text{Max}_m(r_{m,t}) - r_{5,t}$). We construct

predictive densities for each observation in the sample. The observation on that date is compared to the model predictive density, then these are evaluated or averaged over the 351 weeks in the sample. The *hump* function is nonlinear in the yields, so we present this function computed each of two ways. In the first method, we simply ignore the errors, while in the second, we integrate over the error space, so that the predictive density is actually of an expectation from the union of the CIR model and the model of data errors. This integral is evaluated numerically at *each* of the iterations within the Gibbs sampler, by taking 10,000 draws from the error space (a (351×5) matrix, generated from the Σ from that Gibbs draw). An antithetic variate technique is used to reduce the numerical variance. Comparing the two-factor and three-factor models in this metric yields the same qualitative ranking as the likelihood-based comparisons. In all five cases, the root-mean-square error is lower for the three-factor model than for the two-factor model. The three-factor model produces wider predictive density bands than the two-factor model for all five functions of interest. These wider bands contain about twice as many observations as the corresponding bands from the two-factor model.

Bond traders and fixed income portfolio managers may be interested in comparing "bullet" and "barbell" positions. A question such analysts frequently confront is whether to concentrate on a single intermediate term instrument versus buying short- and long-term instruments. The difference between such positions over time is closely related to the notion of convexity to bond traders. A butterfly regression may be used to characterize such positions. In particular, we regress log changes in an intermediate yield on log changes in the shorter and longer term yields. Naturally, under almost any context (especially the posited model), the regressors in this regression are endogenous, which means the OLS estimators are asymptotically biased. Nevertheless, these asymptotic properties are of no concern for the purpose at hand. The slope, for example, of this regression is simply another function of interest. As with all other functions of interest, this also gives us another metric within which to evaluate the precision of the parameter estimates. As above, we integrate numerically over the error space to construct the predictive density of the expectation taken over the error space.

The results of these regressions on the data, as well as under the two- and three-factor models are reported in Table VIII. There are three sets of regressions reported in the table for the different combinations of contiguous durations in our sample. Panel A shows the results—integrating over the errors—of regressing the change in the 180-day bill rate on the contemporaneous changes in the 90-day and 5-year rates. Panel B reports the results for the similar regression of the 5-year rates on the 180-day and 15-year rates, and Panels C and D do the same for the 15-year rate regressed on the 5-year and 25-year rates. We replicated these panels using the change in rates (instead of log changes as reported) and find very similar results.

These tables confirm the ability of the three-factor model to better fit cross-sectional properties than the two-factor model. An important feature

Table VIII

Butterfly Regression Summary: Data and Predictive Densities from the Two- and Three-Factor CIR Models

This table shows actual and predictive densities from the following regressions:

$$\ln r_{m,t} - \ln r_{m,t-1} = \beta_0 + \beta_1 \cdot (\ln r_{m-1,t} - \ln r_{m-1,t-1}) + \beta_2 \cdot (\ln r_{m+1,t} - \ln r_{m+1,t-1}) + \epsilon_t$$

where $r_{1,t}$ is the yield on the 90-day T-bill on week t in the sample; $r_{2,t}$ is the yield on the 180-day T-bill; $r_{3,t}$ is the yield on the 5-Year Principal Only Strip.; $r_{4,t}$ is the yield on the 15-Year Principal Only Strip.; $r_{5,t}$ is the yield on the 25-Year Principal Only Strip. The t statistics are computed using GMM (Newey–West procedure, with Andrews' procedure to compute the optimal lag length (M)). The subscripts on these statistics correspond to the regression coefficients. This regression is estimated on the actual data. The predictive density is constructed by estimating the regression for the reconstructed yields from each of the Gibbs draws. These regressions are integrated numerically over the estimated error structure, so that we compare the data to the predictive densities of expectations taken over the error space.

		Two-Factor			Three-Factor		
Parameter	Actual	2.5%ile	Median	97.5%ile	2.5%ile	Median	97.5%ile
	Panel A: $\ln r_{2,t} - \ln r_{2,t-1} = \beta_0 + \beta_1 \cdot (\ln r_{1,t} - \ln r_{1,t-1}) + \beta_2 \cdot (\ln r_{3,t} - \ln r_{3,t-1}) + \epsilon_t$						
β_0	−0.00003	−0.00005	−0.00004	−0.00003	−0.00001	0.000002	0.00002
β_1	0.669	0.003	0.113	0.283	0.344	0.532	0.751
β_2	0.359	0.420	0.528	0.611	0.515	0.595	0.663
$\beta_1 + \beta_2$	1.028	0.519	0.637	0.825	0.973	1.130	1.284
t_0	−0.064	−1.20	−1.01	−0.66	−0.40	0.06	0.59
t_1	15.25	0.071	2.47	5.79	7.72	12.26	17.22
t_2	9.94	23.63	32.89	42.14	23.57	35.72	48.52
r^2	0.822	0.766	0.865	0.917	0.899	0.943	0.969
σ_ϵ	0.0093	0.0009	0.0011	0.0013	0.0008	0.0009	0.0011
M	3	2.1	2.7	3.5	2.76	3.27	3.63
	Panel B: $\ln r_{3,t} - \ln r_{3,t-1} = \beta_0 + \beta_1 \cdot (\ln r_{2,t} - \ln r_{2,t-1}) + \beta_2 \cdot (\ln r_{4,t} - \ln r_{4,t-1}) + \epsilon_t$						
β_0	0.00009	0.00005	0.00007	0.00008	0.00002	0.00003	0.00004
β_1	0.3048	0.628	0.882	1.176	0.339	0.492	0.664
β_2	0.8780	0.576	0.885	1.142	0.712	0.891	1.054
$\beta_1 + \beta_2$	1.1828	1.608	1.760	1.966	1.311	1.382	1.464
t_0	0.202	1.108	1.262	1.396	0.712	0.891	1.054
t_1	9.79	10.08	15.34	22.64	7.773	12.119	17.112
t_2	22.82	8.15	13.94	19.57	13.956	19.203	24.637
r^2	0.817	0.832	0.915	0.955	0.930	0.961	0.979
σ_ϵ	0.0093	0.0012	0.0015	0.0018	0.00085	0.00100	0.00118
M	2	2.6	3.1	3.5	2.8	3.2	3.4
	Panel C: $\ln r_{4,t} - \ln r_{4,t-1} = \beta_0 + \beta_1 \cdot (\ln r_{3,t} - \ln r_{3,t-1}) + \beta_2 \cdot (\ln r_{5,t} - \ln r_{5,t-1}) + \epsilon_t$						
β_0	−0.00015	−0.00002	−0.00002	−0.00001	−0.00002	−0.00002	−0.00001
β_1	0.3617	0.351	0.402	0.451	0.326	0.439	0.540
β_2	0.6438	0.342	0.415	0.484	0.271	0.416	0.601
$\beta_1 + \beta_2$	1.0055	0.762	0.816	0.868	0.788	0.857	0.946
t_0	−0.499	−0.91	−0.78	−0.66	−0.84	−0.68	−0.50
t_1	11.15	37.74	49.03	61.19	11.284	15.865	21.037
t_2	15.23	15.12	22.63	31.94	5.376	8.577	12.081
r^2	0.894	0.902	0.944	0.969	0.915	0.954	0.976
σ_ϵ	0.0056	0.0006	0.0006	0.0007	0.00065	0.00075	0.00086
M	0	2.75	3.19	3.52	3.07	3.47	3.73

of these regressions is the sum of the two slope coefficients. In none of the three regressions does this sum estimated from the data fall within the 95th percentile bounds of the predictive density of the sum, from the two-factor model. At the shortest end (Panel A), the value of this sum estimated from the data is 1.028, and the 97.5 percentile in the two-factor model predictive density is 0.825—the estimate is 5.1 standard deviations higher than the mean of the predictive density. This value from the data falls within the 95th percentile likelihood density from the three-factor model, however. Similarly, at the longest end of the yield curve, this sum is 1.006 from the data, and the predictive density's 97.5 percentile from the two-factor model is 0.868 (0.946 from the three-factor model). However, in the middle of the yield curve (Panel B), this sum is well below the 2.5 percentile of the predictive density. The estimated sum is 1.1828, while the 2.5 percentile from the predictive density from the two- (three-) factor model is 1.608 (1.311)—the estimate is 6.5 (5.1) standard deviations lower than the mean of the predictive density.

Another important cross-sectional property that has been examined in several earlier empirical analyses of the term structure is the relative importance of each of the factors in explaining the variance-covariance matrix; Dybvig (1997) looks at this property to isolate the relative importance of possible factors, for example. Traders may want to know how much risk can be eliminated by hedging just the first and second factors. The relative importance of the factors is also important in model building. In particular, we have shown that the statistically motivated likelihood comparisons favor the three-factor model, although the two-factor model has a smaller equally weighted pricing error. By looking at the principal components analysis, we can identify the maximum improvement that a three-factor model could afford.

Again, we numerically integrate over the residuals to compare the data with the predictive density of the model's expectation (averaging over the error space). Table IX reports the results from a principal components analysis on the variance–covariance matrix of the five yields—both from the data and the predictive densities. We see that in the data, 71.3 percent of the total variation in the five bonds is explained by the first eigenvalue. In the two-factor model, if errors are ignored, then the predictive density on this percentage is centered around 88 percent, with the 95th percentile range of the predictive density [85.4 percent, 89.8 percent]. When we integrate over the errors, this predictive density is centered at 84.4 percent, and the 95th percentile range of the predictive density is [73.5 percent, 90.4 percent]. The precision of this statistic is low and the predictive density is highly skewed. Nevertheless, the 71.3 percent estimated from the data falls outside of this range—suggesting that along this metric, the data and the model are at odds.

The percentages explained by the first two and three eigenvalues in the data line up well with the predictive density from the two-factor model, however: Both of these statistics fall within the interquartile ranges of the predictive densities. The percentage explained by the first four eigenvalues also falls within the 95th percentile band of the predictive density.

Table IX

Principal Component Analysis: Data and Predictive Densities from the Two- and Three-Factor CIR Models

This table shows actual and predictive densities of the percentage of the total variation in the variance–covariance matrix of the log changes in weekly yields for the five instruments in our sample. In the base case, the predictive densities are constructed by integrating over the error space (numerically). In the No error case, the errors are ignored when constructing the predictive densities.

% Explained by Eigen Value	Actual	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std. Dev.
2-factor model								
1	71.3	73.5	81.7	84.4	86.9	90.4	83.9	4.26
1 No errors		85.4	87.0	87.8	88.5	89.8	87.7	1.14
2	92.4	88.4	91.2	92.5	93.7	95.2	92.4	1.78
2 No errors		99.5	99.5	99.6	99.6	99.6	99.6	0.032
3	96.4	94.6	95.9	96.6	97.1	97.8	96.5	0.83
3 No errors		99.90	99.92	99.93	99.93	99.94	99.93	0.0086
4	99.0	98.9	99.3	99.4	99.5	99.7	99.4	0.18
4 No errors		99.98	99.98	99.98	99.98	99.99	99.98	0.0015
3-factor model								
1	71.3	86.0	89.5	91.2	92.6	94.1	90.9	2.18
1 No errors		84.9	86.6	87.5	88.3	89.8	87.5	1.26
2	92.4	93.6	94.9	95.6	96.3	97.2	95.6	0.96
2 No errors		97.5	97.8	98.0	98.1	98.4	98.0	0.23
3	96.4	97.4	97.9	98.2	98.5	98.9	98.1	0.38
3 No errors		99.8	99.8	99.8	99.9	99.9	99.84	0.019
4	99.0	99.3	99.5	99.6	99.6	99.8	99.6	0.12
4 No errors		99.98	99.98	99.98	99.99	99.99	99.98	0.0020

Comparing the predictive densities from the two-factor and three-factor models in this metric produces surprising results. First, when errors are ignored, the percentage explained by the first factor is just slightly lower for the three-factor (87.5 percent, on average) than the two-factor model (87.8 percent). Without errors, the three-factor model only has 1.8 percent of the total variation (i.e., 99.8 percent–98.0 percent) explained by the third factor. For many purposes, this feature is probably more telling than a portmanteau statistic. Integrating over the errors generates predictive densities that are more striking for the three-factor model. Specifically, the explanatory effect of each of the first four eigenvalues is less in the two-factor model than in the three-factor model. Along this metric, the two factor model looks more like the data than the three-factor model.

D.4. Expectations Hypothesis

There have been several recent attempts to examine whether dynamic term structure models, such as CIR, are consistent with some of the empirical regularities of interest rate dynamics observed in tests of the pure expectations hypothesis (PEH). For example, Campbell and Shiller (1991)

document that unlike the predictions of the PEH, yield changes are often inversely related to lagged yield spreads. Dai and Singleton (2000b) show that these findings are generally compatible with a large class of dynamic term structure models. With two factors, the CIR model may produce negative coefficients in this type of regression, as shown by Roberds and Whiteman (1999). The risk premia are proportional to the factors; thus, relative movements in the two factors produce time-varying risk premia.

We may view the slope coefficient—for example, from a regression like Campbell and Shiller's (1991)—as another function of interest that combines time series and cross-sectional properties of the data, and, as such, we can construct the exact predictive density of the function in the manner described above. The ability to predict the future of interest rates as a function of the current term structure is obviously an important question that involves time-varying risk premia, which may or may not be compatible with a particular specification. Table X looks at the ability to predict future 90-day bill yields given the 90-day forward rate in the spirit of the PEH. In particular, it summarizes results from the following regression:

$$r_{t+13,t+26} - r_{t,t+26} = \beta_0 + \beta_1(r_{t,t+26} - r_{t,t+13}) + \phi_t. \tag{9}$$

Given the construction of our sample, this is the only set of contiguous forwards and yields available. In this regression, there is no overlapping, so there are 26 observations in the sample. Unlike many earlier studies, the point estimate of the slope coefficient is positive (0.12), but estimated very imprecisely using OLS. In fact, the GMM standard error is 0.56, which means that standard inference would be unable to reject the null hypotheses that the true β_1 is 1, 0, or -1.

The means of the predictive density of this coefficient from the two- (three-) factor models when the errors are ignored are -0.50 (-0.66). When we integrate over the errors as above, the mean of the predictive density of this coefficient is -0.84 (-1.48). This is very much in keeping with earlier studies. These predictive densities are also fairly tight, relative to the asymptotic standard errors of the regressions. The interquartile ranges on this slope coefficient are $[-0.93, -0.76]$ ($[-1.62, -1.33]$). The r^2 from these regressions is very small—both in the data and the likelihood. The model estimated on the data has an r^2 of 0.23 percent, and the mean of the predictive density integrating over errors is 12.57 percent (22.20 percent).

Thus, in this panel, we obtain the standard result that the regression slope coefficient is estimated imprecisely using OLS. But the point estimate in the data is positive, and the 95th percentile predictive density band is strictly negative. In this metric—which combines time-series and cross-sectional properties—there is significant inability of the model to accommodate the data, and the two-factor model more closely resembles the data than the three-factor model.

Table X

"Expectations Hypothesis" Projection: Data and Predictive Densities from the Two- and Three-Factor CIR Models

This table shows actual and predictive densities from the following regressions:

$$r_{t+13,t+26} - r_{t,t+26} = \beta_0 + \beta_1(r_{t,t+26} - r_{t,t+13}) + \phi_t.$$

This is a standard regression used to examine the "expectations hypothesis." The sample of 351 weeks has 26 observations for this regression. The t statistics are computed using GMM (Andrews' augmented Newey–West procedure to compute the optimal lag length (M)). This regression is estimated on the actual data. The predictive densities are constructed differently depending on the treatment of the errors in the initial specification. In Panels B and D, the errors are ignored, so the posteriors are constructed from the regression statistics for the reconstructed yields from each of the Gibbs draws. In Panels A and C, the posterior is constructed estimating the mean of the regression statistics for the reconstructed yields from each of the Gibbs draws averaged over 12,500 draws from the error matrix. (So these functions are *expectations* under the error-augmented CIR model.)

Parameter	Actual	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std. Dev.
Panel A: Two-Factor—Integrating over Errors in Posterior								
β_0	−0.00285	−0.0017	−0.0014	−0.0013	−0.0011	−0.0008	−0.0013	0.0002
β_1	0.124	−1.1105	−0.9330	−0.8415	−0.7563	−0.5534	−0.8435	0.1401
t_0	−2.26	−1.4434	−1.2027	−1.0877	−0.9554	−0.7103	−1.0795	0.1876
t_1	0.218	−2.8179	−2.3552	−2.1142	−1.8767	−1.3452	−2.1117	0.3693
r^2	0.0023	0.0693	0.1053	0.1238	0.1446	0.1886	0.1257	0.0303
σ_ϕ	0.0046	0.0048	0.0049	0.0049	0.0050	0.0052	0.0049	0.0001
M	0	0.7131	0.8953	0.9840	1.0732	1.2537	0.9847	0.1358
Panel B: Two-Factor—Ignoring Errors in Posterior								
β_0	−0.00285	−0.0020	−0.0019	−0.0018	−0.0017	−0.0016	−0.0018	0.0001
β_1	0.124	−0.6290	−0.5457	−0.5017	−0.4558	−0.3676	−0.5006	0.0668
t_0	−2.26	−1.58	−1.44	−1.37	−1.31	−1.19	−1.38	0.10
t_1	0.218	−0.91	−0.76	−0.68	−0.61	−0.48	−0.69	0.11
r^2	0.0023	0.0065	0.0105	0.0129	0.0156	0.0220	0.0132	0.0039
σ_ϕ	0.0046	0.0045	0.0047	0.0047	0.0048	0.0050	0.0048	0.0001
M	0	2	3	3	4	4	3	1
Panel C: Three-Factor—Integrating over Errors in Posterior								
β_0	−0.00285	−0.0017	−0.0015	−0.0014	−0.0013	−0.0012	−0.0014	0.0001
β_1	0.124	−1.8617	−1.6225	−1.4709	−1.3303	−1.1461	−1.4811	0.1954
t_0	−2.26	−1.4480	−1.3232	−1.2394	−1.1634	−1.0510	−1.2434	0.1063
t_1	0.218	−3.7984	−3.3253	−3.0187	−2.7374	−2.2691	−3.0263	0.4085
r^2	0.0023	0.1495	0.1927	0.2204	0.2519	0.2989	0.2220	0.0404
σ_ϕ	0.0046	0.0052	0.0053	0.0054	0.0055	0.0057	0.0054	0.0001
M	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Panel D: Three-Factor—Ignoring Errors in Posterior								
β_0	−0.00285	−0.0019	−0.0018	−0.0017	−0.0016	−0.0014	−0.0017	0.0001
β_1	0.124	−0.7818	−0.7012	−0.6580	−0.6143	−0.5251	−0.6571	0.0651
t_0	−2.26	−1.4921	−1.3421	−1.2724	−1.2066	−1.0755	−1.2754	0.1042
t_1	0.218	−1.3973	−1.1818	−1.0793	−0.9778	−0.7939	−1.0841	0.1544
r^2	0.0023	0.0190	0.0277	0.0330	0.0389	0.0528	0.0338	0.0086
σ_ϕ	0.0046	0.0047	0.0049	0.0050	0.0051	0.0053	0.0050	0.0002
M	0	0.0000	1.0000	1.0000	2.0000	3.0000	1.3386	0.6772

D.5. Information about Derivatives and Hedging in Yields

Another function of the model parameters and state variables is a derivative price. In this subsection, we look at pricing an interest rate derivative relative to the estimated term structure. This provides another metric for evaluating the precision of the parameter and state variable estimates. One of the nice features of CIR is that prices of any interest rate claim can be derived. A trader might consider using interest rate data to estimate a model's parameters and then use the estimated model to obtain derivative prices and hedge ratios. The approach in this paper is uniquely suited to evaluate the precision of such an exercise.

Given the state variables and parameters, evaluation of any interest rate derivative's price involves taking an expectation over the *conditional* density of the state variables at the expiration of the derivative in the appropriate measure.[15] Of course, we must first identify the domain in state variable space where the futures price exceeds the option's strike price. In the two-factor model, this involves a two-dimensional integral (see Chen and Scott (1992) for details). For each iteration from the Gibbs sampler, we have to evaluate this integral to construct a single option price. The predictive density on this option price is constructed by evaluating this integral and solving for the option price at each iteration from the Gibbs sampler. We can construct a predictive density on the option price to examine how much information there is in bond yields for the pricing of derivatives. As an example, we consider an option which expires in 6 weeks, on an underlying futures contract which expires in 10 weeks on an eight percent coupon bond that matures 15 years after the expiration of the futures contract. We construct the predictive density on the price of this futures option *relative to the estimated term structure*.[16]

For illustrative purposes, we construct the predictive density on May 24, 1990, when interest rates are high (and this option is deep out of the money). This is a hypothetical option, and we abstract from delivery options, and so on, to focus on the *information* in yields about derivative prices. In fact, the predictive density of the option price is fairly diffuse. We are 90 percent sure (assuming that the two-factor CIR is true) that the option price lies between \$0.36 and \$1.28. If this were a bid-ask spread, it would amount to 112 percent of the midpoint. The median of the predictive density is \$0.69; thus the

[15] The fact that we are evaluating such a distribution highlights the importance of using the appropriate densities in the empirical analysis. Even if approximating the transition density of the state variables as normal does not result in gross pricing errors, it would not provide the appropriate conditional density—once again a noncentral χ^2—to use in option pricing.

[16] While this function of interest is unlike the previous functions of interest, where we compare the predictive density to the data, the precision of the state variable estimates, per se, is still relevant in this case. This is because we construct the density of the state variable at the time the option expires, *conditional on* its current value. Unlike the predictive density–data comparisons above, since the purpose here concerns the amount of information about the state variables and parameters, we ignore the errors in constructing these predictive densities.

predictive density is asymmetric—it exhibits a right tail and looks like a lognormal density function.

Consider the same futures option on April 8, 1993, when interest rates are relatively low (and this option is in the money). Here the 90th percentile band (from \$125.92 through \$146.80) is roughly 17 percent of the median (\$136.55). Here, since the option is so deep in-the-money, the predictive density of the option price is much more symmetric than in the previous case (where the option was deeply out-of-the-money). It is not surprising, given the above discussion of the lack of identification of the individual parameters, that the yield data do not have enough information to pin down derivatives on the long end of the yield curve. While it is conceptually straightforward to include option prices in the information set of the likelihood function, this would be intractable numerically since just evaluating a single pdf (of which thousands must be done for a single Gibbs iteration) involves computing a two-dimensional integral over the state variable space.

The relatively diffuse predictive density on these option prices is consistent with the notion that, whereas the parameters that determine the shape of the yield curve are well-identified, the parameters that determine the dynamics are less well identified. For example, in Table II we see that the sum of the two factors' σs, κs, and λs are known precisely relative to the steady-state parameters θ and the steady-state variance. The ratio of the predictive density interquartile range to the mean for the sums of σ is 3.2 percent, κ 4.6 percent, and λ 3.2 percent. This ratio for the sum of θ is 21 percent, and, for the steady-state variance, it is 95 percent. In general, we would expect that the predictive density would be more diffuse the less-in-the-money the option, since, for a deep in-the-money option, the model influence on the price is less.

Hedging interest rate exposure is at least as important a problem as pricing interest rate derivatives. Just as we construct a predictive density on an option price, we can construct such a density on an option's deltas or elasticities. For a two-factor model, there are two deltas—one for each factor. We construct the predictive densities for the factor elasticities for the option above (the out-of-the money option of May 24, 1990). The mean (standard deviation) of the elasticities with respect to the two factors are -5.89 (0.41) and -12.50 (0.99). The predictive densities are both symmetric around their means and look like Gaussian densities. The elasticity with respect to the first factor is smaller and less diffuse than the option price's elasticity with respect to the second factor. (No doubt because the second factor has both a higher and more diffuse steady-state variance and a more diffuse predictive density on θ (Table II).)

For the deep in-the-money option of April 8, 1993, the factor elasticities are much lower here than they are for the out-of-the-money option. It is still the case that the predictive density of the elasticity of the first factor is much tighter and smaller than that of the second factor. Clearly, these factor elasticities are important for fixed income management and trading–not only the levels, but the entire distributions.

Thus, there is not enough information in this panel of 351 weeks and five bond yields to pin down option prices on long-term bonds well. Further, one of the factor's elasticity is fairly tight, whereas the elasticity on Factor 2 is diffuse. This suggests that to use the CIR model to price and hedge derivatives requires either a longer time series of yields or derivative prices themselves in the sample. For the two-factor model, and the numerical procedures used in this paper, these possibilities would be virtually infeasible on even the fastest computers.

VI. Conclusions

This paper introduces a framework to estimate the model of Cox, Ingersoll, and Ross wherein all interest rates are used to learn about all factors, and the exact discrete-time transition densities from the continuous-time model are used. Because we know the transition densities of the state variables, we use the full conditional densities for these factors. Although they are nonstandard, we demonstrate how to use numerical techniques to sample from them. Thus, we learn about the state variables from all of the data, and the transition from continuous-time theory to discretely sampled data is exact.

We evaluated one-, two-, and three- (orthogonal) factor CIR models. The one-factor model is incapable of explaining the observed volatility of long bond yields. The two-factor model can generate patterns that look like the data, but the goodness of fit is in doubt. The three-factor model uses its higher dimensionality to give a marginally better fit to the shortest- and longest- term bonds, as well as cross-sectional properties of the yield curve, but at the same time the fit in the intermediate range of the yield curve deteriorates. Standard likelihood-based inference agrees with previous findings that the three-factor model is a dramatic improvement on the two-factor model. In spite of this, the two-factor model has lower overall errors when we compare the sample data to the predictive densities and weight the five instruments equally. This insight could not be gained by empirically identifying a yield with a factor (since adding an additional factor a priori means a better in-sample fit). This insight also suggests that in this high-dimensional problem, statistical model comparison may not correspond to comparisons based on what traders find important. A mixture of the two- and three-factor models may be the best specification, and it is possible that the correlated factor structure in Dai and Singleton (2000a) generates such a mixture, although this is not obvious.

We looked closely at the properties of the two-factor model. Unlike other studies, we find that the two factors are uncorrelated. However, the shocks to the factors are negatively correlated. The serial dependence of the model errors is inconsistent with the theoretical properties of the model. Nevertheless, the average standard deviations of the data errors do not seem outrageous: 10 basis points for the 90-day bill; 13 basis points for the 180-day bill; 28 basis points for the 5-year bond; 20 basis points for the 15-year bond;

and 24 basis points for the 25-year bond. The amount of information in the yields for pricing derivatives was also demonstrated. It appears that yields do not adequately identify factor volatilities to allow precise derivative pricing. We also demonstrate construction of predictive densities for option elasticities. Recent research is bridging early tests of pure expectations models of the term structure and dynamic term structure models. We show that the model implies a negative coefficient on the forward rate in predicting future spot rates, but that in the data this coefficient is positive. Thus, the model can certainly accommodate the types of departures from PEH seen in earlier studies, but along this dimension it is different from the data in our sample.

Finally, in the introduction, we noted that future research may explore the possibility of extending the CIR model to allow for parameter and state variable uncertainty on the part of the representative agent in the economy. The noted shortcomings of the model suggest that this might indeed be a fruitful tack. The serially dependent errors in long bonds seen in the benchmark two-factor case, for example, could occur because the agent is unable to isolate where the individual factors are. Short rates are pinned down because the representative agent knows what the sum of the two factors is, but with finite information, this agent is unable to separately locate each factor. It is easy to imagine how such learning errors could be serially dependent. Pursuit of such an extension to CIR would seem well motivated by the results in this paper. Of course, in such a setting, the model restrictions on factor correlations would hold conditional on agents' information sets and perhaps not unconditionally.

Appendix

A. Incorporation of Prior Beliefs

In the text, all results are presented using only the joint likelihood function. Of course, the tools used and the likelihood approach are commonly associated with Bayesian, or subjectivist, analysis. Mapping out the surface of the likelihood is tantamount to specifying a prior that is proportional to a constant—that is, a diffuse, improper prior. While diffuse priors may serve to facilitate scientific communication, they may entail problems because the flat prior on the parameter space has peculiar implications for functions of the parameters that are of fundamentally more importance than the parameters themselves. Lamoureux and Zhou (1996) provide an example of this problem and show that it may be "solved" by using proper priors. The specification in Lamoureux and Zhou involved highly persistent time series, and, as such, raises concerns for interest rate data.

Such concerns turn out to be unfounded for the analysis in this paper, as the qualitative analysis in the text is robust to use of a proper prior. In this appendix, we show the proper priors used to validate the results in the text for the two-factor model. Note that there is no need to specify a prior on the z process. In fact, Geweke (1995b) notes that latent time-series models have

the form of a hierarchical prior, where the latent time-series process corresponds to an intermediate level of the hierarchy.

B. Residual Variance–Covariance Matrix

Since the conditional likelihood is inverse Wishart from the perspective of Σ, a natural conjugate prior for Σ is therefore inverse Wishart, or Σ^{-1} is Wishart, (i.e., $\Sigma^{-1} \sim W(v, \mathbf{V})$). The predictive density is then simply

$$\Sigma^{-1} \sim W\left(v + T, \mathbf{V} + \sum_{t=1}^{T} \epsilon_t \epsilon_t'\right).$$

Here the parameters of this density are $v = 20$, and

$$\mathbf{V} = \begin{bmatrix} 0.00000025 & 0 & 0 & 0 & 0 \\ 0 & 0.00000045 & 0 & 0 & 0 \\ 0 & 0 & 0.00000065 & 0 & 0 \\ 0 & 0 & 0 & 0.00000085 & 0 \\ 0 & 0 & 0 & 0 & 0.000001 \end{bmatrix}.$$

C. Model Parameters

With a proper prior,

$$f(\kappa | \text{data, all other parameters}) \propto p(\kappa) \mathcal{L}(r | \kappa, z, \cdot) f(z | \kappa, \cdot),$$

where $p(\kappa)$ is the prior density on κ. Since κ and σ must be nonnegative, a natural prior density is the inverted gamma with parameters ν and s.[17]

In this case, the full conditional density has the form

$$\begin{aligned} &f(\kappa | \text{data, all other parameters}) \\ &\quad \propto \frac{1}{\kappa^{v+1}} e^{-vs^2/2\kappa^2 + \Sigma_{t=1}^{T} (2 r_t' \Sigma^{-1} A_t + A_t' \Sigma^{-1} A_t - 2 A_t' \Sigma^{-1} B_t \cdot z_t)} \\ &\qquad \times \prod_{t=1}^{T} c\, e^{-u_t - \nu_t} \left(\frac{\nu_t}{u_t}\right)^{q/2} I_q(2(u_t \nu_t)^{1/2}), \end{aligned}$$

where the constants in the last kernel are defined following (2) (with time subscripts added here since the parameter depends on the entire history of z). The proper prior on λ is Gaussian.

[17] The inverted gamma density is obtained by viewing a standard normal density from the perspective of the variance of the normal. It has two parameters and looks a lot like a χ^2 density in shape, that is, it originates at 0 and is skewed to the right.

Table AI

Priors and Posteriors for Hyperparameters and Functions of Interest: 2-Factor Model (Proper Prior)

This table highlights features of the posterior distribution of the eight hyperparameters and other functions of interest from the two-factor CIR model, (updating the proper prior using the 351 weeks of yields on the five bonds in the sample).

Prior **Posterior**	5%ile	25%ile	Median	Mean	75%ile	95%ile	Std. Dev.	corr(x_1,x_2)
θ_1	0.0247	0.0341	0.0450	0.0548	0.0628	0.1139	0.0379	0.0000
	0.0415	**0.0427**	**0.0463**	**0.0461**	**0.0482**	**0.0523**	**0.0035**	**–0.3225**
θ_2	0.0123	0.0170	0.0222	0.0273	0.0311	0.0584	0.0192	
	0.0061	**0.0068**	**0.0074**	**0.0073**	**0.0078**	**0.0083**	**0.0007**	
$\theta_1 + \theta_2$	0.0441	0.0576	0.0720	0.0829	0.0941	0.1547	0.0455	ND
	0.0485	**0.0504**	**0.0537**	**0.0534**	**0.0560**	**0.0592**	**0.0034**	
σ_1	0.0163	0.0215	0.0274	0.0312	0.0361	0.0593	0.0158	0.0
	0.0368	**0.0384**	**0.0400**	**0.0405**	**0.0427**	**0.0449**	**0.0026**	**0.0312**
σ_2	0.0163	0.0215	0.0274	0.0312	0.0361	0.0593	0.0158	
	0.0878	**0.0894**	**0.0905**	**0.0906**	**0.0917**	**0.0935**	**0.0017**	
$\sigma_1 + \sigma_2$	0.0381	0.0479	0.0576	0.0627	0.0712	0.1036	0.0231	ND
	0.1261	**0.1287**	**0.1308**	**0.1310**	**0.1332**	**0.1365**	**0.0032**	
κ_1	0.1854	0.2551	0.3389	0.4148	0.4746	0.8906	0.2854	0.0
	0.5894	**0.6378**	**0.6740**	**0.6786**	**0.7289**	**0.7599**	**0.0543**	**–0.3384**
κ_2	0.2149	0.2966	0.3939	0.4842	0.5520	1.0181	0.3625	
	0.1033	**0.1096**	**0.1170**	**0.1186**	**0.1256**	**0.1423**	**0.0115**	
$\kappa_1 + \kappa_2$	0.4850	0.6320	0.7875	0.8999	1.0234	1.6596	0.4798	ND
	0.7193	**0.7572**	**0.7883**	**0.7972**	**0.8447**	**0.8806**	**0.0515**	
λ_1	−0.2319	−0.1547	−0.1000	−0.1000	−0.0457	0.0327	0.0800	0.0
	–0.2135	**–0.1932**	**–0.1347**	**–0.1429**	**–0.1102**	**–0.0571**	**0.0512**	**–0.3365**
λ_2	−0.2035	−0.1246	−0.0700	−0.0700	−0.0168	0.0606	0.0800	
	–0.1851	**–0.1674**	**–0.1591**	**–0.1607**	**–0.1511**	**–0.1451**	**0.0119**	
$\lambda_1 + \lambda_2$	−0.3556	−0.2466	−0.1704	−0.1700	−0.0943	0.0158	0.1131	ND
	–0.3798	**–0.3475**	**–0.2946**	**–0.3037**	**–0.2626**	**–0.2277**	**0.0485**	
$\sigma^2_{ss,1}$·1000	0.1077	0.2671	0.5156	1.0628	1.0488	3.3623	3.6872	0.0000
	0.0397	**0.0461**	**0.0549**	**0.0568**	**0.0655**	**0.0810**	**0.0129**	**–0.2634**
$\sigma^2_{ss,2}$·1000	0.0462	0.1146	0.2208	0.4546	0.4490	1.4401	1.5611	
	0.1779	**0.2209**	**0.2590**	**0.2567**	**0.2900**	**0.3328**	**0.0468**	
$(\sigma^2_{ss,1} + \sigma^2_{ss,2})$·1000	0.2635	0.5299	0.8954	1.5125	1.6015	4.3294	4.1416	ND
	0.2418	**0.2783**	**0.3140**	**0.3135**	**0.3444**	**0.3906**	**0.0452**	

Notes: Prior and posterior (**bold**) distribution characteristics of hyperparameters and functions of interest from the two-factor CIR model. The parameters are as defined in the text (e.g., equations (1) and (3)). The variable $\sigma^2_{ss,j}$ is the steady-state variance of factor j. The variable corr(x_1,x_2) is the correlation between a parameter of Factor 1 and the same parameter of Factor 2. The closer this correlation is to one in absolute value, the less well identified are the separate factors.

The prior on κ is inverted gamma. As noted above, the intractability of the conditional densities requires that we have univariate priors on the κ for each factor. Here the parameters are $v_1 = 3$, and $s_1 = 0.3$; $v_2 = 3$, and $s_2 = 0.35$. Table AI shows percentiles for this prior. For example, we see that the inverted gamma prior on κ_1 is skewed to the right: The 2.5 percentile is 0.19, the median is 0.34, the mean is 0.42, and the 95th percentile is 0.89. The mean and standard deviation of the prior on κ_2 are 0.48 and 0.36.

The univariate priors on the standard deviations of the factors, σ, are inverted gamma. In the highlighted case, the parameters are $v_1 = v_2 = 4$, and $s_1 = s_2 = 0.025$. As shown in Table AI, these priors imply a mean and standard deviation of the inverted gamma distribution of 3.12 percent and 1.58 percent.

The prior for λ is normal. Here the prior on $\lambda_1 \sim N(-0.1, 0.08^2)$, and $\lambda_2 \sim N(-0.07, 0.08^2)$.

The univariate priors on the long-term means of the factors, θ, are inverted gamma. The two parameters in the priors are $v_1 = 3$, and $s_1 = 0.04$; and $v_2 = 3$, and $s_2 = 0.02$. We can see the implication of this density and parameter choice in Table AI: for θ_1 the mean, median, and standard deviation are 5.5 percent, 4.5 percent, and 3.8 percent, and the corresponding statistics for θ_2 are 2.7 percent, 2.3 percent, and 3.8 percent.

Table AI reports properties of the proper prior and the posterior obtained by using the likelihood to update this prior using Bayes Rule. We can compare Tables AI and II to evaluate the *direct* effect of the prior on the estimation of the parameters.

The values of σ_j are robust to the prior. But there are material differences in the other parameters. The biggest difference appears in the λ_j parameters. The two factors' median values on this parameter under the diffuse prior are -25.1 percent and -6.9 percent, whereas, from the highlighted proper prior, the median values are -13.5 percent and -15.9 percent. This is not surprising since λ is the only parameter that can take on positive and negative values. The proper prior is centered below 0 (as seen in Table AI), whereas the diffuse prior naturally is centered at 0. It seems that the proper prior pulls down both factors' λ values, and this affects θ and κ. The lower and more diffuse prior on κ_2 is responsible for the much larger and more diffuse posterior density for the steady-state variance of factor 2 in Table II, relative to its counterpart from the proper prior of Table AI. While the parameters on the individual factors do exhibit some sensitivity to the prior, the sums of the two factors' parameters are more robust. For example, the posterior median values of κ_1 under the proper and diffuse priors are 0.67 and 0.80, respectively. The corresponding values for κ_2 are 0.12 and 0.02. The posterior medians of the sums of the two κs from these two different priors are 0.79 and 0.82.[18]

Despite the differences in the parameters, the effect of the proper prior relative to the diffuse prior is very small for functions that might interest either a bond trader or analyst. For example, Table III shows the predictive densities of the weekly yields over the sample and characterizes the goodness-of-fit of the models. Here we see that the predictive posterior densities on the sample yields are generally the same widths using either prior. The overall fit is actually better using the proper prior. As seen in the text when we compared the two- and three-factor models, the likelihood per se is weighted toward the short rate and cross-sectional fit. Thus, the proper prior provides a better fit to the three intermediate yields while the diffuse prior (i.e., the likelihood) provides a better fit for the 90-day bill yield and the 25-year PO

[18] As in Section D.5, the question of robustness of the posterior to the prior involves the amount of information contained in the data about the particular aspects of the model. This data provide relatively precise information about the total risk premium (the sum of λ_1 and λ_2, for example), but much less information as to how to assign this across the two factors.

yield. The biggest difference is at the middle yield (the 5-year PO), where the root-mean-square error of the median of the predictive posterior is 34.4 basis points, using the diffuse prior, and 25.5 basis points from the proper prior.

REFERENCES

Aït-Sahalia, Yacine, 1996, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.

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

Casella, George, and Edward I. George, 1992, Explaining the Gibbs sampler, *The American Statistician* 46, 167–174.

Chan, K. C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.

Chapman, David A., and Neil D. Pearson, 2000, Is the short rate drift actually nonlinear? *The Journal of Finance* 55, 355–388.

Chen, Ren-Raw, and Louis Scott, 1992, Pricing interest rate options in a two-factor Cox–Ingersoll–Ross model of the term structure, *Review of Financial Studies* 5, 613–636.

Chen, Ren-Raw, and Louis Scott, 1994, Multi-factor Cox-Ingersoll–Ross models of the term structure: Estimates and tests from a Kalman filter, Unpublished working paper, University of Georgia.

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 J. Singleton, 2000a, Specification analysis of affine term structure models, Unpublished working paper, Stanford University.

Dai, Qiang, and Kenneth J. Singleton, 2000b, Expectation puzzles, time-varying risk premia, and dynamic models of the term structure, Unpublished Working paper, NYU and Stanford.

David, Alexander, and Pietro Veronesi, 1999, Option prices with uncertain fundamentals: Theory and evidence on the dynamics of implied volatilities and over-/underreaction in the options market, Working paper, GSB, The University of Chicago.

de Jong, Frank, 2000, Time-series and cross-section information in affine term structure models, *Journal of Business and Economic Statistics* 18, 300–314.

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

Dybvig, Phil, 1997, Bond and bond option pricing based on the current term structure, in Michael Dempster and Stanley Pliska, eds: *Mathematics of Derivative Securities* (Cambridge University Press, Cambridge, U.K.).

Gelman, Andrew, Iven Van Mechelen, Geert Verbeke, Daniel Heitjan, and Michel Mulders, 2000, Graphical model diagnostics in missing and latent data settings, Working paper, Columbia University, Department of Statistics.

Geweke, John, 1994, Bayesian comparison of econometric models, Working Paper 532, Federal Reserve Bank of Minneapolis.

Geweke, John, 1995a, Monte Carlo simulation and numerical integration, Staff Report 192, Federal Reserve Bank of Minneapolis.

Geweke, John, 1995b, Posterior simulators in econometrics, Working Paper 555, Federal Reserve Bank of Minneapolis.

Gibbons, Michael R., and Krishna Ramaswamy, 1993, A test of the Cox, Ingersoll, and Ross model of the term structure, *Review of Financial Studies* 6, 620–658.

Lamoureux, Christopher G., and Guofu Zhou, 1996, Temporary components of stock returns: What do the data tell us? *The Review of Financial Studies* 9, 1033–1059.

Meng, Xiao-Li, 1994, Posterior predictive *p*-Values, *The Annals of Statistics* 22, 1142–1160.

Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory* (John Wiley & Sons, New York).

Pearson, Neal, and T. S. Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279–1303.

Ritter, Christian, and Martin A. Tanner, 1992, Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy–Gibbs sampler, *Journal of the American Statistical Association* 87, 861–868.

Roberds, William, and Charles H. 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.

Santa-Clara, Pedro, and Didier Sornette, 1999, The dynamics of the forward interest rate curve with stochastic string shocks, Working Paper, Anderson Graduate School of Management, University of California, Los Angeles.

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

Stanton, Richard, 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973–2002.

Veronesi, Pietro, 1999, Stock market overreaction to bad news in good times: A rational expectations equilibrium model, *Review of Financial Studies* 12, 975–1007.