Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests From a Kalman Filter Model

By

Ren-Raw Chen* and Louis Scott**

July 1993
Revised July 1995
Revised October 1998
Revised February 2002
Revised September 2002

*   Associate Professor of Finance, Rutgers University.
** Executive Director, Morgan Stanley & Co.

Ren-Raw Chen
Rutgers Business School
94 Rockafeller Rd., Room 121
Piscataway, NJ 08854
U.S.A.
732-445-4236
rchen@rci.rutgers.edu

Louis Scott
Morgan Stanley & Co.
25 Cabot Square, Canary Wharf
London E14 4QA
United Kingdom
44 207 425 6581
Louis.Scott@morganstanley.com

Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests From a Kalman Filter Model

REN-RAW CHEN
School of Business, Rutgers University, Piscataway, NJ 08854

LOUIS SCOTT
Morgan Stanley & Co., London E14 4QA

Abstract

This paper presents a method for estimating multi-factor versions of the Cox, Ingersoll, Ross (1985b) model of the term structure of interest rates. The fixed parameters in one, two, and three factor models are estimated by applying an approximate maximum likelihood estimator in a state-space model using data for the U.S. treasury market. A nonlinear Kalman filter is used to estimate the unobservable factors. Multi-factor models are necessary to characterize the changing shape of the yield curve over time, and the statistical tests support the case for two and three factor models. A three factor model would be able to incorporate random variation in short term interest rates, long term rates, and interest rate volatility.

Key Words: interest rates, term structure, Kalman filter

1. Introduction

The Cox, Ingersoll, Ross (1985b) model is an equilibrium asset pricing model for the term structure of interest rates. The model provides solutions for bond prices and a complete characterization of the term structure which incorporates risk premiums and expectations for future interest rates. The model is frequently presented as a one factor model, but in sections 6 and 7 of their paper, Cox, Ingersoll, and Ross, hereafter CIR, show how to incorporate multiple factors and how to extend the model to value nominal bonds and nominal claims. The model is important for several reasons: it provides a link between intertemporal asset pricing theory and the term structure of interest rates, preserves the requirement that interest rates remain nonnegative, and produces relatively simple closed form solutions for bond prices. The model is also useful as a tool for valuing interest rate derivative securities.[1]

In this paper, we estimate multi-factor versions of the CIR model by using a state space model in which estimates of the unobservable state variables are generated by a Kalman filter. One, two, and three factor models are estimated, and several tests are performed to determine whether these models can accurately capture the variability of the term structure over time. The estimation technique is different from methods previously used to estimate CIR models, and the econometric model is able to capture several important features of the term structure.[2] The Kalman filter model does not require the additional restrictive assumptions associated with previous work based on maximum likelihood estimation.[3] Our results support the case for multi-factor models and the empirical tests identify several advantages associated with a three factor model. The three factor model that we estimate represents an extension of the two factor model analyzed by Longstaff and

Schwartz (1992). In that model, Longstaff and Schwartz interpret the two factors as the short term interest rate and interest rate volatility. An alternative interpretation is one in which the two factors are the short term rate and a long term rate, which is similar in spirit to the work of Brennan and Schwartz (1979). The two factor model that we estimate seems to fit more closely to this latter interpretation. If the short term rate, a long term rate, and interest rate volatility are three distinct, important influences for the term structure, then a three factor model is necessary. Using factor analysis applied to a sample of bond returns, Litterman and Scheinkman (1991) have found that three factors are necessary to characterize empirically the intertemporal variation of the term structure: the general level of interest rates, the slope of the yield curve, and curvature, which is associated with volatility.[4] The three factor model that we estimate captures these three features of the term structure within a model that can be used for asset pricing. We directly estimate the fixed parameters of the CIR model, and we effectively impose all of the restrictions implied by the model. These fixed parameters determine the cross correlations and the dynamic behavior of bond rates. In addition, the econometric model produces estimates of the parameter combinations that are relevant for asset pricing.

Finally, it is necessary to observe that the CIR model does not satisfy all of the normality assumptions required for statistical consistency in the maximum likelihood estimation of a state-space model. As we explain in section 2, the Kalman filter for estimating the unobservable state variables requires modification. As a result, the filter for the CIR model is not linear and may have a bias even though it is a minimum mean squared error estimator. The innovations in the CIR model have noncentral χ^2 distributions, in contrast to the normal distribution that is assumed for maximum likelihood estimation of Kalman filter models. Maximum likelihood estimation under the assumption of normality is often applied in cases where the fundamental innovations are not normally distributed. When these estimators are consistent, they are classified as quasi maximum likelihood estimators. Consistency for these estimators can be verified by setting up the first order conditions for the maximization,

$$\frac{\partial \ln L}{\partial \beta} = 0 ,$$

and checking the large sample properties. In the case of quasi maximum likelihood estimators that impose normality assumptions, statistical consistency depends on correctly modeling the first and second moments. In the application of the CIR model here, the first and second moments are modeled correctly, but statistically consistency cannot be established because the Kalman filter produces estimates of the unobservable state variables that may be conditionally biased. The Kalman filter estimates are, however, minimum mean squared error estimates and are unconditionally unbiased. We examine the seriousness of this potential bias by performing a Monte Carlo analysis of the approximate maximum likelihood estimator. This analysis reveals significant biases for the parameters that determine the time series properties of interest rates, but the biases for the parameter combinations that are relevant for asset pricing are found to be either small or insignificant.

The paper is organized as follows. In section 2 we present a multi-factor model for pricing nominal bonds which follows from sections 6 and 7 of CIR, and we show that the model can be set up in discrete time as a state space model, which is estimated by approximate maximum likelihood. Estimates of the unobservable state variables are computed with a nonlinear Kalman filter. In sections 3 and 4, we present the estimates for one, two, and three factor models, and we perform a variety of tests on the models. In section 5, we present the Monte Carlo analysis of the approximate maximum likelihood estimator.

2. The State Space Model for Parameter Estimation

2.1 The CIR Model of the Term Structure

The model for the analysis is the nominal pricing model in equations (57) - (60) of CIR (1985b). The instantaneous nominal interest rate is assumed to be the sum of *K* state variables,

$$r = \sum_{j=1}^{K} y_j$$

and the state variable are assumed to be independent and generated as square root diffusion processes:

$$dy_j = \kappa_j(\theta_j - y_j)\,dt + \sigma_j \sqrt{y_j}\,dz_j, \text{ for } j = 1,\ldots,K \;.$$

The solution for the nominal price at time *t* of a nominally risk-free bond that pays \$1 at time *s* is determined as follows:

$$N_t(s) = A_1(t,s)\cdots A_K(t,s)\exp\left\{-B_1\, y_{1t} - \cdots - B_K\, y_{Kt}\right\}.$$

where $A_j(t,s)$ and $B_j(t,s)$ have the form that is given in CIR:

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

$$B_j(t,s) = \frac{2(1-e^{-\gamma_j(s-t)})}{2\gamma_j\, e^{-\gamma_j(s-t)} + (\kappa_j+\lambda_j+\gamma_j)(1-e^{-\gamma_j(s-t)})}$$

and $\gamma_j = \sqrt{(\kappa_j+\lambda_j)^2 + 2\sigma_j^2}$. Each state variable has a risk premium, $\lambda_j\, y_j$, and each λ_j is treated as a fixed parameter.[5] The continuously compounded yield for a discount bond is defined as follows:

$$R_t(s) = \frac{-\ln N_t(s)}{s-t}$$

which is a linear function of the unobservable state variables. Given a set of yields on K discount bonds, one can conceptually invert to infer values for the state variables.[6]

2.2 The State Space Model for Estimation

To estimate the CIR model, we use the state space model, which is described in Engle and Watson (1981) and Watson and Engle (1983).[7] Because the unobservable state variables are distributed conditionally as noncentral χ^2 variates, adjustments must be made to the Kalman filter. If we consider observations for the state variables and bond rates sampled at discrete time intervals, the continuous time model can be expressed as follows:

$$y_t = a + \Phi\, y_{t-1} + v_t$$

$$R_t = A + B\, y_t \qquad ,$$

where y_t, v_t, and a are $K \times 1$ vectors, R_t and A are $M \times 1$ vectors, Φ is a $K \times K$ diagonal matrix, and B is an $M \times K$ matrix. y_t contains the unobservable state variables and R_t contains the continuously compounded yields for various discount bonds, $R_t(s_i)$, i = 1, ..., M. A, B, a, and Φ are functions of the fixed parameters in the stochastic processes for the state variables. The individual elements of y_t and R_t are as follows:

$$y_{jt} = \theta_j\, (1 - e^{-\kappa_j \Delta t}) + e^{-\kappa_j \Delta t}\, y_{j,t-1} + v_{jt}\,, \quad j = 1, \ldots, K$$

$$R_t(s_i) = -\sum_{k=1}^{K} \frac{\ln A_k(t, s_i)}{s_i - t} + \sum_{k=1}^{K} \frac{B_k(t, s_i)}{s_i - t}\, y_{kt}\,, \quad i = 1, \ldots, M$$

where Δt is the size of the time interval in the discrete sample. The equations for the state variables follow directly from the noncentral χ^2 distribution. The expectation for y_{jt} conditional on information at t - 1 is

$$\theta_j\, (1 - e^{-\kappa_j \Delta t}) + e^{-\kappa_j \Delta t} y_{j,t-1} \quad .$$

The error term v_{jt} represents the unanticipated change in y_{jt} and it has a conditional expected value of zero and a conditional variance equal to

$$\sigma_j^2 \left(\frac{1 - e^{-\kappa_j \Delta t}}{\kappa_j} \right) \left(\tfrac{1}{2}\, \theta_j\, (1 - e^{-\kappa_j \Delta t}) + e^{-\kappa_j \Delta t}\, y_{j,t-1} \right) \; .$$

There is no serial correlation in v_{jt}, but there is serial dependence in the variance. The model described up to this point is an exact discrete time representation of the CIR model, without any of the typical approximations that are applied in deriving discrete time representations of continuous time models.

This model can be expressed in the state space form by adding error terms to the equations for the observable bond rates:

$$\begin{aligned} y_t &= a + \Phi\, y_{t-1} + v_t \\ R_t &= A + B\, y_t + \varepsilon_t \ , \end{aligned} \tag{1}$$

where $\varepsilon_t' = (\varepsilon_{1t}, \ldots, \varepsilon_{Mt})$. Each error term is a measurement error, or noise term, that is introduced to allow for small errors and imperfections in the observed bond rates. Bond rates are typically computed from averages of bid and ask prices, and in many samples, the rates for long term discount bonds must be computed from various coupon bond issues.[8] We assume that there is no serial correlation and no cross correlation in these measurement errors for the bond rates. This simple structure for the measurement errors is imposed so that the serial correlation and the cross correlation in bond rates is attributed to the variation of the unobservable state variables. With these assumptions, the covariance matrix for the error terms in (1) can be written as follows:

$$E_{t-1} \begin{pmatrix} v_t \\ \varepsilon_t \end{pmatrix} \begin{pmatrix} v_t \\ \varepsilon_t \end{pmatrix}' = \begin{pmatrix} Q_t & 0 \\ 0 & U \end{pmatrix} \ ,$$

where Q_t is a diagonal matrix with the conditional variances of the state variables on the diagonal, and U is a diagonal matrix with the variances of the measurement errors on the diagonal.

2.3 The Kalman Filter

The fixed parameters of a state space model are typically estimated by the method of maximum likelihood using the Kalman filter to compute estimates of the unobservable state variables. The model in (1) fits into the framework of the state space model described in equations (1) - (3) in Watson and Engle (1983), but the innovations for the state variables in the CIR model are not normally distributed. If the innovations in a state space model are not normally distributed, the standard linear Kalman filter is no longer conditionally unbiased as an estimator of the unobservable state variables.[9] The fixed parameters in a state space model are typically estimated by using the Kalman filter to compute innovations in the unobservable state variables and maximizing a likelihood function that imposes normal distributions for all of the innovations. There are models in which the normality assumptions can be relaxed and the maximum likelihood estimator is still consistent. These estimators are known as quasi maximum likelihood estimators. In this application for the CIR model, the quasi maximum likelihood estimation is not consistent because there is a bias in the Kalman filter. To develop a consistent quasi maximum likelihood estimator, one must develop an unbiased estimator for the unobservable state variables.

To clarify some of these issues, we begin with a review of the linear model. The Kalman filter is an algorithm for computing estimates of the state variables at each time period during the sample.[10] The innovations for the observed bond rates are defined as:

$$u_t \ = \ R_t \ - \ \left(A \ + \ B\,(\,a + \Phi\,\hat{y}_{t-1})\right) \ ,$$

where $\hat{y}_{t-1}$ is an estimate of y_{t-1} based on u_{t-1} and $\hat{y}_{t-2}$. For the initial estimate $\hat{y}_0$, one can use the unconditional means for the state variables. The innovations for the state variables, given the previous estimates, are defined as:

$$\eta_t \ = \ y_t - a - \Phi\,\hat{y}_{t-1} \ .$$

The Kalman filter is a linear model for computing estimates of the state variables:

$$\hat{y}_t \ = \ a \ + \ \Phi\,\hat{y}_{t-1} \ + \ D_t\,u_t \ , \tag{2}$$

where D_t is an $M \times K$ matrix of coefficients which are set to minimize the mean squared error between y_t and $\hat{y}_t$. If the innovations are normally distributed, this estimator is also the expectation conditional on the current and past values of the observed variables. In the estimation of the unobservable state variables, the fixed parameters of the model are presumed to be known. The estimator is formed by solving the following minimization:

$$\min_{\hat{y}_t} \ E\,(y_t - \hat{y}_t)'\,(y_t - \hat{y}_t) \ = \ \sum_{j=1}^{K} E\,(y_{jt} - \hat{y}_{jt})^2$$

Because $y_t - \hat{y}_t \ = \ \eta_t - D_t u_t$, the minimization can be restated as

$$\min_{D_t} E\,(\eta_t - D_t\,u_t)'\,(\eta_t - D_t\,u_t) \ .$$

Here, the expectation is conditional on the observations available at time t (R_t, $\hat{y}_{t-1}$, R_{t-1}, ...). The Kalman filter uses a least squares projection of η_t on u_t to estimate the coefficients in D_t, which determine the current innovations for the state variables. The first order conditions for this minimization problem are

$$E\left[(\eta_t - D_t\,u_t)\,u_t'\right] \ = \ E\,(\eta_t\,u_t') - D_t\ E\,(u_t\,u_t') \ = \ 0 \ \ .$$

The covariance matrices are defined as follows:

$$E\,(\eta_t\,u_t') \ = \ \hat{\Sigma}_t\ B' \quad \text{and} \quad E\,(u_t\,u_t') \ \equiv \ H_t \ = \ B\,\hat{\Sigma}_t\ B' \ + \ U \ ,$$

where $E\,(\eta_t\,\eta_t') \ = \ \hat{\Sigma}_t$ and $\hat{\Sigma}_t$ is determined recursively

$$\Sigma_{t-1} \;=\; \hat{\Sigma}_{t-1} \;-\; \hat{\Sigma}_{t-1}\, B'\, H_{t-1}^{-1}\, B\, \hat{\Sigma}_{t-1}$$

$$\hat{\Sigma}_t \;=\; \Phi\, \Sigma_{t-1}\, \Phi' \;+\; \hat{Q}_t \qquad .$$

In the CIR model, the diagonal elements of Q_t are linear functions of y_{t-1}, and $\hat{Q}_t$ is formed by replacing y_{t-1} in Q_t with $\hat{y}_{t-1}$. The solution that minimizes the mean squared errors is $D_t = \hat{\Sigma}_t\, B'\, H_t^{-1}$ and the estimates are computed as follows:

$$\hat{y}_t \;=\; a \;+\; \Phi\, \hat{y}_{t-1} \;+\; \hat{\Sigma}_t\, B'\, H_t^{-1}\, u_t \quad .$$

To start this algorithm, one sets $\hat{y}_0$ equal to the unconditional mean for y and Σ_0 equal to the unconditional variance, and the calculations are done recursively. This filter is the standard Kalman filter with one important difference: H_t depends on $\hat{y}_{t-1}$, which depends on observations through time t - 1 (R_{t-1}, R_{t-2}, ...).

The model also has one more important difference because there is an extra restriction on the state variables, $y_t \geq 0$. If the Kalman filter produces a negative estimate for y_{jt}, one can generate a better estimate, in the sense of minimizing the mean squared error, by setting $\hat{y}_{jt}$ equal to zero. One can add this nonnegativity constraint to the minimization of the mean squared errors and use the Kuhn-Tucker conditions.

$$\min_{D_t} E\,(\eta_t - D_t\, u_t)'\,(\eta_t - D_t\, u_t)$$

$$\text{s.t. } \hat{y}_t \;=\; a \;+\; \Phi\, \hat{y}_{t-1} \;+\; D_t\, u_t \;\geq\; 0 \quad .$$

The first order conditions are now modified as follows:

$$\hat{\Sigma}_t\, B' \;-\; D_t\, H_t \;\geq\; 0 \text{ as } \hat{y}_t \geq 0 \ .$$

If $\hat{y}_t > 0$, the corresponding equation holds as an equality. The solution to these first order conditions can be found by first computing the linear solution $D_t = \hat{\Sigma}_t\, B'\, H_t^{-1}$. If an element of $\hat{y}_t$ is negative, set that estimate equal to zero and drop the corresponding row from the system of equations. The result is

$$\hat{\Sigma}_t^*\, B' \;-\; D_t^*\, H_t \;=\; 0 \ ,$$

where $\hat{\Sigma}_t^*$ is $(K-1)\times K$ and the row dimension of D_t^* is K - 1. The resulting solution, D_t^* = $\hat{\Sigma}_t^*\, B'\, H_t^{-1}$, produces the same estimates for the nonnegative elements of $\hat{y}_t$ found in the original solution. The net result is the linear estimator from the standard Kalman filter, with any negative estimates replaced with zeros. We refer to this Kalman filter as a quasi linear Kalman filter, but it is nonlinear.

The quasi linear Kalman filter minimizes the mean squared errors subject to the restriction that the estimates must be nonnegative. Even though we have retained the linear structure of the standard Kalman filter, the resulting filter is nonlinear in two respects: the nonnegativity restriction and the dependence of the coefficients in D_t on $\hat{y}_{t-1}$. Because the estimator is not strictly linear, it is not a best linear estimator that minimizes mean squared error, and it is possible that there are other nonlinear estimators that produce smaller mean squared errors. This estimator, like the standard Kalman filter, is computed by inverting a matrix and performing several matrix multiplications, and it does not require an iterative solution to a set of nonlinear equations.

2.4 The Approximate Maximum Likelihood Estimator

The maximum likelihood estimator is obtained by maximizing the following log-likelihood function:

$$\max_{\beta} \ln L = \sum_{t=1}^{T} L_t = \sum_{t=1}^{T} -\tfrac{1}{2}\left(\ln\left|H_t\right| + u_t' \, H_t^{-1} \, u_t \right), \tag{3}$$

where β is a vector containing all of the fixed parameters to be estimated. In our application of the state space model, u_t is not normally distributed, but this approximate maximum likelihood estimator, based on the normality assumption, is a method of moments estimator. The estimates are found by solving the likelihood equations, the first order conditions for the maximization problem in (3):

$$-\frac{\partial \ln L}{\partial \beta_j} = \sum_{t=1}^{T} \left\{ \left(\frac{\partial u_t}{\partial \beta_j} \right)' H_t^{-1} u_t + \tfrac{1}{2} \operatorname{tr}\left(H_t^{-1} \frac{\partial H_t}{\partial \beta_j} \right) - \tfrac{1}{2} u_t' H_t^{-1} \frac{\partial H_t}{\partial \beta_j} H_t^{-1} u_t \right\} = 0 \ ,$$

for $j = 1, \ldots, N$. Setting these derivatives equal to zero is equivalent to setting

$$\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\partial u_t}{\partial \beta_j} \right)' H_t^{-1} u_t + \frac{1}{2\,T} \sum_{t=1}^{T} \left\{ \operatorname{tr}\left(H_t^{-1} \frac{\partial H_t}{\partial \beta_j} \right) - \operatorname{tr}\left(H_t^{-1} \frac{\partial H_t}{\partial \beta_j} H_t^{-1} u_t u_t' \right) \right\} = 0 \ ,$$

for $j = 1, \ldots, N$. Here we have used the result that tr (AB) = tr (BA). In each one of the first order conditions, the approximate maximum likelihood estimator effectively sets the sum of two sample moments equal to zero. The partial derivatives in the equations are functions of the fixed parameters and past values of the random variables, R_{t-1}, R_{t-2}, …, etc. Consistency could be established if the following results were to hold:

$$E\left(u_t \,\middle|\, \hat{y}_{t-1}\right) = E\left(u_t \,\middle|\, R_{t-1}, R_{t-2}, \ldots\right) = 0 \text{ and}$$

$$E\left(u_t u_t' \,\middle|\, \hat{y}_{t-1}\right) = E\left(u_t u_t' \,\middle|\, R_{t-1}, R_{t-2}, \ldots\right) = H_t \ .$$

The first condition is not satisfied because the quasi linear Kalman filter for the CIR model does not necessarily equal the conditional expectation for the state variables.[11] If the Kalman filter is conditionally unbiased, then both conditions hold and the estimator is statistically consistent.

In quasi maximum likelihood estimation, the covariance matrix for the parameter estimates must be adjusted and the likelihood ratio statistics for model restrictions do not have asymptotic χ^2 distributions.[12] As shown in White (1982), the covariance matrix for $\sqrt{T}\,(\hat{\beta} - \beta)$ converges to

$$\left[E\left(\frac{\partial^2 \ln L_t(\beta)}{\partial\beta\,\partial\beta'} \right) \right]^{-1} E\left(\frac{\partial \ln L_t(\beta)}{\partial\beta} \right)\left(\frac{\partial \ln L_t(\beta)}{\partial\beta} \right)' \left[E\left(\frac{\partial^2 \ln L_t(\beta)}{\partial\beta\,\partial\beta'} \right) \right]^{-1} . \qquad (4)$$

If the innovations are normally distributed as in the standard model, then

$$E\left(-\frac{\partial^2 \ln L_t(\beta)}{\partial\beta\,\partial\beta'} \right) = E\left(\frac{\partial \ln L_t(\beta)}{\partial\beta} \right)\left(\frac{\partial \ln L_t(\beta)}{\partial\beta} \right)' ,$$

and the covariance matrix becomes the familiar inverse of the information matrix. These statistical properties do not necessarily carry over for the approximate maximum likelihood estimator, but we use equation (4) to compute standard errors for the parameters estimates. In section 3, we proceed as if the biases are not serious. In section 5, we examine the magnitudes of the biases. To find the approximate maximum likelihood estimator, we use a modified method of scoring which is described in Berndt, Hall, Hall, and Hausman (1974) and Engle and Watson (1981). The covariance matrix for β is computed by replacing the expectations in (4) with sample moments. The first derivatives and the expectation of the second derivative matrix are given in Engle and Watson.

2.5 Comparison with Other Estimation Techniques in the Literature

Several approaches have been used in the previous research on the empirical estimation of CIR models. Gibbons and Ramaswamy (1993) and Heston (1989) have used unconditional sample moments in a generalized method of moments (GMM) framework to estimate and test different versions of the model. GMM estimators of this form are typically less efficient than alternative estimators and the resulting estimates for the model parameters have relatively large standard errors. This approach does have some advantages: specific distributional assumptions are not required, and as Gibbons and Ramaswamy have noted, the GMM estimators allow for measurement errors in the bond rates. Longstaff and Schwartz (1992), in their estimation of a two factor model, use one month T-Bill rates as a proxy for the instantaneous interest rate and estimates of interest rate volatility generated from a GARCH model.[13] Their parameters are estimated by regressing bond yield changes on the changes in these two estimated factors. This approach depends on the assumption that the estimates for the two factors do not contain measurement error.

A third approach is maximum likelihood estimation, which has been used in Chen and Scott (1993) and Pearson and Sun (1994). The likelihood function for the observed bond rates is developed from the conditional density functions for the state variables, which have noncentral χ^2 distributions. Because there are typically more bond rates than unobservable factors or state variables, Chen and Scott introduce measurement errors for selected bond rates. To obtain tractable likelihood functions, they assume that some of the bond rates are measured without error. This approach to maximum likelihood is not tractable if all of the bond rates are measured with some error, and it requires a specific distribution for the measurement errors. Duffie and Singleton (1997) have applied this maximum likelihood estimator to rates in the interest rate swap market. Pearson and Sun have circumvented this problem by assuming no measurement errors and by restricting the number of cross sections for the bond rates to be equal to the number of factors. In the estimator developed in this section, we allow for measurement errors on all of the bond rates and the likelihood function remains relatively simple and tractable. This estimator differs from the previous maximum likelihood estimators because we do not use the non-central χ^2 density function and we are not required to impose a specific assumption for the distribution of the measurement errors. The estimator uses the structure imposed by the CIR model on the first and second moments of the conditional distributions, and it imposes more of the model structure than the previously implemented GMM estimators. The negative feature of the estimator developed in this section is the potential large sample bias. The magnitudes of the biases are examined in section 5.

More recently, a simulated method of moments (SMM) estimator has been developed by Dai and Singleton (2000) for exponential affine term structure models that include the CIR model as a special case. Their SMM estimator is statistically consistent, but the estimator is computationally slow as it requires iterative solutions coupled with Monte Carlo simulations. Two recent papers have used Kalman filters to estimate multi-factor term structure models. Babbs and Nowman (1999) use the standard Kalman filter, state-space model to estimate a generalized Vasicek model, in which all of the innovations are normally distributed. They use weekly data on 8 maturities over the period 1987 to 1996 to estimate 1, 2, and 3 factor models. Their zero coupon rates are extracted from quotes for U.S. LIBOR and U.S. swap rates, and the longest maturity is 10 years. Geyer and Pichler (1999) have estimated multi-factor CIR models using the estimation technique developed in this section. They use monthly data for 16 maturities in the U.S. Treasury market over the period 1964 to 1993, and they estimate models with up to 5 factors. The longest maturity in their study is 5 years. In the next section, we use 4 maturities from the U.S. Treasury market to estimate 1, 2, and 3 factor models, and the longest maturity varies from 15 to 30 years. We emphasize the use of long maturities because we feel that the length of the maturities used in the analysis will have a significant impact on the estimation of the mean reversion parameters.

3. Estimation of One, Two, and Three Factor Models

3.1 Data

We use two data sets for the estimation of the CIR model. The first data set includes yields for discount bonds calculated by McCulloch, and presented in Shiller and McCulloch (1990). We use the rates from the table for the zero coupon yield curves for 3 months, 6 months, 5 years, and the longest maturity available (10-25 years). The rates are annualized and stated on a continuously compounded basis and represent rates for discount bonds. These rates have been computed from month end prices in the Treasury bond market, and we use the data for the period 1960 to 1987. This monthly data set serves as a longer sample for the estimation of the parameters in the interest rates processes.

Interest rate volatility appears to have changed over the last 30 years, and Treasury bond prices are also available on a daily or weekly basis. The second data set consists of bond prices on Thursdays from January 1980 to December 1988. Prices for 13 week and 26 week T-bills, 5 year Treasury notes, and the longest maturity noncallable bonds available have been collected from the *Wall Street Journal*. For this period, there are 470 weekly observations. This particular set of observations was selected for several reasons. T-bills mature on Thursdays and there are only a few holidays that fall on Thursday during the sample period.[14] By using weekly data instead of monthly data, we have a much larger sample size. The four different maturities were chosen so that different points along the yield curve could be used to estimate the factors and the parameters in the processes that determine the factors. The two Treasury bills are discount bonds, but the two longer term bonds are coupon bonds. The yields for 5 year discount bonds and long term discount bonds have been approximated from the coupon bonds by assuming two forward rates: one to apply from 6 months to 5 years and one from 5 years to 30 years. The longest maturities on noncallable bonds were 15 years at the beginning of the 1980's and 30 years at the end of the sample period.

3.2 Empirical Results

The results of the estimation are presented in Tables I and II. Table I contains the estimates for the monthly data set, 1960-87, and Table II contains the estimates for the weekly data set, 1980-88. Time is measured in years so that all of the parameter values are expressed on an annual basis. All of the variance parameters are statistically significant, but the results are mixed for the individual estimates of the κ, θ, and λ parameters. Most of the estimates for the risk premiums are negative and approximately half are statistically significant. Only one of the risk premium estimates is positive and it is not significant. The κ and θ estimates are statistically significant for the first factor, but they are generally insignificant for the second and third factors in the multi-factor models. The log-likelihood values increase dramatically as the number of factors is increased. The standard deviations for the measurement errors naturally decrease as the number of factors is increased.

In most of the cases, the standard deviations for the measurement errors on the 6 month T-Bill rate go to zero. This is a common phenomenon in factor analysis, and the estimates for the variances, and the standard deviations, are constrained to be nonnegative. In the weekly data set, the variance for the measurement error on the long term bond rate also goes to zero in the three factor model. The estimated standard deviations for the measurement errors are much larger for the one factor model. In the monthly data set, these standard deviations are 33 basis points for the 3 month T-Bill rate, 102 basis points for the 5 year bond rate, and 132 basis points for the long term bond rate. In the weekly data set, the standard deviations are 40 basis points for the 3 month T-Bill rate, 104 basis points for the 5 year bond rate, and 122 basis points for the long term bond rate. The measurement errors for the 5 year bond rate and the long term bond rate are quite large in the one factor model. The largest standard deviations for measurement errors in the multi-factor models are 30 to 37 basis points, and most of the estimates are much smaller. For example, in the three factor model estimated from the weekly data set, these standard deviations are 32 basis points for the 3 month T-Bill rate, 0 for the 6 month T-Bill rate, 7 basis points for the 5 year bond rate, and 0 for the long term bond rate. These results indicate that most of the variation in the observed bond rates is explained by the common factors in the multi-factor models.

The estimates for the κ parameters are close to zero for the extra factors in the multi-factor models, and almost all of the estimates are smaller than their standard errors. The only exception is κ_2 in the two factor model estimated from the weekly data, but the estimate, .021185, is relatively small. The stochastic processes for the state variables resemble first order autoregressions if they are sampled at discrete time intervals, and the κ parameters determine the rate of mean reversion. The coefficient on y_{t-1} in the autoregression is equal to $\exp(-\kappa\Delta t)$, where Δt is the size of the time interval over which the data are sampled. If the κ parameter is close to zero, then the autoregression coefficient is close to one, which is equivalent to having a root close to the unit circle in the time series representation. If κ is zero, the state variable is a random walk and the bond rates are not stationary time series. Cooley, LeRoy, and Parke (1992) have argued from a theoretical perspective that interest rates should be stationary time series.[15] If $\kappa = 0$ for a square root process in the CIR model, the factor behaves like a pure random walk but zero becomes an absorbing barrier for the process; if the process hits zero, it disappears. If the κ parameter is small so that $2\kappa\theta < \sigma^2$, the process can hit zero, but zero serves as a reflecting barrier so that the process continues. The small κ estimates indicate that some of the factors in the multi-factor models do exhibit characteristics similar to random walks, and these factors are the ones that explain the variation of the long term bond rates. The rate of mean reversion for each factor can be measured by computing mean half lives.[16] In the two factor model estimated from the weekly data set, the mean half lives are .95 years for the first factor and 32.7 years for the second factor. In the three factor model estimated from the weekly data set, the mean half lives for the three factors are .48 years, 40.9 years, and 19.7 years. The factors with long mean half lives are the ones that determine the variation of the longer term bond rates. The estimates for the mean half lives are much shorter in the studies by Babbs and Nowman(1999) and Geyer and Pichler (1999). We attribute the difference to our inclusion of 30 year maturities.

The estimator developed in section 2 is based on the assumption that the bond rates and the state variables are stationary time series, but the Kalman filter can be applied to nonstationary time series.[17] If the series are not stationary, one uses a starting value for y_0 which is treated as a fixed parameter and Σ_0 is set equal to zero. We have treated the bond rates as stationary time series and the unconditional means, θ, are used for the initial estimate $\hat{y}_0$, and the unconditional variances and covariances are used for Σ_0. If κ_j is close to zero, the corresponding variance in Σ_0 is large so that the estimate, $\hat{y}_{jt}$, for the first period is allowed to have a large deviation from the unconditional mean. The Kalman filter uses conditional variances for the subsequent observations in the sample. The large variance associated with a small κ parameter affects the first observation only. We have run the models in Tables I and II with the nonstationary Kalman filter and the changes in the parameter estimates are very small.

An analysis of the factor loadings can be used to determine the nature of the factors calculated by the Kalman filter. In this model, the factor loadings are the coefficients in the matrix *B* defined in equation (1). In Figures 1-4, we present graphs of these coefficients across different maturities in the two and three factor models. The coefficients for the two and three factor models computed from the estimates in the 1960-87 sample are presented in Figures 1 and 2. The coefficients computed from the estimates for the weekly data set, 1980-88, are in Figures 3 and 4. The patterns are similar for both sets of graphs. In the two factor model, the coefficients on the first factor decrease quickly as time to maturity increases. The coefficients for the second factor are approximately one for all maturities. The first factor has a strong influence on short term rates, but a diminished effect on long term rates, and this factor determines the slope of the term structure. The second factor affects all rates and determines the general level of interest rates. We find that estimates of the first factor are highly correlated with the slope of the term structure, specifically the short term rate minus the long term rate, and estimates of the second factor are highly correlated with the long term interest rate. In the three factor model, the coefficients for the first factor are similar to those for the first factor in the two factor model. The coefficients decrease sharply as time to maturity increases, and this factor determines the slope of the term structure. The coefficients for the second factor decrease slowly as time to maturity increases, and the coefficients for the third factor increase with time to maturity and then level off between 20 and 30 years. The second and third factors determine the general level of interest rates and the relationship between medium and long term rates. We find that the sum of the second and third factors is highly correlated with medium and long term rates. In these model estimates, the second factor has a higher volatility and the third factor has a lower volatility. The interaction of these two factors determines the curvature of the term structure and the volatility of bond rates.

In all of these models, the relevant combinations of parameters for valuing bonds and interest rate derivative assets are $(\kappa_j + \lambda_j)$, $\kappa_j \theta_j$, and σ_j. The estimates for these parameter combinations in the multi-factor models are presented with their asymptotic standard errors in Table III. Most of these parameter combinations are statistically significant in the sense that the estimates are large relative to their standard errors, but several of the estimates for $\kappa_j \theta_j$ are close to zero and are smaller than their standard errors. All of the estimates for $(\kappa_j + \lambda_j)$ and σ_j are several times greater than their standard errors. Most of the parameter combinations that are relevant for asset pricing are estimated with a high degree of precision.

We turn now to statistical tests of the different models. Because the three models are nested, one could use the likelihood ratio statistic for hypothesis testing, but this statistic does not have the standard χ^2 distribution when the innovations are not normally distributed. The asymptotic distribution for the likelihood ratio statistic would be a weighted sum of χ^2 distributions, as described in Vuong (1989). We have already noted that the approximate maximum likelihood estimator is potentially biased and the statistical results for quasi maximum likelihood estimators may not apply. Comparisons of log likelihood functions across the models do serve as indicators of the model performance in fitting the data. The values for the log likelihood function in the monthly data set are 5,828.77 for the one factor model, 6,730.11 for the two factor model, and 6,954.48 for the three factor model. In the weekly data set, these values are 8505.09 for the one factor model, 10008.65 for the two factor model, and 10,424.24 for the three factor model. The likelihood ratio statistics for tests of the one factor model versus the two factor model are 1,803 in the monthly data and 3,007 in the weekly data. The likelihood ratio statistics for tests of the two factor model versus a three factor model are 449 in the monthly data and 831 in the weekly data. The values for these likelihood ratio statistics are extremely large and would indicate rejection of the null hypotheses at low significance levels if one could apply either the standard χ^2 distribution or the results for weighted sums of χ^2 distributions in Vuong. There is also a substantial reduction in the standard deviation of the measurement error for the 5 year bond rate as we move from the two factor model to the three factor model. In the monthly data set, this standard deviation is reduced from 37 basis points to 16 basis points. In the weekly data set, it is reduced from 34 basis points to 7 basis points. There are reductions in the standard deviations for the other measurement errors in the three factor model, but the improvements are smaller.

During the period 1979-82, there was a shift in Federal Reserve policy toward an emphasis on growth rates of the money supply, and interest rate volatility increased. In Tables IV and V, we present estimates of the CIR model with this period removed from the two samples. The results are similar to those reported in Tables I and II, except for the estimates of the volatility parameters. The estimates for the σ parameters are smaller and in some cases the estimates are much smaller. For example, in the weekly data set the σ estimates for the two factor model decrease from .16885 and .054415 to .08515 and .04579 when the 1979-82 period is removed. The standard deviations for the measurement errors are also smaller when this period is not included in the samples. The other aspects of the results remain the same. In the two and three factor models, there are factors with slow mean reversion (κ estimates close to zero). Most of the risk premium estimates are negative, and the log likelihood function increases significantly as the number of factors is increased.

4. Can the CIR Model Explain the Term Structure of Interest Rates Over Time?

In this section we examine the ability of the CIR models to fit actual bond prices and the different points along the yield curve. For selected days from 1980 through 1992, we have collected all of the available bond prices for the U.S. Treasury market. The flower bonds, the callable bonds, and the coupon issues with less than a year to maturity have been excluded. The specific dates are given in the note to Table VI. Eight dates from June 1989 to December 1992 fall outside of the estimation period. On each day, we compute prices and yields using the three CIR models with the parameter estimates from Table II for the weekly data set, which covers the period 1980-88. The estimates for the state variables are computed from the Kalman filter. The same values for the fixed parameters have been used in the post sample period, 1989-92. To calculate the estimates for the state variables during the post sample period, we use weekly observations on the same bond rates from 1989 to 1992. We then compare the prices from the three CIR models with actual bond prices, and we compare the yields-to-maturity computed from the CIR model prices with yields-to-maturity computed from actual bond prices.

To measure the fit for the three models, we compute root mean square errors for the errors in prices and yields. Both absolute pricing errors and percentage pricing errors are computed. The calculations are summarized in Table VI. The root mean squared errors are similar for the absolute pricing errors and the percentage pricing errors. The relative ranking of the three models is the same across all of the measures of model fit: the three factor model performs marginally better than the two factor model, and both multi-factor models perform much better than the one factor model. For the 1980-88 sample of 2,304 bonds, the multi-factor models have root mean squared errors which are much smaller than the root mean squared errors for the one factor model. The multi-factor models continue to outperform the one factor model in the post sample period. During the 1980-88 sample period, the root mean squared error for prices with the three factor model is 35% lower than the root mean squared error for prices from the two factor model. In terms of yields, the root mean squared error for the three factor model is 19% lower. The three factor model outperforms the two factor model by a small margin in the post sample observations from 1989 to 1992.

What is the nature of the pricing errors of these models? In Figure 5, we present graphs of the yield curve for selected days on which we have complete sets of observations for the Treasury market.[18] In the graphs we plot yield-to-maturity versus duration, a common measure of maturity for bonds. The actual yields are plotted as asterisks against curves for the three CIR models. The one-factor model performs poorly and on many days there are large misses for the long end of the yield curve. In all of the graphs, the two factor model is able to fit both the short end of the yield curve and the long end, but on a few days it is unable to fit the intermediate points. The two factor model captures the general slope of the yield curve, but it occasionally misses the shape or curvature of the yield curve. In almost all of the cases the three factor model fits the general shape as well as the slope of the yield curve. For example, in June 1984, the two factor model misses the yields for bonds with durations of one to five years, whereas the three factor model generally prices these bonds correctly. For the post sample dates, the yield curves for the multi-factor models are close and the actual yields fall very close to the two curves. For many of the days in the sample, the two factor model provides an adequate characterization of the actual yield curve, but there are occasions when a three factor model is necessary to capture the curvature of the yield curve. The one factor model is unable to characterize the changes in the yield curve over time and the errors of the model are economically significant. To make the one-factor model perform well over time, one must regularly adjust the parameter values, but such a procedure is internally inconsistent and suggests that some of the parameters should be treated as state variables. Geyer and Pichler (1999) run additional diagnostic tests to show that the multi-factor CIR models are rejected by the term structure data.

5. *A Monte Carlo Analysis of the Approximate Maximum Likelihood Estimator*

In section (2), we have modified the Kalman filter to impose a nonnegativity restriction and to account for the dependence of the variance of the state variables on previous levels, and the resulting filter is not strictly linear. At a theoretical level, the approximate maximum likelihood estimator can have large sample biases. In this section, we simulate a two factor model to study the properties of the modified Kalman filter and the approximate maximum likelihood estimator. The two factor model has been chosen because it captures much of the variation of the term structure over time, and the approximate maximum likelihood estimator for the two factor model converges at a much faster rate than the estimator for the three factor model.

As shown in CIR (1985b), if *y* is generated as a square root process, the distribution of y_t conditional on y_s is a noncentral χ^2 if we perform the transformation $x = 2c\,y_t$, where

$$c = \frac{2\,\kappa}{\sigma^2\,(1-e^{-\kappa(t-s)})} \quad .$$

The random variable *x* is distributed as a noncentral χ^2 with degrees of freedom, $\nu = {}^{4\kappa\theta}\!/_{\sigma^2}$, and a noncentrality parameter, $2\,c\,e^{-\kappa(t-s)}y_t$. There are several methods available for simulating non-central χ^2 variates, and we have used two methods described in Johnson and Kotz (1970, Ch. 28). They note the following property of this distribution:

$$\chi_\nu^{2\,/}(\delta) \;=\; \chi_1^{2\,/}(\delta) \;+\; \chi_{\nu-1}^2\;,$$

where $\chi_\nu^{2\,/}(\delta)$ is a noncentral χ^2 with ν degrees of freedom and noncentrality parameter δ, and $\chi_{\nu-1}^2$ is a central χ^2 with $\nu-1$ degrees of freedom. The $\chi_1^{2\,/}(\delta)$ variate can be simulated by simulating a standard normal, *Z*, and performing the following transformation: $\chi_1^{2\prime}(\delta)$ = $(Z + \sqrt{\delta})^2$. Then a second simulation from a $\chi^2_{\nu-1}$ is added to this value to produce the simulation. Some of the estimates in our multi-factor models produce degrees of freedom that are less than one and this method cannot be used in these cases. Another method is based on the observation that a noncentral χ^2 variate is a mixture of central χ^2 variates. First simulate the degrees of freedom from a Poisson distribution that has an expected value equal to $\frac{1}{2}\delta$ and then simulate a χ^2 variate with the simulated degrees of freedom. We have used the first method in those cases where the degrees of freedom are greater than one and the second method when the degrees of freedom are less than one. The random number generators for the standard normal, the chi-squared, and the Poisson distributions contained in the International Mathematical and Statistical Library (IMSL) have been used for the simulations.

For the Monte Carlo analysis, we have simulated the model in (1) with two factors and four bond rates. The simulations for the unobservable state variables have been drawn from the noncentral χ^2 distribution as described above, and the measurement errors have been simulated as normal random variables with zero means. The fixed parameter values have been set at the values given in Table 7, which are very close to the estimates for the two factor model in Table 2. The maturities for the bond rates in the simulations are 3 months, 6 months, 5 years, and 30 years.

We checked first the properties of the Kalman filter by running the filter with true values for the parameters. In each sample, we simulated 470 weeks of observations for the bond rates and then calculated the estimates for the state variables. Five hundred independent samples were simulated, and the results are summarized in Table 7 where we report the means and the root mean squared errors for $y_t - \hat{y}_t$. The means for both state variables are very close to zero, which confirms the result that the unconditional expectation of the bias in the quasi linear Kalman filter is zero. The possibility of a conditional bias has not been examined. The root mean squared errors are also small, .00098 and .00065. In basis points, the root mean squared errors are 10 and 7, which suggest that the estimates from the filter are close to the true values, and that the conditional biases may be relatively small.

We have examined the behavior of the approximate maximum likelihood (ML) estimator in two different samples. The first set of simulations is for samples representing 10 years of monthly data. The second set is for samples of weekly data with 470 weeks of observations, which is approximately 9 years of data. In the initial simulations, we found that there were some samples in which the estimate for $\sigma(\varepsilon_2)$ was approaching zero and the estimator did not converge. In these cases, we have fixed $\sigma(\varepsilon_2)$ at a value close to zero and we have restarted the estimator for the remaining parameters. The results are summarized in Panels A and B of Table 8. In both sets of simulations, there are clearly biases in the estimates of the κ, θ, and λ parameters, but no significant biases in the variance parameters, σ_1 and σ_2, and no biases in the standard deviations for the measurement errors. In section (3), we noted that the relevant parameters for asset pricing are the σ, $\kappa+\lambda$, and $\kappa\,\theta$ combinations. We have also reported the simulation results for $\kappa+\lambda$ and $\kappa\,\theta$ and there are no significant biases for these combinations. The approximate ML estimator produces reliable estimates for the standard deviations of the measurement errors and the parameters that are relevant for asset pricing, but there are significant biases in the separate estimates of the κ, θ, and λ parameters. The κ and θ parameters determine some of the time series properties of the factors: the κ parameter measures the rate of mean reversion and the θ

parameter is the long run average.

6. *Summary and Conclusions*

In this paper we have used a state space model to estimate multi-factor versions of the CIR model of the term structure of interest rates. Estimates of the unobservable state variables have been generated by a nonlinear Kalman filter. We find that multi-factor models are necessary to explain the changes over time in the slope and shape of the yield curve. In statistical tests, the two factor model is rejected with the three factor model as the alternative hypothesis. The diagnostic tests in section 4 suggest that the two factor model frequently performs as well as the three factor model, but there are periods when the three factor model is needed to capture the general shape of the yield curve. The three factor model has the added flexibility necessary to explain the random variation in short term interest rates, long term rates, and volatility. In the multi-factor models, the variation of long term rates is explained by factors that experience very slow mean reversion. This aspect of the empirical results is a reflection of the near random walk behavior of long term rates.

The approximate maximum likelihood estimator for the CIR model is one that is potentially biased, even in large samples. The Monte Carlo simulations confirm that there are significant biases in some of the parameter estimators. The significant biases occur in the estimates of κ, θ, and λ. The κ and θ parameters, along with the σ parameters, control the time series properties of interest rates. The κ parameters control the rates of mean reversion and the θ parameters control the long run averages. The λ parameters control the risk adjustments when moving from the real world distribution to the risk neutral distribution for asset pricing. The parameter combinations, $\kappa+\lambda$, $\kappa\theta$, and σ, determine the risk neutral distribution for asset pricing. In contrast, there is no evidence of significant biases in these parameter estimates. Geyer and Pichler (1999) report that the inclusion of more maturities improves the precision of the parameter estimates; the standard errors of the estimates decrease as more maturities are added. The inclusion of additional maturities increases the number of cross sections in the sample, and this would improve the precision of the risk neutral parameter combinations. This conclusion cannot be applied to the estimation of the long run means, the mean reversion parameters, or the risk premia. Throughout the paper, we have alluded to the potential conditional bias in the quasi linear Kalman filter as the cause of the biases in the approximate maximum likelihood estimator. The estimation errors in the Kalman filter appear to be small in the Monte Carlo simulations for the CIR model, and the biases in the κ and θ parameters could be nothing more than the familiar finite sample biases found in autoregressive models. The biases in the λ parameters are the result of the biases in the κ parameters, because the estimates for the $\kappa+\lambda$ combinations do not have significant biases.

Notes

1. Solutions for option and futures prices in two factor versions of the CIR model can be found in Beaglehole and Tenney (1991), Chen and Scott (1992), and Longstaff and Schwartz (1992). Solutions for the multi-factor model can be found in Chen and Scott (1995).
2. Geyer and Pichler (1999) use the same estimation technique developed here to estimate multi-factor CIR models.
3. Specifically the estimation techniques found in Chen and Scott (1993), Longstaff and Schwartz (1992), and Pearson and Sun (1994). For other work on the estimation of CIR models, see Brown and Dybvig (1986), Gibbons and

Ramaswamy (1993), and Stambaugh (1988).

4. There is a relationship between volatility and the curvature of the yield curve, and this aspect of the term structure is examined in Litterman, Scheinkman, and Weiss (1991).
5. This model can be derived by applying arbitrage methods or by using the utility based model in CIR (1985b). The risk premiums are determined endogenously in a utility based model by the covariability of the state variables with marginal utility of wealth. The form for the risk premium used here is consistent with a log utility model.
6. This inversion of bond rates to infer values for the state variables has been used in Chen and Scott (1993), Duffie and Kan (1993), and Pearson and Sun (1994).
7. For an application of the state space model in the finance literature, see Pennacchi (1991).
8. A similar allowance for measurement errors in bond rates was used by Stambaugh (1988).
9. In this case, the Kalman filter is a linear minimum mean squared error estimator and it is unconditionally unbiased. See Harvey (1991, pp. 109-113).
10. For references on Kalman filters, see Chow (1975, pp. 186-95) and Harvey (1991).
11. A minimum mean squared error estimator can be biased.
12. See White (1982) and Vuong (1989).
13. They estimated interest rate volatility by applying a GARCH model to one month T-Bill rates. GARCH is an acronym for generalized autoregressive conditional heteroskedasticity.
14. For those weeks during which Thursday is a holiday, we use Wednesday or Friday prices.
15. Statistical tests for unit roots have very little power and one cannot distinguish empirically in a finite sample whether a time series has a root on or just close to the unit circle.
16. The mean half life is the expected time for the process to return halfway to its long run average. The mean half life is defined as follows: or, where is the mean half life.
17. See Watson and Engle (1983, p.387).
18. We present graphs for 7 days. The graphs for the other 19 days in the sample can be obtained from the authors.

References

Babbs, Simon H., and K. Ben Nowman. (1999). "Kalman Filtering of Generalized Vasicek Term Structure Models," *Journal of Financial & Quantitative Analysis* 34: 115-130.

Beaglehole, David R., and Mark S. Tenney. (1991). "General Solutions of Some Interest Rate Contingent Claim Pricing Equations," *Journal of Fixed Income* 1: 69-83.

Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman. (1974). "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement* 3: 653-65.

Bollerslev, Tim, and Jeffrey M. Wooldridge. (1992). "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews* 11: 143-72.

Brennan, M.J., and E.S. Schwartz. (1979). "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance* 3: 133-55.

Brown, S.J., and P.H. Dybvig. (1986). "The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates," *Journal of Finance* 41: 617-30.

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: 613-36.

Chen, Ren-Raw, and Louis Scott. (1993). "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income* 3: 14-31.

Chen, Ren-Raw, and Louis Scott. (1995). "Pricing Interest Rate Options in a Multi-Factor Cox-Ingersoll-Ross Model of the Term Structure," *Journal of Derivatives* : 53-71.

Chow, G.C. (1975). *Analysis and Control of Dynamic Economic Systems*. New York: John Wiley.

Cooley, T.F., S.F. LeRoy, and W.R. Parke. (1992). "Pricing Interest-Sensitive Claims When Interest Rates Have Stationary Components," *Journal of Fixed Income* 2: 64-73.

Cox, J.C., J.E. Ingersoll, and S.A. Ross. (1985a). "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica* 53: 363-84.

Cox, J.C., J.E. Ingersoll, and S.A. Ross. (1985a). "A Theory of the Term Structure of Interest Rates," *Econometrica* 53: 385-408.

Dai, Q., and K.J. Singleton (2000). “Specification Analysis of Affine Term Structure Models,” *Journal of Finance* 55: 1943-78.

Duffie, Darrell, and Rui Kan. (1996). "A Yield-Factor Model of Interest Rates," *Mathematical Finance* 6: 379-406.

Duffie, Darrell, and K. Singleton (1997). "An Econometric Model of the Term Structure of Interest Rate Swap Yields," *Journal of Finance* 52: 1287-1321.

Engle, R., and M. Watson. (1981). "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the American Statistical Association* 76: 774-81.

Geyer, Alois L.J., and Stefan Pichler. (1981). "A State-Space Approach to Estimate and test Multifactor Cox-Ingersoll-Ross Models of the Term Structure," *Journal of Financial Research* 22: 107-130.

Gibbons, M.R., and K. Ramaswamy. (1993). "A Test of the Cox, Ingersoll, and Ross Model of the Term Structure," *Review of Financial Studies* 6: 619-58.

Hansen, L.P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50: 1029-54.

Harvey, Andrew C. (1991). *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press.

Johnson, N.L., and S. Kotz. (1970). *Distributions in Statistics: Continuous Univariate Distributions* II. Boston: Houghton Mifflin.

Longstaff, Francis, and Eduardo S, Schwartz. (1992). "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance* 47: 1259-82.

Litterman, Robert, and José Scheinkman. (1991). "Common Factors Affecting Bond Returns," *Journal of Fixed Income* 1: 54-61.

Litterman, Robert, José Scheinkman, and Laurence Weiss. (1991). "Volatility and the Yield Curve," *Journal of Fixed Income* 1: 49-53.

Pearson, N.D., and T.S. Sun. (1994). "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, Ross Model," *Journal of Finance* 49: 1279-1304.

Pennacchi, George G. (1991). "Identifying the Dynamics of Real Interest Rates and Inflation: Evidence Using Survey Data," *Review of Financial Studies* 4: 53-86.

Richard, S.F. (1978). "An Arbitrage Model of the Term Structure of Interest Rates," *Journal of Financial Economics* 6: 33-57.

Sheil, J. and I. O'Muircheartaigh. (1977). "The Distribution of Non-Negative Quadratic Forms in Normal Variables," Algorithm AS 106, *Applied Statistics* 26: 92-98.

Singleton, K.J. (1980). "A Latent Time Series Model of the Cyclical Behavior of Interest Rates," *International Economic Review* 32: 559-75.

Stambaugh, R.F. (1988). "The Information in Forward Rates," *Journal of Financial Economics* 21: 41-70.

Vuong, Q.H. (1989). "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica* 57: 307-33.

Watson, M., and R. Engle. (1983). "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC, and Varying Coefficient Regression Models," *Journal of Econometrics* 23: 385-400.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50: 1-25.

Table I

Estimates from Monthly Data, 1960-87
Sample Size, T=326

	One Factor Model	Two Factor Model		Three Factor Model		
κ	.07223	.6402	.01700	1.3683	.08433	.008428
	(.02646)	(.1576)	(.01901)	(.1649)	(.07688)	(.02389)
θ	.03739	.03080	.00003265	.02979	.0006553	.0007228
	(.01426)	(.00721)	(.00003647)	(.00296)	(.0006024)	(.0008492)
σ	.07540	.1281	.05547	.1231	.1355	.04883
	(.003801)	(.0099)	(.00263)	(.0061)	(.00803)	(.00217)
λ	-.07892	-.1744	-.04076	-.3229	-.04425	-.05830
	(.02877)	(.1516)	(.01886)	(.1375)	(.07810)	(.02372)
$\sigma(\varepsilon_1)$	.003324	.003103		.002974		
	(.0001956)	(.0002370)		(.0001167)		
$\sigma(\varepsilon_2)$	0.0	.0007315		0.0		
		(.0003378)				
$\sigma(\varepsilon_3)$	.01022	.003709		.001624		
	(.0004331)	(.0001894)		(.00008175)		
$\sigma(\varepsilon_4)$	.01320	.0009302		.0005319		
	(.0003583)	(.00006960)		(.00006981)		
ln L	5,828.771	6,730.114		6,954.484		
AIC	-11,642	-13,436		-13,877		

NOTES: The numbers in parentheses are standard errors. AIC is the Akaike information criterion.

Table II

Estimates from Weekly Data, 1980-88
Sample Size, T=470

	One Factor Model	Two Factor Model		Three Factor Model		
κ	.13974	.7298	.021185	1.4298	.01694	.03510
	(.03408)	(.3013)	(.004139)	(.2761)	(.11319)	(.02776)
θ	.08480	.04013	.022543	.04374	.002530	.003209
	(.02050)	(.01660)	(.003616)	(.00838)	(.016991)	(.002543)
σ	.10001	.16885	.054415	.16049	.1054	.04960
	(.003846)	(.01015)	(.002786)	(.01047)	(.00679)	(.003148)
λ	-.07132	-.01731	-.044041	-.2468	.03411	-.1569
	(.03475)	(.30105)	(.005692)	(.2635)	(.11281)	(.03077)
$\sigma(\varepsilon_1)$	.0039566	.0034988		.0031873		
	(.0002126)	(.0002000)		(.0001041)		
$\sigma(\varepsilon_2)$	0.0	0.0		0.0		
$\sigma(\varepsilon_3)$	.010443	.0033555		.0007033		
	(.0003792)	(.0001149)		(.0000461)		
$\sigma(\varepsilon_4)$	.012249	.0007003		0.0		
	(.0002952)	(.0000582)				
ln L	8,505.09	10,008.65		10,424.24		
AIC	-16,994	-19,993		-20,816		

NOTES: The numbers in parentheses are standard errors. AIC is the Akaike information criterion.

Table III
Estimates of Parameter Combinations for Asset Pricing

A. The Two Factor Model

1. Monthly Data Set, 1960-87

$\kappa_1 + \lambda_1 =$	0.4659 (0.03605)	$\kappa_1\theta_1 =$	0.01972 (0.001386)	$\sigma_1 =$	0.1281 (0.009914)
$\kappa_2 + \lambda_2 =$	-0.02376 (0.003677)	$\kappa_2\theta_2 =$	0.0000005551 (0.000001152)	$\sigma_2 =$	0.05547 (0.002626)

2. Weekly Data Set, 1980-88

$\kappa_1 + \lambda_1 =$	0.7125 (0.04084)	$\kappa_1\theta_1 =$	0.029287 (0.001759)	$\sigma_1 =$	0.16885 (0.01015)
$\kappa_2 + \lambda_2 =$	-0.02286 (0.002782)	$\kappa_2\theta_2 =$	0.0004776 (0.00009256)	$\sigma_2 =$	0.054415 (0.002786)

B. The Three Factor Model

1. Monthly Data Set, 1960-87

$\kappa_1 + \lambda_1 =$	1.0454 (0.1094)	$\kappa_1\theta_1 =$	0.04076 (0.003988)	$\sigma_1 =$	0.1231 (0.006088)
$\kappa_2 + \lambda_2 =$	0.04008 (0.01536)	$\kappa_2\theta_2 =$	0.00005526 (0.00005830)	$\sigma_2 =$	0.1355 (0.008030)
$\kappa_3 + \lambda_3 =$	-0.04988 (0.003933)	$\kappa_3\theta_3 =$	0.000006092 (0.00001563)	$\sigma_3 =$	0.04883 (0.002173)

2. Weekly Data Set, 1980-88

$\kappa_1 + \lambda_1 =$	1.1830 (0.08895)	$\kappa_1\theta_1 =$	0.06253 (0.006666)	$\sigma_1 =$	0.16049 (0.01047)
$\kappa_2 + \lambda_2 =$	0.05105 (0.01868)	$\kappa_2\theta_2 =$	0.00004286 (0.00058632)	$\sigma_2 =$	0.105392 (0.006794)
$\kappa_3 + \lambda_3 =$	-0.12177 (0.01394)	$\kappa_3\theta_3 =$	0.0001126 (0.00008353)	$\sigma_3 =$	0.049599 (0.003148)

NOTE: The numbers in parentheses are standard errors.

Table IV

Estimates from Monthly Data, 1960-78
Sample Size, T=228

	One Factor Model	Two Factor Model		Three Factor Model		
κ	.06462	.7544	$.2652 \times 10^{-8}$	1.4849	.08472	.00003807
	(.01985)	(.1439)	(.2115)	(.1797)	(.06500)	(.07716)
θ	.04015	.03385	$.1794 \times 10^{-10}$	.02835	.003658	$.25 \times 10^{-8}$
	(.01241)	(.00531)	(.001428)	(.00254)	(.002873)	$(.51 \times 10^{-5})$
σ	.06514	.08632	.04395	.09085	.1240	.04041
	(.003053)	(.00768)	(.00408)	(.00639)	(.01034)	(.00268)
λ	-.05576	-.1213	-.02022	-.3888	.07104	-.03795
	(.02072)	(.1202)	(.2143)	(.1351)	(.07061)	(.07766)
$\sigma(\varepsilon_1)$	.003002	.002659		.002519		
	(.0001642)	(.0002196)		(.0001400)		
$\sigma(\varepsilon_2)$	0.0	.0008400		.0002344		
		(.0002357)		(.0006357)		
$\sigma(\varepsilon_3)$	.007394	.002763		.001410		
	(.0003287)	(.0001948)		(.00008235)		
$\sigma(\varepsilon_4)$	.009943	.0008082		.0007660		
	(.0003387)	(.00006897)		(.00007019)		
ln L	4,312.428	4,944.305		5,074.019		
AIC	-8,609	-9,865		-10,116		

NOTES: The numbers in parentheses are standard errors. AIC is the Akaike information criterion.

Table V

Estimates from Weekly Data, 1983-88
Sample Size, T=313

	One Factor Model	Two Factor Model		Three Factor Model		
κ	.07951	1.5446	.01265	2.07588	.004553	.002329
	(.008763)	(.2530)	(.003292)	(.2601)	(.09144)	(.05131)
θ	.04109	.02638	.02120	.02786	.001125	.0005463
	(.004081)	(.006711)	(.004718)	(.00503)	(.02262)	(.01201)
σ	.1033	.08515	.04579	.08388	.08477	.04550
	(.001388)	(.01120)	(.005044)	(.009100)	(.006595)	(.003724)
λ	-.1428	-.7289	-.02881	-.9228	-.01284	-.09760
	(.008840)	(.2388)	(.007158)	(.2462)	(.09022)	(.05391)
$\sigma(\varepsilon_1)$	.002472	.001711		.001414		
	(.0001306)	(.0001103)		(.0000705)		
$\sigma(\varepsilon_2)$	0.0	.0005836		.0007931		
		(.00008857)		(.00006275)		
$\sigma(\varepsilon_3)$	.004964	.002897		.0005658		
	(.0001716)	(.0001418)		(.00004888)		
$\sigma(\varepsilon_4)$	.007055	.0007848		.0001197		
	(.0002334)	(.00005392)		(.0002590)		
ln L	6,308.50	7,104.68		7,439.12		
AIC	-12,601	-14,185		-14,846		

NOTES: The numbers in parentheses are standard errors. AIC is the Akaike information criterion.

Table VI

Comparisons of Actual Bond Prices with Prices from the Estimated CIR Models

	Root Mean Squared Errors		
	Absolute Pricing Errors	Percentage Pricing Errors	Yield-to-Maturity (Basis Points)
A. 1980-88 (2304 Bonds)			
One Factor Model	$3.00	3.12%	87
Two Factor Model	1.34	1.35	47
Three Factor Model	0.87	0.89	38
B. 1989-92, Out-of-Sample Period (1160 Bonds)			
One Factor Model	$2.38	2.24%	56
Two Factor Model	.71	.66	23
Three Factor Model	.68	.64	21

Note: All prices are calculated for $100 of par value. 100 basis points represents 1% in the yield. Our rule for selecting days has been to pick a Thursday every six months beginning with June 1980. The 1980-88 sample contains bond prices from the following days:

June 5, 1980, Dec. 4, 1980, June 4, 1981, Dec. 3, 1981,
June 3, 1982, Dec. 2, 1982, June 2, 1983, Dec. 1, 1983,
June 7, 1984, Dec. 6, 1984, June 6, 1985, Dec. 5, 1985,
June 5, 1986, Dec. 4, 1986, June 4, 1987, Dec. 3, 1987,
June 2, 1988, Dec. 1, 1988.

The 1989-92 sample contains bond prices for June 1, 1989, Dec. 7, 1989, and June 7, 1990, December 6, 1990, June 6, 1991, December 5, 1991, June 4, 1992, and December 3, 1992. Over the selected days all Treasury bond prices were used except flower bonds, callable bonds, and coupon bonds with maturities less than a year.

Table 7
A Simulation Analysis of the Kalman Filter

A. Parameter Values for Simulations

Factor	κ	θ	σ	λ
1	.7298	.04013	.1688	-.0173
2	.02118	.02254	.05442	-.04404

$\sigma(\varepsilon_1) = .003499$ $\sigma(\varepsilon_2) = .0005$ $\sigma(\varepsilon_3) = .003355$ $\sigma(\varepsilon_4) = .0007$

B. Results for 500 Independent Simulations
The Number of Time Series Observations for Each Simulation is 470

$y_1 - \hat{y}_1$		$y_2 - \hat{y}_2$	
Mean	Root Mean Squared Error	Mean	Root Mean Squared Error
$-.74 \times 10^{-8}$	.00098	$.23 \times 10^{-7}$	.00065

Table 8
A Simulation Analysis of the Approximate ML Estimator

A. 500 Simulations of Monthly Samples
Each Sample Covers 10 Years of Monthly Data, $T = 120$

	κ_1	θ_1	σ_1	λ_1	$\kappa_1 + \lambda_1$	$\kappa_1\theta_1$
Parameter Value	.7298	.04013	.1688	-.0173	.7125	.029287
Mean	.8676	.03757	.1682	-.1381	.7295	.030129
Std. Error	.2609	.01201	.0187	.2421	.0746	.004601

	κ_2	θ_2	σ_2	λ_2	$\kappa_2 + \lambda_2$	$\kappa_2\theta_2$
Parameter Value	.02118	.02254	.05442	-.04404	-.02286	.000477
Mean	.05218	.01021	.05486	-.07598	-.02381	.000490
Std. Error	.01683	.00422	.00071	.02438	.01165	.000155

	$\sigma(\varepsilon_1)$	$\sigma(\varepsilon_2)$	$\sigma(\varepsilon_3)$	$\sigma(\varepsilon_4)$
Parameter Value	.003499	.0005	.003355	.0007
Mean	.003405	.000603	.003289	.000708
Std. Error	.000278	.000587	.000239	.000194

B. 500 Simulations of Weekly Samples
Each Sample is 470 Weeks, $T = 470$

	κ_1	θ_1	σ_1	λ_1	$\kappa_1 + \lambda_1$	$\kappa_1\theta_1$
Parameter Value	.7298	.04013	.1688	-.0173	.7125	.029287
Mean	.8526	.03748	.1679	-.1348	.7178	.029713
Std. Error	.2419	.01065	.0101	.2377	.0348	.002710

Table 8 cont.

	κ_2	θ_2	σ_2	λ_2	$\kappa_2 + \lambda_2$	$\kappa_2\theta_2$
Parameter Value	.02118	.02254	.05442	-.04404	-.02286	.000477
Mean	.04899	.01017	.05458	-.07248	-.02348	.000476
Std. Error	.01015	.00290	.00462	.01469	.00723	.000079

	$\sigma(\varepsilon_1)$	$\sigma(\varepsilon_2)$	$\sigma(\varepsilon_3)$	$\sigma(\varepsilon_4)$
Parameter Value	.003499	.0005	.003355	.0007
Mean	.003484	.000494	.003339	.000702
Std. Error	.000123	.000263	.000105	.000050

NOTE: The standard errors are the standard deviations for the simulated values. The standard error for the mean is equal to the standard error (in the table) divided by the square root of T, where T is 120 in panel A and 470 in panel B.

Figure 1

Factor Loadings, 2 factor Model Estimated from Monthly Data

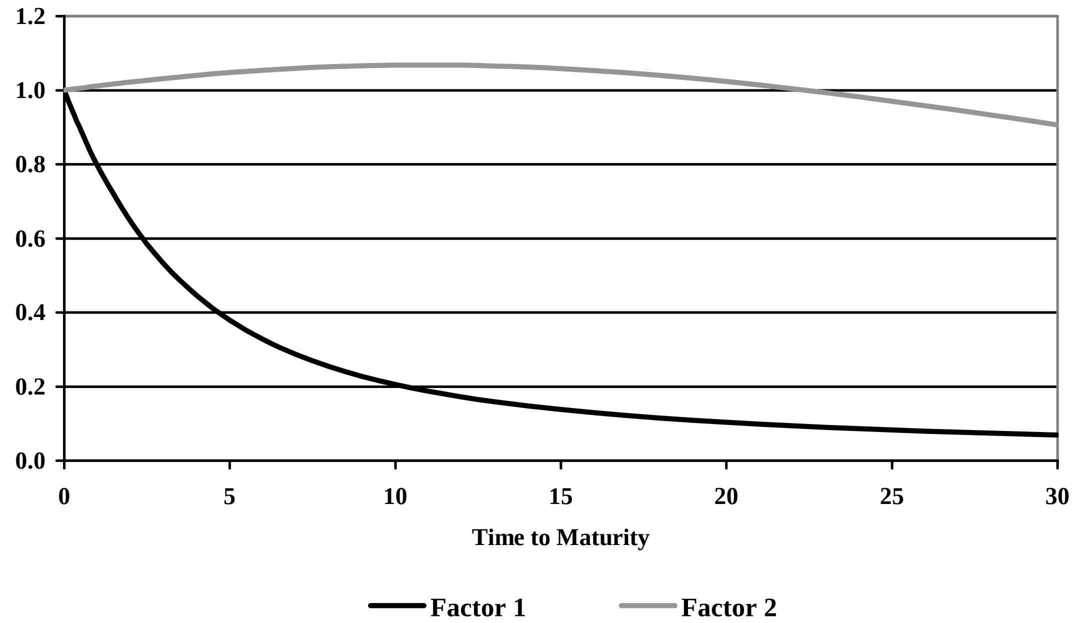


Figure 2

Factor Loadings, 3 factor Model Estimated from Monthly Data

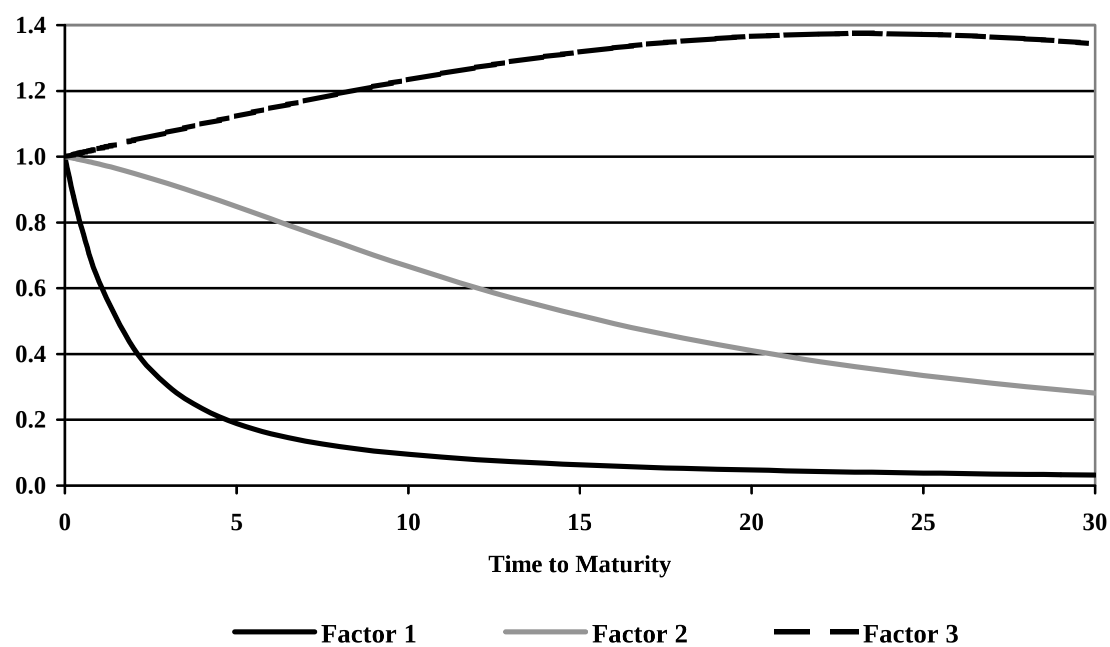


Figure 3
Factor Loadings, 2 factor Model Estimated from Weekly Data

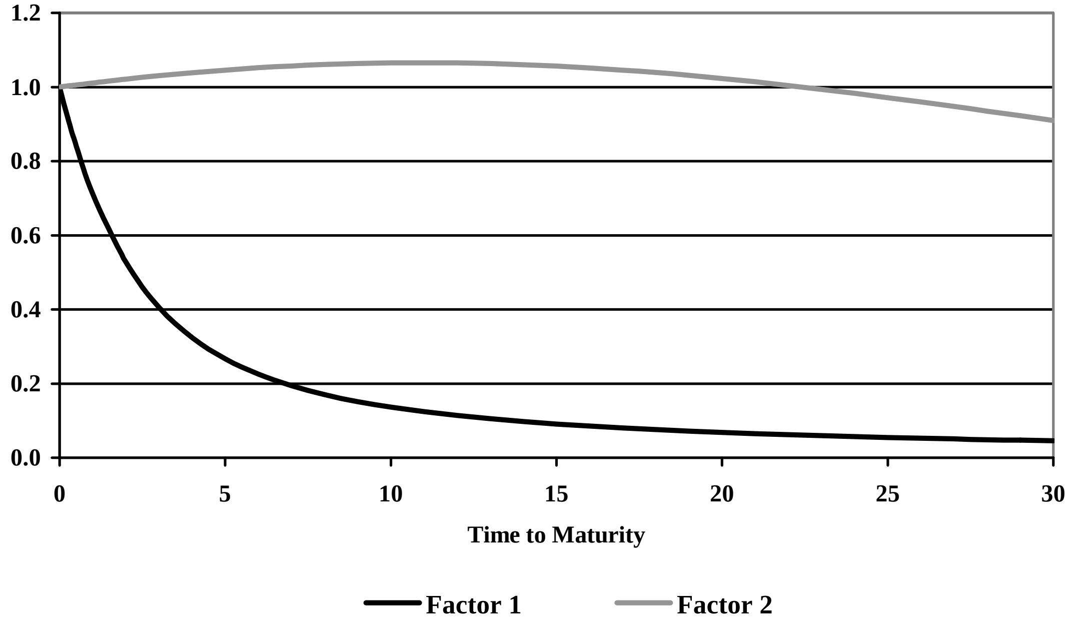


Figure 4
Factor Loadings, 3 factor Model Estimated from Weekly Data

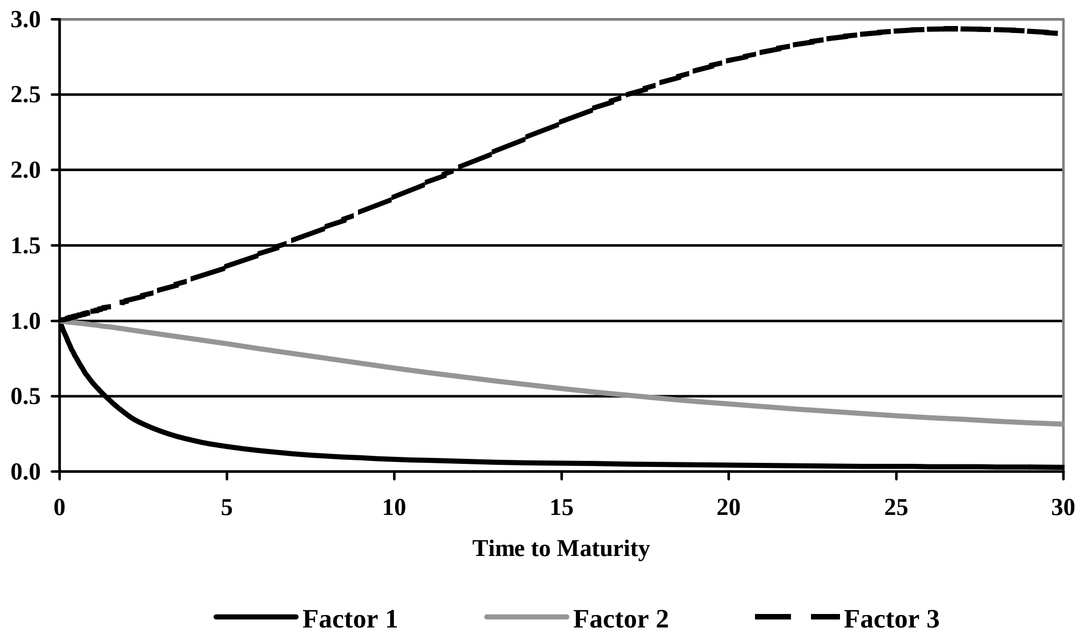
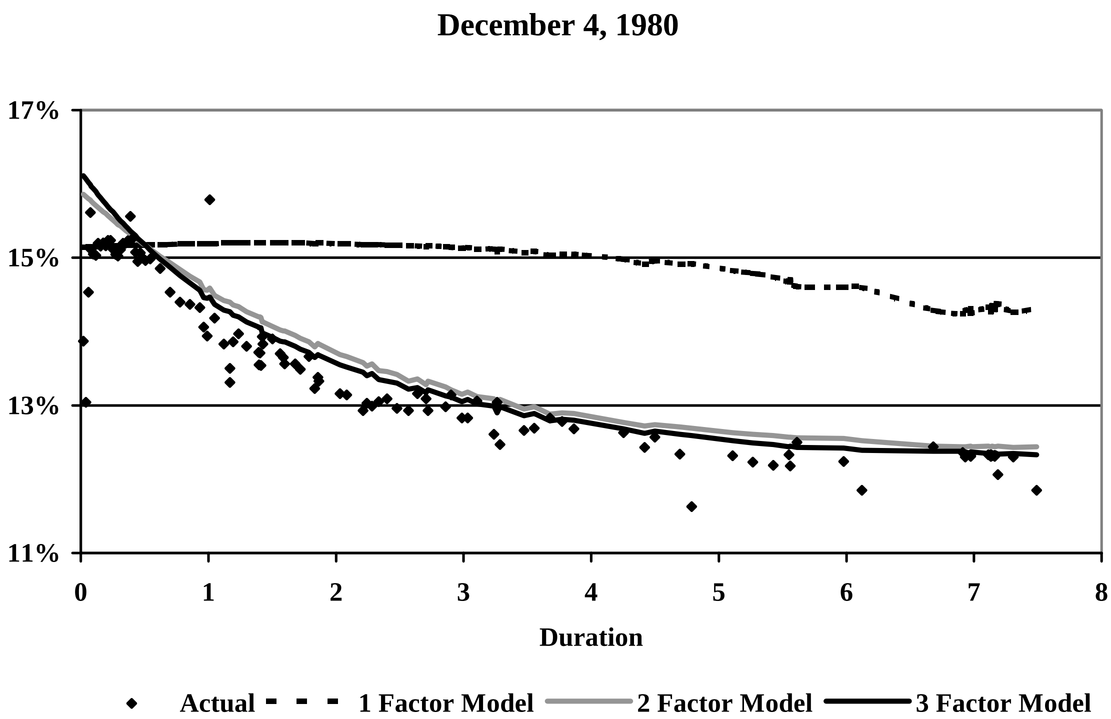


Figure 5
Yield Curve Plots, Actual Yields with Model Yields



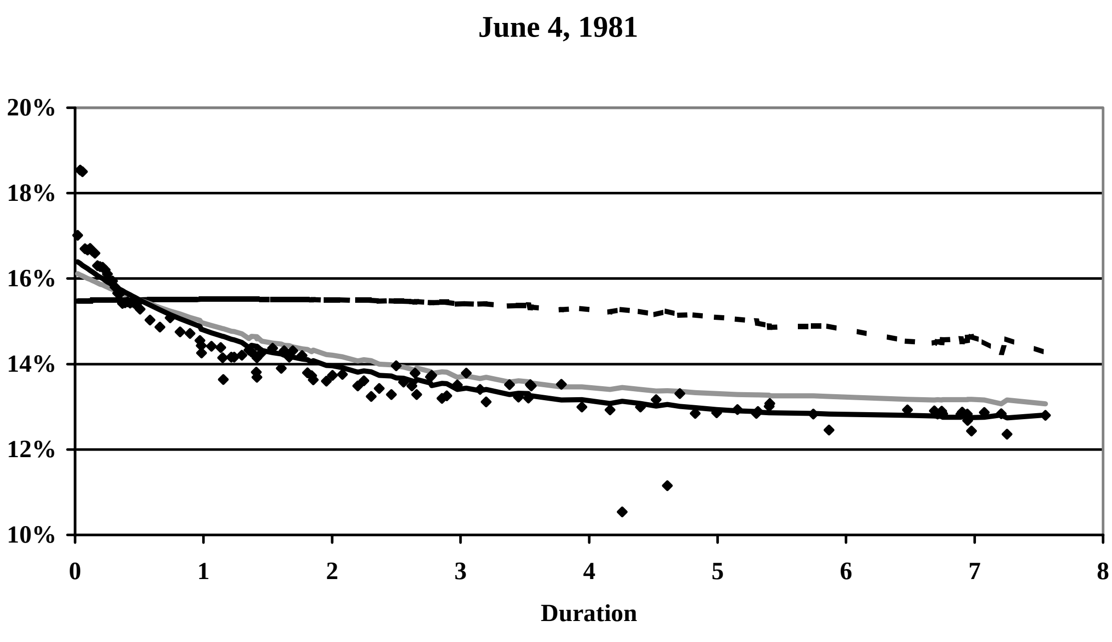
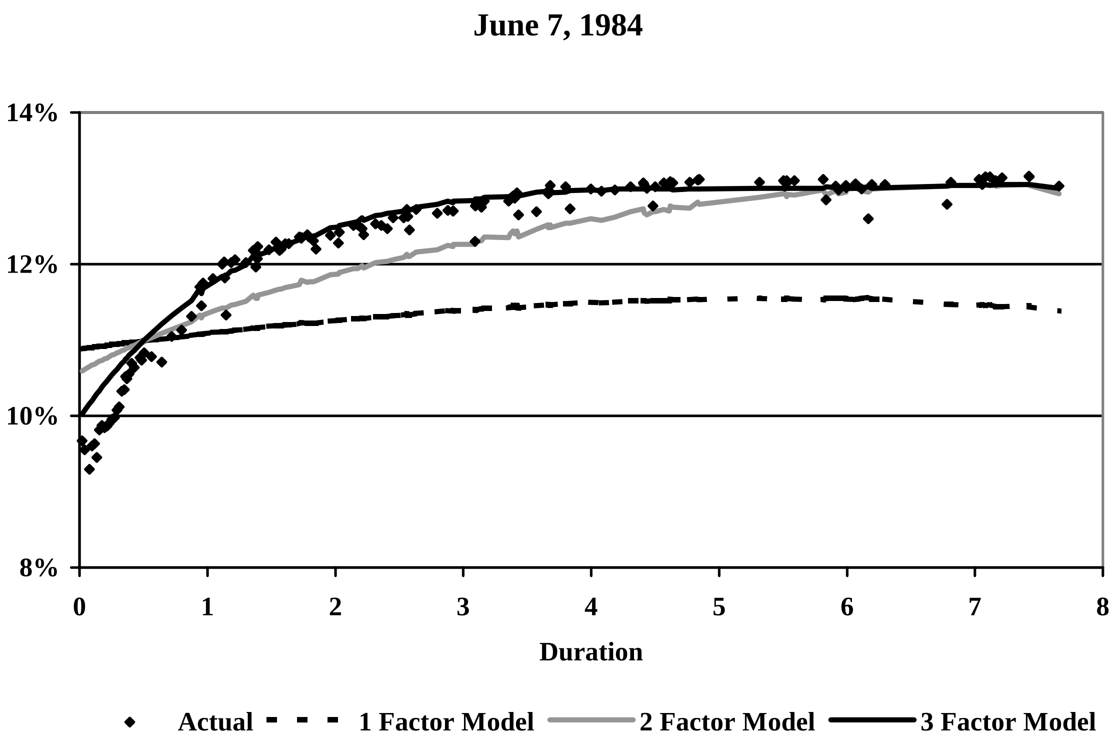


Figure 5 (continued)
Yield Curve Plots, Actual Yields with Model Yields



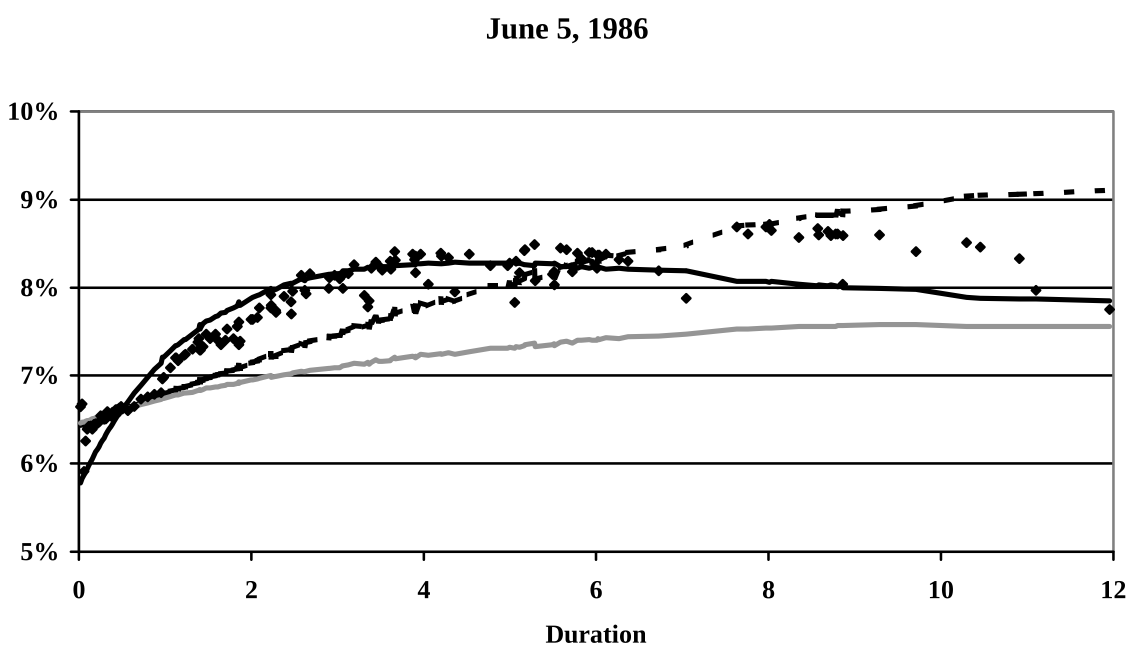
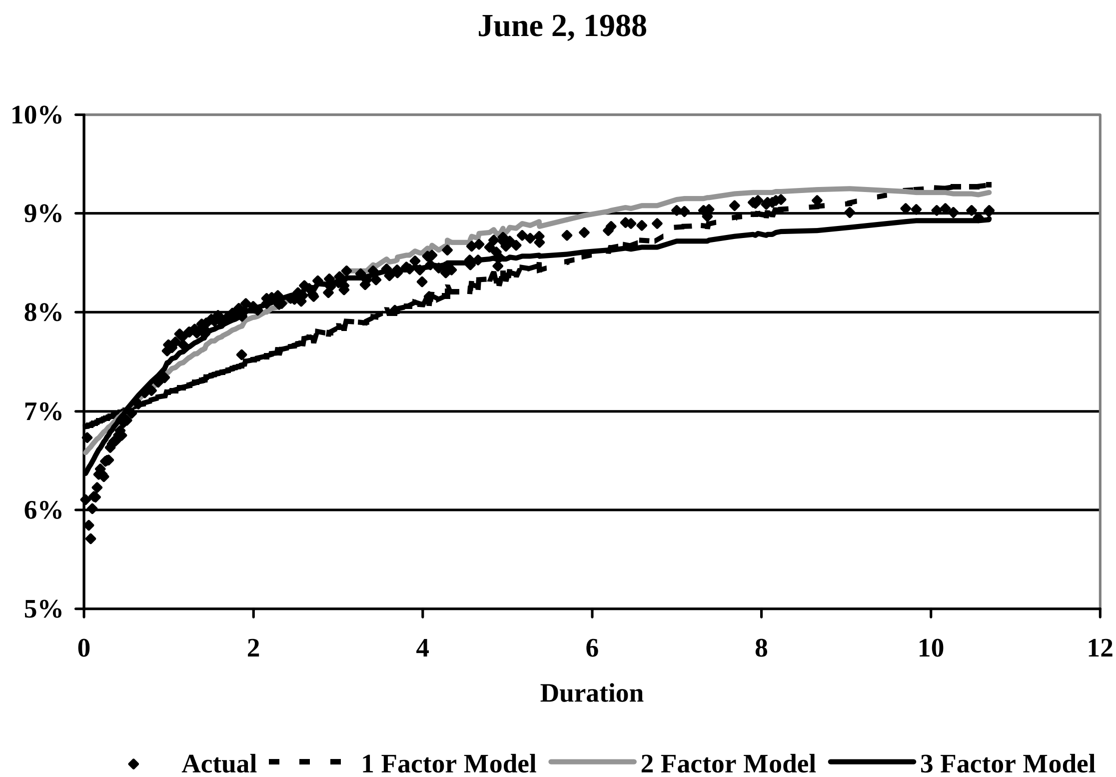


Figure 5 (continued)
Yield Curve Plots, Actual Yields with Model Yields



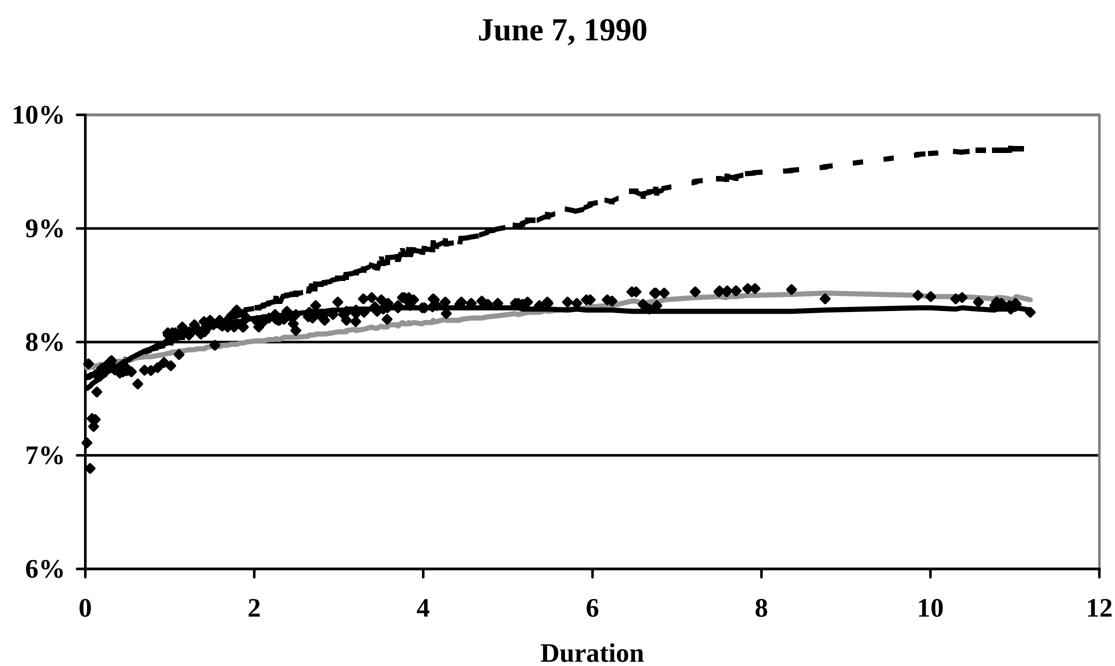


Figure 5 (continued)
Yield Curve Plots, Actual Yields with Model Yields