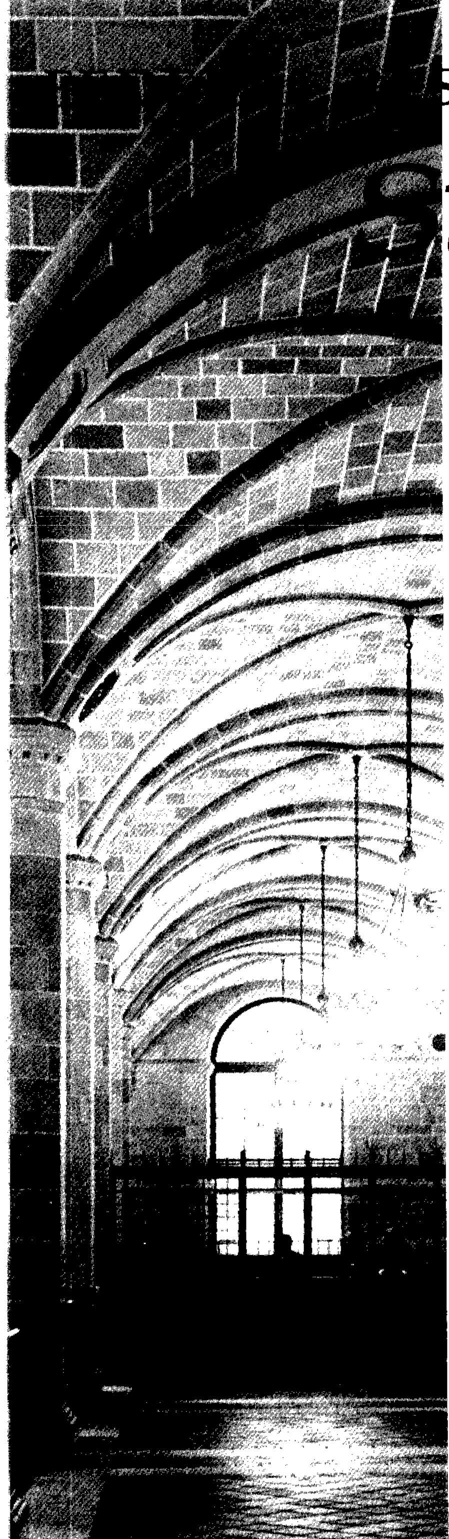

SERVE BANK OF NEW YORK

Staff Reports

A Three-Factor Econometric Model of the U.S. Term Structure

Frank F. Gong and Eli M. Remolona

Number 19
January 1997

A Three-Factor Econometric Model of the U.S. Term Structure

FRANK F. GONG and ELI M. REMOLONA

Capital Markets Function
Federal Reserve Bank of New York
New York, NY 10045
Tel: (212)720-6943(Gong)
(212)720-6328(Remolona)
E-mail: frank.gong@frbny.sprint.com
eli.remolona@frbny.sprint.com

The authors thank John Clark, Young Ho Eom, Arturo Estrella, Ken Kuttner, Francis Longstaff, Jim Mahoney, Tony Rodrigues, and Christopher Sims for helpful comments, and Debbie Gruenstein for research assistance.

A Three-Factor Econometric Model of the U.S. Term Structure

Abstract

We estimate a three-factor model to fit both the time-series dynamics and cross-sectional shapes of the U.S. term structure. In the model, three unobserved factors drive a discrete-time stochastic discount process, with one factor reverting to a fixed mean and a second factor reverting to a third factor. To exploit the conditional density of yields, we estimate the model with a Kalman filter, a procedure that also allows us to use data for six maturities without making special assumptions about measurement errors. The estimated model reproduces the basic shapes of the average term structure, including the hump in the yield curve and the flat slope of the volatility curve. A likelihood ratio test favors the model over a nested two-factor model. Another likelihood ratio test, however, rejects the no-arbitrage restrictions the model imposes on the estimates. An analysis of the measurement errors suggests that the three factors still fail to capture enough of the comovement and persistence of yields.

JEL Codes: E43, G12, G13.

Keywords: Term structure, pricing kernel, affine yields, mean reversion, time-varying mean, Kalman filter.

A Three-Factor Econometric Model of the U.S. Term Structure

I. Introduction

The reconciliation of the time-series dynamics of yields with the cross-sectional shapes of the term structure continues to be a challenge for equilibrium models. When such models relied on a single factor, the problem was that estimates based on time-series data failed to reproduce the average shapes of actual curves.[1] With estimates based on cross-section data, the problem was that parameter values necessarily changed with shifts in the term structure.[2] The advent of multi-factor models has transformed the challenge from one of capturing the yield curve's unconditional moments into one of capturing its conditional moments, which indeed is what the pricing of actual fixed-income instruments would require. In this paper, we estimate a three-factor model of the U.S. term structure to exploit the yields' conditional density and test the

[1] Backus and Zin (1994) and Campbell, Lo, and Mackinlay (1994) emphasize this point. The latter state, "But in simple term structure models, there also appear to be systematic differences between the parameter values needed to fit cross-section term structure data and the parameter values implied by the time-series behavior of interest rates." Part of the problem was that the models assumed that yields were affine, i.e., linear functions of the short rate. However, Chan, et al. (1992), Ait-Sahalia (1996), Eom (1995), Stanton (1996), Andersen and Lund (1996), and Gallant and Tauchen (1996b) provide evidence on the short rate's movements that suggest non-affine yields.

[2] To market participants, it is more critical that their model be consistent with the cross section than with the time series, particularly for pricing contingent claims. However, as Black and Karasinski (1991) point out, relying solely on cross section data means having a different model from one moment to the next. Among the popular cross-section models are Ho and Lee (1986), Black, Derman, and Toy (1990), Hull and White (1990), and Heath, Jarrow, and Morton (1992).

model's ability to meet the current challenge.

The efforts to reconcile time-series data with the average shapes of the term structure have tended to rely on the generalized method of moments (GMM), and in practice this meant placing cross-sectional restrictions on unconditional moments. Gibbons and Ramaswamy (1993), for example, impose restrictions on both the unconditional first and second moments of the short rate to fit the one-factor model of Cox, Ingersoll, and Ross (1985), or the CIR model. Backus and Zin (1994) place restrictions on the unconditional first moments of several yields up to the ten-year maturity. Longstaff and Schwartz (1992) estimate a two-factor model, with the short rate and its volatility as the two factors, and impose restrictions on unconditional first moments from a reduced-form model. The use of a model for the pricing of actual instruments, however, would require the far more stringent test of fitting the conditional moments, that is, of capturing the term structure's shapes at each point in time.

Maximum likelihood procedures have recently allowed the use of conditional moments to estimate term structure models but have often required special assumptions about measurement errors. Chen and Scott (1993) estimate a one-factor model, a two-factor model, and a three-factor model by maximizing a likelihood based on the factors' conditional moments. To derive the factors, however, their procedure requires the assumption of zero measurement error for as many yields as they have factors. Their estimates lead to the puzzling result that a likelihood ratio test for non-nested hypotheses rejects their three-factor model in favor of their two-factor model at the same time that the three-factor model outperforms the two-factor model in pricing bonds over time. Pearson and Sun (1994) also exploit the factors' conditional density in estimating a two-factor CIR model. They derive the factors by using two yields at a time and assuming no measurement error for either yield. They find that a likelihood ratio test rejects the CIR model and an extended model performs as poorly as a naive model.

The number of factors required for an adequate term structure model is a related issue. The point is to build a consistent model with as few factors as possible. Litterman and Scheinkman (1991) show that three factors can explain nearly all the variation in bond returns. They interpret their factors as representing the level of interest rates, the slope of the yield curve, and the curvature of the yield curve. However, they do not ensure that their factor loadings are consistent with no arbitrage. Gong and Remolona (1996) are careful to impose no arbitrage when they fit three alternative two-factor models to U.S. quarterly yield data. However, they fail to find a model that is adequate for explaining the whole term structure, and they conclude that at least three factors would be required for that purpose. Using Gallant and Tauchen's (1996a) efficient method of moments (EMM), Andersen and Lund (1996) find that a three-factor model explains the movements of the short-term rate well. However, they do not fit their model to the term structure.

In this paper, we specify a discrete-time, affine-yield, three-factor model. We follow Backus and Zin (1994) and Campbell, Lo, and MacKinlay (1994), or CLM, by specifying the model in terms of a discrete-time stochastic discount process or pricing kernel. Specifying the model in discrete time avoids the pitfalls of estimating a continuous time model with discrete-time data.[3] In this model, three unobserved factors drive the pricing kernel: one factor reverts over time to a fixed mean while a second factor reverts to a time-varying mean

[3] Ait-Sahalia (1996) points out that the approximation of a continuous time process by discretization methods is hard to justify even for daily data.

that serves as the third factor. Shocks to the third factor represent a risk priced by the market. To maintain tractability, we write the model to satisfy Duffie and Kan's (1993) conditions for affine yields. As in Gong and Remolona (1996), we keep the model affine in part by specifying a single source of priced risk.[4]

To estimate the model, we exploit the conditional density of the yields by using a Kalman filter while imposing restrictions implied by arbitrage conditions. Jegadeesh and Pennacchi (1996) and Gong and Remolona (1996) implement this maximum likelihood procedure for two-factor models but do not test the arbitrage conditions. The procedure has the advantage of letting the data determine the measurement errors, and we use these errors to test the model. The data consist of monthly U.S. Treasury zero-coupon yields from January 1984 to March 1995. We use six maturities: three months, six months, one year, two years, five years, and ten years.

The estimated model performs well in certain respects but fails its most critical test. The model reproduces the basic shapes of the average term structure, including the hump in the yield curve and the flat slope of the volatility curve. A likelihood ratio test for nested hypotheses favors the model over a two-factor version. The real test of the model, however, is whether the arbitrage restrictions it implies are consistent with yield movements across the term structure. A strict likelihood ratio test rejects these arbitrage restrictions. Our analysis of the measurement errors suggests that the three factors are still insufficient to capture enough of the comovement and persistence of yields.

The paper is organized as follows: In Section II, we discuss pricing-kernel

[4]It is possible to specify a discrete-time affine-yield model with more than one source of priced risk if these sources are uncorrelated. Here, however, we opted to allow correlation.

affine yield models in general and specify our three-factor model. In Section III, we estimate the model by means of a Kalman filter. In Section IV, we interpret the results in terms of reproducing the average shapes of the term structure. In Section V, we test the model against a nested two-factor version. In Section VI, we test the arbitrage restrictions of the model. In Section VII, we discuss the reasons for the model's failure.

II. Theory: Pricing Kernels, Affine Yields, and Three Factors

A. Background Literature

Theoretical work with equilibrium models, notably by Vasicek (1977) and Cox, Ingersoll, and Ross (1985), or CIR, show how the term structure at a moment in time would reflect regularities in interest rate movements over time. In the simplest such models, the short-term interest rate is the single factor driving movements in the term structure. Vasicek assumes that the short-rate's volatility is constant, while CIR assume that it is proportional to the square root of the short rate itself. The absence of arbitrage requires that the ratio of expected excess return to return volatility be the same for different bonds. This arbitrage condition, the assumption of lognormal bond prices, and either Vasicek's or CIR's short-rate volatility produce an affine-yield solution in which all bond yields (or log bond prices) are linear functions of the short-term rate. Such linearity simplifies the pricing of fixed-income securities and contingent claims. In the two-factor models of Brennan and Schwartz (1982), Schaefer and Schwartz (1984), Longstaff and Schwartz (1992), and CLM (1994) and in the three-factor model of Chen and Scott (1993), similar assumptions generate bond yields that are also linear in the factors. Duffie and Kan (1993) establish the conditions that produce such affine yields in general.

Rather than model the short-term interest rate directly, Backus and Zin (1994) and CLM (1994) specify a discrete-time stochastic discount process or pricing kernel to price assets in general. The discrete-time specification avoids the pitfalls of estimating continuous-time processes with discrete-time data. Ait-Sahalia (1996) points out that the approximation of continuous-time models by discretization methods is hard to justify even with high-frequency data. Arbitrage opportunities are avoided by applying the same pricing kernel to different assets. In this approach, the factors are unobservable state variables that serve to forecast discount rates. Pricing kernel models can also be specified so that bond yields are affine in the factors and, with a linear transformation, affine in the short rate as well. We describe below such a pricing-kernel affine-yield model with K factors.

B. The Pricing Kernel

The pricing kernel approach relies on a no-arbitrage condition common to intertemporal asset pricing models.[5] In the case of zero-coupon bonds, the price of an n-period bond is

$$P_{nt} = E_t[M_{t+1} P_{n-1,t+1}], \tag{1}$$

where M_{t+1} is the stochastic discount factor. The condition expresses the price of the bond as the expected discounted value of the bond's next-period price. It rules out arbitrage opportunities by applying the same discount factor to

[5]Singleton (1990) provides a critical survey of these models, particularly their empirical performance. Duffie (1992) relates arbitrage conditions to concepts of optimality and equilibrium.

all bonds. We will model P_{nt} by modeling the stochastic process for M_{t+1}, a process called the pricing kernel.[6] Indeed we can solve (1) forward to get $P_{nt} = E_t[M_{t+1}...M_{t+n}]$, which specifies bond prices to be simply functions of the future discount factors. By convention we normalize $P_{0t} = 1$ to ensure the equality of a bond's price at maturity to its par value.

C. K-Factor Affine Yields

We assume that M_{t+1} is conditionally lognormal, so that we can take logs of (1) and write it as

$$p_{nt} = E_t(m_{t+1} + p_{n-1,t+1}) + \frac{1}{2}Var_t(m_{t+1} + p_{n-1,t+1}), \tag{2}$$

where lower-case letters denote logarithms of upper-case letters. Furthermore, if we have K factors, $x_{1t}, x_{2t}, \ldots, x_{Kt}$, that forecast m_{t+1}, an affine yield model can be written as

$$-p_{nt} = A_n + B_{1n}x_{1t} + B_{2n}x_{2t} + \ldots + B_{Kn}x_{Kt}. \tag{3}$$

Since the n-period bond yield is $y_{nt} = -p_{nt}/n$, yields will also be linear in the factors. The coeficients $A_n, B_{1n}, B_{2n}, \ldots, B_{Kn}$ will depend on the stochastic processes of $x_{1t}, x_{2t}, \ldots, x_{Kt}$. Since the number of factors is usually smaller than the number of maturities on the curve, the factor structure would imply restrictions across coefficients for bond prices of different maturities. In practice, specifying $A_n, B_{1n}, B_{2n}, \ldots, B_{Kn}$ involves solving (2) based on the stochastic processes of $x_{1t}, x_{2t}, \ldots, x_{Kt}$ and verifying that (3) holds.

[6]The term "pricing kernel" is due to Sargent (1987). In consumption-based equilibrium models, M_{t+1} would represent the marginal rate of substitution between present and next-period consumption (Lucas 1978, for example).

D. Previous Results with Two Factors

In the effort to reconcile time-series data with cross-section data on interest rates, models with fewer than three factors have not fared well. Backus and Zin (1994) and CLM (1994) argue that the basic problem with one-factor models is that the yield curve's steep slope near the short end requires swift mean reversion by the factor while the curve's flat slope near the long end requires slow mean reversion. The flat slope of the volatility curve also requires slow mean reversion. Gong and Remolona (1996) demonstrate that the problem is not solved with two factors. They find that the data favor models in which one of the factors is a time-varying mean, which serves to produce the characteristic hump in the U.S. yield curve around the one-year to two-year maturities. The other factor must then revert rapidly to this mean to create a steep yield curve near the short end but revert slowly to create both a flat yield curve near the long end and a flat volatility curve.

E. Three-Factor Affine Yields: Model Specification

We now propose a three-factor model. We follow Backus and Zin (1994), CLM (1994), and Gong and Remolona (1996) by specifying the model in terms of a discrete-time pricing kernel. To maintain tractability, we write the model to satisfy Duffie and Kan's (1993) conditions for affine yields. In this model, three unobserved factors drive the pricing kernel: one factor reverts over time to a fixed mean while a second factor reverts to a time-varying mean that serves as the third factor. Two-factor time-varying mean models have been estimated by Balduzzi, Das, and Foresi (1996), Gong and Remolona (1996), and Jegadeesh and Pennacchi (1996).

Three factors drive the pricing kernel, and the factors are not directly observable. Two of the factors affect expectations of the stochastic discount factor for the next period, while the third factor affects the ultimate destination of the stochastic discount factor. Specifically, the conditional expectation of the negative of the log stochastic discount factor depends on the sum of two factors:

$$-m_{t+1} = x_{1t} + x_{2t} + w_{t+1}, \tag{4}$$

where w_{t+1} represents the unexpected change in the log stochastic discount factor and will be related to risk. The shock w_{t+1} has mean zero and a variance that will be specified to depend on the time-varing mean of the second factor. Each of these factors follows a univariate AR(1) process with heteroskedastic shocks described by a square-root process:

$$\begin{aligned} x_{1,t+1} &= (1-\phi_1)\theta + \phi_1 x_{1,t} + \mu_t^{0.5} u_{1,t+1} \\ x_{2,t+1} &= (1-\phi_2)\mu_t + \phi_2 x_{2,t} + \mu_t^{0.5} u_{2,t+1} \\ \mu_{t+1} &= (1-\phi_3)\mu + \phi_3 \mu_t + \mu_t^{0.5} u_{3,t+1}, \end{aligned} \tag{5}$$

where $1-\phi_1$, $1-\phi_2$, and $1-\phi_3$ are the rates of mean reversion, θ and μ are the long-run means to which the factors revert, and $u_{1,t+1}$, $u_{2,t+1}$, and $u_{3,t+1}$ are shocks with mean zero, volatilities σ_1, σ_2 and σ_3 and covariances $\sigma_{12}, \sigma_{13}, \sigma_{23}$. The shocks have a common source of heteroskedasticity in that they are all proportional to $\mu_t^{0.5}$. This common square-root process permits affine yields while allowing correlations among the shocks.

As in Gong and Remolona (1996), we specify the shock to m_{t+1} to be proportional to the shock to μ_{t+1}, which in turn depends on the level of μ_t:

$$w_{t+1} = \lambda \mu_t^{0.5} u_{3,t+1}, \tag{6}$$

where λ represents the market price of risk. When λ is negative, bond returns are inversely correlated with the stochastic discount factor and risk premia are positive. The model can thus be characterized as a three-factor single-risk model.[7]

We now verify that yields are affine in the factors, that is, we can write

$$y_{n,t} = \frac{1}{n}(A_n + B_{1n}x_{1t} + B_{2n}x_{2t} + B_{3n}\mu_t). \tag{7}$$

The normalization $p_{0t} = 0$ gives us coefficients of $A_0 = B_{10} = B_{20} = B_{30} = 0$. We can then derive the one-period yield or short rate as

$$\begin{aligned} y_{1,t} &= -p_{1,t} = -E_t(m_{t+1}) - \frac{1}{2}Var_t(m_{t+1}) \\ &= x_{1,t} + x_{2,t} - \frac{1}{2}\lambda^2\sigma_3^2\mu_t, \end{aligned} \tag{8}$$

which is also linear in the factors, with the coefficients $A_1 = 0$, $B_{11} = B_{21} = 1$, and $B_{31} = -\frac{1}{2}\lambda^2\sigma_3^2$.

In Appendix A we verify that the yield of an n-period bond is linear in the factors with the coefficients restricted by

$$\begin{aligned} A_n &= A_{n-1} + (1-\phi_1)\theta B_{1,n-1} + (1-\phi_3)\mu B_{3,n-1} \\ B_{1,n} &= 1 + \phi_1 B_{1,n-1} \\ B_{2,n} &= 1 + \phi_2 B_{2,n-1} \\ B_{3,n} &= \phi_3 B_{3,n-1} + (1-\phi_2)B_{2,n-1} \\ &- \frac{1}{2}[(\lambda + B_{3,n-1})^2\sigma_3^2 + B_{1,n-1}^2\sigma_1^2 + B_{2,n-1}^2\sigma_2^2 \\ &+ 2(\lambda + B_{3,n-1})B_{1,n-1}\sigma_{13} + 2(\lambda + B_{3,n-1})B_{2,n-1}\sigma_{23} + 2B_{1,n-1}B_{2,n-1}\sigma_{12}], \end{aligned} \tag{9}$$

The coefficients B_{1n}, B_{2n}, and B_{3n} are factor loadings for x_{1t}, x_{2t}, and μ_t. The coefficient A_n represents the pull of the factors to the means θ and μ.

[7]With discrete-time models, accommodating more than one risk would require zero correlation between shocks that are sources of risk.

These recursive equations impose cross-sectional restrictions to be satisfied by twelve parameters: ϕ_1, ϕ_2, ϕ_3, θ, μ, σ_1, σ_2, σ_3, σ_{12}, σ_{13}, σ_{23}, and λ.

The volatility curve is derived from the conditional variance of the n-period yield:

$$\begin{aligned} Var_t(y_{n,t+1}) &= \frac{1}{n^2}(B_{1n}^2\sigma_1^2 + B_{2n}^2\sigma_2^2 + B_{3n}^2\sigma_3^2 \\ &+ 2B_{1n}B_{2n}\sigma_{12} + 2B_{1n}B_{3n}\sigma_{13} + 2B_{2n}B_{3n}\sigma_{23})\mu_t. \end{aligned} \tag{10}$$

We would have a downward-sloping volatility curve given that ϕ_1, ϕ_2, and ϕ_3 are less than unity. Mean reversion by the factors serves to dampen yield volatilities as maturity is lengthened.

A linear transformation of this model will give us a reduced form which expresses the yield for any given maturity as a linear function of any three other yields, or of one other yield, the conditional variance of any yield, and the conditional expectation of any yield. This follows from the fact that yields, conditional variances, and conditional expectations are different linear functions of the factors. Such a reduced form will therefore nest the one-factor models of Vasicek (1977) and CIR (1985) and the two-factor models of Brennan and Schwartz (1982), Schaefer and Schwartz (1984), and Longstaff and Schwartz (1992).

The model also allows us to measure term premia. We can derive term premia in the form of the expected excess bond return:

$$\begin{aligned} E_t p_{n-1,t+1} - p_{nt} &- y_{1t} = -\lambda(B_{1,n-1}\sigma_{13} + B_{2,n-1}\sigma_{23} + B_{3,n-1}\sigma_3^2)\mu_t \\ &- \frac{1}{2}[B_{1,n-1}^2\sigma_1^2 + B_{2,n-1}^2\sigma_2^2 + B_{3,n-1}^2\sigma_3^2 \\ &+ 2B_{1,n-1}B_{2,n-1}\sigma_{12} + 2B_{1,n-1}B_{3,n-1}\sigma_{13} + 2B_{2,n-1}B_{3,n-1}\sigma_{23}]\mu_t \\ &= -(\lambda\Pi + \Gamma)\mu_t, \end{aligned} \tag{11}$$

where

$$\Pi \equiv (B_{1,n-1}\sigma_{13} + B_{2,n-1}\sigma_{23} + B_{3,n-1}\sigma_3^2) \tag{12}$$

$$\Gamma \equiv \frac{1}{2}(B_{1,n-1}^2\sigma_1^2 + B_{2,n-1}^2\sigma_2^2 + B_{3,n-1}^2\sigma_3^2 + 2B_{1,n-1}B_{2,n-1}\sigma_{12} + 2B_{1,n-1}B_{3,n-1}\sigma_{13} + 2B_{2,n-1}B_{3,n-1}\sigma_{23})\mu_t. \tag{13}$$

The term Π represents a risk premium that depends on the covariance between the stochastic discount factor and bond returns, while the second term Γ represents Jensen's inequality arising from the use of logarithms. Positive term premia require that λ be so negative that

$$-\lambda > \frac{\Gamma}{\Pi}. \tag{14}$$

Note also that if $\sigma_3 = 0$, we will have homoskedastic shocks, term premia will be constant, and the pure expectations hypothesis will hold.

III. Estimating the Model

A. An Econometric Approach

Because the factors driving the dynamic movements of the stochastic discounting factor are not directly observable, the model lends itself to estimation by a Kalman filter. This maximum likelihood procedure exploits the conditional density of observed yields to extract conditional forecasts of the unobserved factors. In this application, we impose the model's arbitrage conditions as overidentifying restrictions.

Other maximum likelihood procedures also allow the use of conditional moments to estimate term structure models, but these procedures require special assumptions about measurement errors. Chen and Scott (1993) estimate a one-factor model, a two-factor model, and a three-factor model by maximizing a likelihood based on the factors' conditional moments. To derive the factors, however, their procedure requires the arbitrary assumption of zero measurement error for as many yields as they have factors. Pearson and Sun (1994) also exploit the factors' conditional density in estimating a two-factor CIR model. They derive the factors by using two yields at a time and assuming no measurement error for either yield.

The Kalman filter provides a way to exploit the conditional moments of the yields while allowing measurement errors for all the yields used in the estimation. In an early application, Fama and Gibbons (1982) used a Kalman filter to extract estimates of expected real returns from ex post inflation and three-month Treasury-bill rates. More recently, Jegadeesh and Pennacchi (1996) and Gong and Remolona (1996) applied the procedure to two-factor term structure models. In the present application, the yields as affine functions of the factors serve as the measurement equations of the Kalman filter and the factors' stochastic processes as the transition equations. The model's arbitrage conditions, however, imply strong restrictions between the measurement and transition equations, and we take careful account of these restrictions. While the procedure has the advantage of letting the data determine the measurement errors, these errors prevent us from solving the measurement equations to derive the factors directly. Hence, the likelihood function is based on the conditional density of the yields rather than the factors. The data consist of monthly U.S. Treasury zero-coupon yields from January 1984 to March 1995. We use six maturities: three months, six months, one year, two years, five years, and ten years.

In spirit, our work is close to Backus and Zin (1994) in that we use the observed yields to determine the dynamics of the underlying stochastic discount factor. Backus and Zin estimate a reduced form in the sense that they study various ARMA processes for the stochastic discounting factor. We estimate a structural model by specifying the underlying factors that drive the pricing kernel. An $ARMA(3,2)$ yield process is generated by our three $AR(1)$ factors.[8]

B. Data and Summary Statistics

We obtain end-of-month zero-coupon yield data from McCulloch and Kwon (1993) for 1984:1 to 1990:12 and from the Federal Reserve Bank of New York for 1991:1 to 1995:3. In the case of the Federal Reserve data, each zero curve is generated by fitting a cubic spline to prices and maturities of about 160 outstanding coupon-bearing U.S. Treasury securities. The securities are limited to off-the-run Treasuries to eliminate the most liquid securities and reduce the possible effect of liquidity premia. Fisher, Nychka, and Zervos (1995) explain the procedure in detail. Summary statistics for the yields with maturities of three months, six months, one year, two years, five years, and ten years for the sample period 84:1-95:3 are reported in Table 1. The average yield curve is upward sloping, with mean yields rising from 6.25 percent to 8.47 percent. Its slope is steep near the short end and flat near the long end. This curve is somewhat hump-shaped, with the hump located near the two year maturity. The average volatility curve is downward sloping with a relatively flat slope. The average volatility as measured by sample standard deviation is 2.01 percent at the short end and 1.67 percent at the long end. The yields across the curve are all very persistent, with first-order monthly auto-correlations of 0.98.

[8]Engel (1984) derives the sums, products, and time aggregations of ARMA processes.

C. Kalman Filtering and Maximum Likelihood Estimation

We write the model in the linear state-space form, with the measurement equation

$$\begin{bmatrix} y_{l,t} \\ \vdots \\ y_{n,t} \end{bmatrix} = \begin{bmatrix} a_l \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_{1,l} & \beta_{2,l} & b_{3,l} \\ \vdots & \vdots & \vdots \\ b_{1,n} & b_{2,n} & b_{3,n} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \mu_t \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ \vdots \\ v_{n,t} \end{bmatrix}, \tag{15}$$

where $y_{l,t}, \ldots, y_{n,t}$ are zero-coupon yields at time t with maturities $l, \ldots, n$ and the measurement errors $v_t = \{v_{1t}, \ldots, v_{nt}\}'$ are assumed to be i.i.d.:

$$v_t \sim N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} e_1^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & e_n^2 \end{pmatrix} \right), \tag{16}$$

and $a_k = A_k/k, b_{1,k} = B_{1,k}/k, b_{2,k} = B_{2,k}/k,\ b_{3,k} = B3, k/k, k = l, \ldots, n.$

Allowing such measurement errors is an important difference between our Kalman filter procedure and other maximum likelihood procedures based on conditional moments.

The transition equation is

$$\begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \\ \mu_{t+1} \end{bmatrix} = \begin{bmatrix} (1-\phi_1)\theta \\ 0 \\ (1-\phi_3)\mu \end{bmatrix} + \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 1-\phi_2 \\ 0 & 0 & \phi_3 \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \mu_t \end{bmatrix} + \mu_t^{0.5} \begin{bmatrix} u_{1,t+1} \\ u_{2,t+1} \\ u_{3,t+1} \end{bmatrix}$$

with shocks to the state variables distributed as

$$\begin{bmatrix} u_{1,t+1} \\ u_{2,t+1} \\ u_{3,t+1} \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23}^2 & \sigma_3^3 \end{pmatrix} \right). \tag{17}$$

In standard linear state-space models, no restrictions link the measurement equation and the transition equation. This time, however, the measurement equation comes from the transition equation and the no-arbitrage conditions, and the restrictions are given by equation (15).

After putting the restrictions into the measurement equations, the preceding model is estimated by maximum likelihood using the Kalman filter. The algorithm is discussed in Appendix C. For more detailed discussions of the Kalman filtering procedure, see, for example, Hamilton (1994).

IV. Estimates and Factors

The estimated parameters describe three factors that differ in their rates of mean reversion and volatility and combine to reproduce the average shape of the U.S. term structure. Table 3 reports parameter estimates and measurement errors based on three sample periods: the full sample from January 1984 to March 1995 and the subsamples January 1984 to December 1990 and January 1991 to March 1995. The different samples produce reasonably similar parameter estimates, suggesting a degree of stability in the pricing kernel. These estimates characterize an incredibly stable and slow moving first factor and a rather volatile second factor that reverts rapidly to a less volatile third factor. We also find sizeable measurement errors for each of the six yields.

A. Parameters

The mean reversion rates are sharply different for the three factors. These rates also seem reasonable compared with other estimates. Recall that the rates of mean reversion are given by $1 - \phi_1$ for the first factor, $1 - \phi_2$ for the second factor, and $1 - \phi_3$ for the third factor. The estimates for these parameters are quite precise and all are statistically significant. In the full-sample estimates, the first factor reverts to its fixed mean at the rate of 0.2 percent a month. Such a slow rate implies a mean half life of 27.5 years. Even this rate, however, is not so slow compared with other estimates. Chen and Scott (1993) report a half life of 771 years for one factor in a two-factor model. Our second factor reverts at the rate of 8.8 percent a month, implying a relatively short mean half life of 7.5 months. Pearson and Sun (1994) report a half life of roughly a month for one of their factors in one of their estimates. Our third factor reverts at the rate of 5.3 percent a month, implying a mean half life of 12.8 months, which is longer than that of the second factor but shorter than that of the first.

For our estimates, the slower the mean reversion, the less volatile the factor. The first factor's volatility is only 0.1 percent, the second factor's volatility 26.9 percent, the third factor's volatility 18.6 percent. It makes sense for a factor representing a time-varying mean to be less volatile than the factor that reverts to it. At the same time, it is interesting that even this time-varying mean is much more volatile than the first factor. The first factor has little correlation with other factors, while the second and third factors are positively correlated. In the model, the time-varying mean is also the source of the time variation in the risk premium. The price of risk has the expected negative sign and is statistically significant. This price of risk and the volatility of the third factor show that there is significant time variation in the term premium.

B. The Implied Factors

The Kalman filter procedure allows us to back out the three factors for each month by conditioning on the time series of six yields up to that month. Figure

1 plots these implied factors. The first factor is depicted as an almost straight line with a slowly declining trend, reflecting the factor's small volatility and slow mean reversion. The second and third factors exhibit wide swings, with the second factor showing wider swings than the third factor. The reversion of the second factor to the third factor is evident. We suspect that the first factor may represent expectations of inflation, a process that has been characterized as a long-memory process (Backus and Zin 1993). We suspect that the third factor represents the expected real return over the medium term consistent with the outlook for the business cycle, while the second factor reflects the currently expected real return.

C. Shaping the Term Structure

How well do the three factors reproduce the actual shapes of the average term structure? The term structure is represented by both the yield curve and the volatility curve. The average U.S. yield curve can be characterized as having a steep slope near the short end, a flat slope near the long end, and something like a hump around the one-year to two-year maturities. The average U.S. volatility curve can be characterized as downward sloping but with a rather flat slope. Figure 2 shows the average yield curve for our full sample and the yield curve based on the unconditional first moments implied by the model. The chart shows that the factors capture the basic shape of the yield curve, including the hump around the two-year maturity. Figure 3 shows the average volatility curve and the volatility curve based on the unconditional second moments implied by the model. Again the chart shows that the factors capture the basic shape of the volatility curve, particularly its relatively flat slope.

The slopes of the yield curve near either end are the easiest characteristics of the term structure to explain. The first factor determines the slope near the long end, because it is the factor with the slowest mean reversion. In fact, a rather slow mean reversion rate is required to keep the curve from declining. It would have the most influence on the long end because a shock to the factor would persist the longest and would be the shock most likely to be reflected in long-term yields. The second factor determines the slope near the short end, because it is the factor with the most rapid mean reversion, and it would have its greatest effect on the shorter term yields. These explanations are consistent with the factor loadings shown in Figure 4. The picture portrays the first factor as having an effect on yields that decays very slowly as maturity is lengthened and the second factor as having an effect that dies down so swiftly that its effect at the ten-year maturity is only an eighth of that of the first factor.

The hump in the yield curve is a feature associated with the time-varying mean and the effect it has on the risk premium. Such an association is evident in our time-varying mean models with two factors (Gong and Remolona 1996). In these models, the time-varying mean factor induces hetoreskedasticity in volatility, and this source of risk is priced. In the present three-factor model, the loading for this factor, as shown in Figure 4, is negative, rising rapidly near the short-end. It then becomes flat after the three-year maturity.

To produce the right curvature, however, requires the right mean reversion rate as well as the right volatility and price of risk. Figure 5, for example, shows the implied yield curve produced with the third factor having a mean reversion rate of 10 percent instead of 5 percent. The curve overshoots the average yields between three-month to six-month yields and one-year to five-year maturities.

In general, a downward sloping volatility curve requires mean reversion and a flat curve slow mean reversion. To produce the right shape for this curve, however, requires slow mean reversion for the first factor. Figure 6, for example, shows that the volatility curve drops too fast once the mean reversion rate of the first factor is changed to 5 percent.

V. Testing Three Factors Against Two

We now test whether adding a third factor represents a significant improvement over a two-factor model. Chen and Scott (1993) find that Vuong's likelihood ratio test favors their non-nested two-factor model over their three-factor model. Using quarterly data and two yields at a time, Gong and Remolona (1996) fail to find a two-factor model that is adequate to explain the whole yield curve. Their preferred two-factor model is a time-varying mean model that is nested by our three-factor model. Hence, we can compare the two models by relying on the standard likelihood ratio test.

In our three-factor model, setting $\phi_1 = 0$, $\theta = 0$, and $\sigma_1 = 0$, then $x_{1,t} = 0$ reduces the model to the two-factor time-varying mean model (Gong and Remolona 1996), which has the following pricing kernel:

$$\begin{aligned} -m_{t+1} &= x_{2t} + w_{t+1}, \\ x_{2,t+1} &= (1-\phi_2)\mu_t + \phi_2 x_{2,t} + \mu_t^{0.5} u_{2,t+1} \\ \mu_{t+1} &= (1-\phi_3)\mu + \phi_3 \mu_t + \mu_t^{0.5} u_{3,t+1}, \\ w_{t+1} &= \lambda \mu_t^{0.5} u_{3,t+1}. \end{aligned} \tag{18}$$

The yields are affine in the factors so that we can write

$$y_{n,t} = \frac{1}{n}(A_n + B_{2n} x_{2t} + B_{3n}\mu_t), \tag{19}$$

where the coefficients are restricted by

$$\begin{aligned} A_n &= A_{n-1} + (1-\phi_3)\mu B_{3,n-1} \\ B_{2,n} &= 1 + \phi_2 B_{2,n-1} \\ B_{3,n} &= \phi_3 B_{3,n-1} + (1-\phi_2) B_{2,n-1} \\ &- \frac{1}{2}[(\lambda + B_{3,n-1})^2\sigma_3^2 + B_{2,n-1}^2\sigma_2^2 + 2(\lambda + B_{3,n-1})B_{2,n-1}\sigma_{23}]. \end{aligned} \tag{20}$$

We use the same monthly sample and the same six maturities used in the estimation of the three-factor model to estimate the two-factor model. The estimates of the two-factor model and those of the three-factor model are shown in Table 4. For the two-factor model, one factor reverts to a time-varying mean rather quickly, at a rate of 10 percent a month, for a mean half life of 6.6 months. The time-varying mean is more persistent, reverting to a fixed mean at a rate of 2.5 percent a month for a mean half life of 27.4 months. All the model parameter estimates are significant. Compared with the three-factor model, however, the estimated standard deviations of the measurement errors tend to exceed those of the three-factor model, particularly near the long end.

The likelihood ratio statistic of the three-factor model against the nested two factor model is $2 * 135 * (2.18 - 0.817) = 368$. At the 1 percent level the critical value of this statistic is 11.3. Thus, the likelihood ratio test soundly rejects the two-factor model in favor of the three-factor model.

To visually compare the two-factor and three-factor models, we plot the implied unconditional first and second moments of the two models and the actual sample average yield curve and the volatility curve. As shown in Figures 7 and 8, the three-factor model captures the shapes of the sample average curves better than the two-factor model. In particular, the implied volatility of the two-factor model near the long end declines too sharply, because the factors revert too quickly to the mean to retain sufficient volatility near the long end.

VI. Testing the Arbitrage Restrictions

The real test of the three-factor model is whether the arbitrage restrictions it implies are consistent with yield movements across the term structure. The arbitrage restrictions take the form of cross-section restrictions imposed on the coefficients of the measurement equations. To test the model, we estimate one without these arbitrage restrictions, where we let all the coefficients of the measurement equations be estimated independently. To identify the unrestricted model, we impose the parameters implied by the transition equations on the coefficients of the measurement equation for the the three-month yield only. Thus, in the unrestricted estimation, there are no restrictions placed across maturities.

Because the three-factor model is nested by the unrestricted model, the standard likelihood ratio test applies. The likelihood functions are based on the conditional moment estimates from the Kalman filter, which provides quite a powerful test. The estimated mean log likelihood for the unrestricted model is 2.76, while the mean log likelihood for the three-factor model is 2.18. The likelihood ratio statistic is $2 * 135 * (2.76 - 2.18) = 157$, which is a χ^2 with 20 degrees of freedom. The critical value at the 1 percent level is 45.3. Thus the no-arbitrage restrictions imposed by the three-factor model are soundly rejected.

To understand the model's failure, we examine the measurement errors. Table 5 reports parameter estimates and the standard deviation of the measurement errors for the six maturities used in the restricted and unrestricted model estimation. For the restricted estimates, the measurement errors are relatively large for long-term yields; for the unrestricted estimates the measurement errors are relatively small, except for the three-month maturity. Note that our estimation procedure assumes i.i.d. errors. Table 6 reports cross-sectional correlations of these measurement errors. We see that measurement errors near the long end are highly correlated. These correlations suggest that there are significant common movements in the long maturity yields that are not captured by the three factors of the model. The smoothing imposed by the cubic spline to generate the zero curve data may induce some of these correlations. Table 7 reports first-order autocorrelations of the measurement errors. We see that the autocorrelations are significant and tend to be stronger near the long end. These autocorrelations suggest that the three factors fail to capture enough of the persistence of the yields. The Q-test implies no higher-order autocorrelations, suggesting that a fourth factor may capture such persistence.

VII. Conclusion

We estimated and tested a three-factor econometric model of the U.S. term structure. It is a model in which three factors drive a pricing kernel and shocks to one factor serve as a risk priced by the market. One factor reverts over time to a fixed mean, a second factor reverts to a time-varying mean, and the time-varying mean itself is a mean-reverting factor that induces time-varying term premia. We estimate the model with monthly data for six maturities across the yield curve. The parameter estimates seem to be robust to sample periods. A likelihood ratio test favors the three-factor model over a nested two-factor model. However, another likelihood ratio test soundly rejects the arbitrage restrictions imposed by the model. An analysis of measurement errors suggests that the factors fail to capture enough of the comovement and persistence of yields.

While fitting the conditional moments of the term structure remains a challenge, the three-factor model shows promise in that it seems to capture the basic shapes of the term structure. The restricted estimates describe three factors with very different rates of mean reversion. The first factor reverts to its fixed mean at a rate of less than 1 percent a month, while the second factor reverts to a time-varying mean about forty times faster. The third factor serving as the time-varying mean reverts to its own mean about twenty times faster than the first factor. Something seems key about these parameter values, because small deviations from our range of estimates produce very bad yield and volatility curves. Each factor has a role in the shape of the term structure. The first factor explains the yield curve's flat slope near the long end and the volatility curve's flat slope for most of its length. The second factor explains the yield curve's steep slope near the short end. The time-varying mean produces the yield curve's hump.

Appendix A

A1. Model I: Recursive Restrictions

We start with the general pricing equation:

$$p_{nt} = E_t(m_{t+1} + p_{n-1,t+1}) + \frac{1}{2}Var_t(m_{t+1} + p_{n-1,t+1}). \tag{21}$$

The short rate is derived by setting $p_{0,t} = 0$:

$$\begin{aligned} y_{1t} = -p_{1t} &= -E_t(m_{t+1}) - \frac{1}{2}Var_t(m_{t+1}) \\ &= x_{1,t} + x_{2,t} - \frac{1}{2}\lambda^2\sigma_3^2\mu_t, \end{aligned}$$

showing the short rate to be linear in the factors.

Now we suppose that the price of an n-period bond is affine:

$$-p_{n,t} = A_n + B_{1,n}x_{1,t} + B_{2,n}x_{2,t} + B_{3,n}\mu_t \tag{22}$$

We verify that there exist A_n, $B_{1,n}$, $B_{2,n}$, and $B_{3,n}$ that satisfy the general pricing equation:

$$\begin{aligned} E_t(m_{t+1} + p_{n-1,t+1}) &= -A_{n-1} - (1-\phi_1)\theta B_{1,n-1} - (1-\phi_3)\mu B_{3,n-1} \\ &- (1+\phi_1 B_{1,n-1})x_{1,t} - (1+\phi_2 B_{2,n-1})x_{2,t} \\ &- [(1-\phi_2)B_{2,n-1} + \phi_3 B_{3,n-1}]\mu_t \qquad (23) \\ Var_t(m_{t+1} + p_{n-1,t+1}) &= [(\lambda + B_{3,n-1})^2\sigma_3^2 + B_{1,n-1}^2\sigma_1^2 + B_{2,n-1}^2\sigma_2^2 \\ &+ 2(\lambda + B_{3,n-1})B_{1,n-1}\sigma_{13} + 2B_{1,n-1}B_{2,n-1}\sigma_{12} \\ &+ 2(\lambda + B_{3,n-1})B_{2,n-1}\sigma_{23}]\mu_t. \qquad (24) \end{aligned}$$

Now substitute (26) (27) into (24) and match the coefficients of equations (24) and (25), we have

$$A_n = A_{n-1} + (1-\phi_1)\theta B_{1,n-1} + (1-\phi_3)\mu B_{3,n-1}$$

$$
\begin{aligned}
B_{1,n} &= 1 + \phi_1 B_{1,n-1} \\
B_{2,n} &= 1 + \phi_2 B_{2,n-1} \\
B_{3,n} &= \phi_3 B_{3,n-1} + (1 - \phi_2) B_{2,n-1} \\
&- \frac{1}{2}[(\lambda + B_{3,n-1})^2 \sigma_3^2 + B_{1,n-1}^2 \sigma_1^2 + B_{2,n-1}^2 \sigma_2^2 \\
&+ 2(\lambda + B_{3,n-1}) B_{1,n-1} \sigma_{13} + 2(\lambda + B_{3,n-1}) B_{2,n-1} \sigma_{23} + 2 B_{1,n-1} B_{2,n-1} \sigma_{12}].
\end{aligned}
$$

Appendix B

B1. The Term Premia

Term premia can be derived from the expected excess bond return over the short rate:

$$
\begin{aligned}
E_t & \quad p_{n-1,t+1} - p_{n,t} - y_{1t} \\
= & \quad -A_{n-1} - B_{1,n-1}E_t x_{1,t+1} - B_{2,n-1}E_t x2, t+1 - B_{3,n-1}E_t\mu_{t+1} \\
+ & \quad A_n + B_{1,n}x_{1,t} + B_{2,n}x_{2,t} + B_{3,n}\mu_t \\
- & \quad x_{1,t} - x_{2,t} + \frac{1}{2}\lambda^2\sigma_3^2\mu_t \\
= & \quad (A_n - A_{n-1} - B_{1,n-1}(1-\phi_1)\theta - B_{3,n-1}\mu(1-\phi_3)\mu) \\
+ & \quad (B_{1,n} - 1 - \phi_1 B_{1,n-1})x_{1t} + (B_{2,n} - 1 - \phi_2 B_{2,n-1})x_{2t} \\
+ & \quad (B_{3n} + \frac{1}{2}\lambda^2\sigma_3^2 - \phi_3 B_{3,n-1} - (1-\phi_2)B_{2,n-1})\mu_t \\
= & \quad -\lambda(B_{1,n-1}\sigma_{13} + B_{2,n-1}\sigma_{23} + B_{3,n-1}\sigma_3^2)\mu_t \\
- & \quad \frac{1}{2}[B_{1,n-1}^2\sigma_1^2 + B_{2,n-1}^2\sigma_2^2 + B_{3,n-1}^2\sigma_3^2 \qquad (25) \\
+ & \quad 2B_{1,n-1}B_{2,n-1}\sigma_{12} + 2B_{1,n-1}B_{3,n-1}\sigma_{13} + 2B_{2,n-1}B_{3,n-1}\sigma_{23}]\mu_t.
\end{aligned}
$$

B2. Expected Change in the Short Rate

The conditional expectation of the short rate n periods in the future is

$$
E_t y_{1,t+n} = E_t x_{1,t+n} + E_t x_{2,t+n} - \frac{1}{2}\lambda^2\sigma_3^2 E_t\mu_{t+n}
$$

$$
E_t \begin{bmatrix} x_{1,t+1} - \theta \\ x_{2,t+1} - \mu \\ \mu_{t+1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 1-\phi_2 \\ 0 & 0 & \phi_3 \end{bmatrix} \begin{bmatrix} x_{1,t} - \theta \\ x_{2,t} - \mu \\ \mu_t - \mu \end{bmatrix}
$$

$$E_t \begin{bmatrix} x_{1,t+n} - \theta \\ x_{2,t+n} - \mu \\ \mu_{t+n} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 1 - \phi_2 \\ 0 & 0 & \phi_3 \end{bmatrix}^n \begin{bmatrix} x_{1,t} - \theta \\ x_{2,t} - \mu \\ \mu_t - \mu \end{bmatrix}$$

$$= \begin{bmatrix} \phi_1^n & 0 & 0 \\ 0 & \phi_2^n & \frac{1-\phi_2}{\phi_2-\phi_3}(\phi_2^{n+1} - \phi_3^{n+1}) \\ 0 & 0 & \phi_3^n \end{bmatrix} \begin{bmatrix} x_{1,t} - \theta \\ x_{2,t} - \mu \\ \mu_t - \mu \end{bmatrix}.$$

We can then write

$$E_t y_{1,t+n} - y_{1,t} = \alpha_n + \beta_{1n}(\theta - x_{1t}) + \beta_{2n}(\mu_t - x_{2t}) + \gamma_n \mu_t,$$

where

$$\begin{aligned} \alpha_n &= [1 - \phi_2^n - \frac{1-\phi_2}{\phi_2 - \phi_3}(\phi_2^{n+1} - \phi_3^{n+1}) - \frac{1}{2}\lambda^2\sigma_3^2(1 - \phi_3^n)]\mu \\ \beta_{1n} &= 1 - \phi_1^n \\ \beta_{2n} &= 1 - \phi_2^n \\ \gamma_n &= \frac{1-\phi_2}{\phi_2 - \phi_3}(\phi_2^{n+1} - \phi_3^{n+1}) + \frac{1}{2}\lambda^2\sigma_3^2(1 - \phi_3^n) - (1 - \phi_2^n). \end{aligned}$$

Appendix C: The Kalman Filter Algorithm

For the state-space models in section III, the measurement and transition equations can be written in the following matrix form.

Measurement Equation:

$$y_t = A + BX_t + v_t,$$

where $v_t \sim N(0, R)$.

Transition Equation:

$$X_{t+1} = C + FX_t + u_{t+1}, \tag{26}$$

where $u_{t+1|t} \sim N(0, Q_t)$.

The Kalman filter algorithm of this state-space model is the following:

1. Initialize the state-vector S_t:

The recursion begins with a guess $S_{1|0}$, usually given by

$$\hat{S}_{1|0} = E(S_1). \tag{27}$$

The associated MSE is

$$\begin{aligned} P_{1|0} &\equiv E[(S_1 - \hat{S}_{1|0})(S_1 - \hat{S}_{1|0})'] \\ &= Var(S_1). \end{aligned}$$

The initial state S_1 is assumed to be $N(\hat{S}_{1|0}, P_{1|0})$.

2. Forecast y_t:

Let I_t denote the information set at time t. Then

$$\begin{aligned}\hat{y}_{t|t-1} &= A + BE[S_t|I_{t-1}] \\ &= A + B\hat{S}_{t|t-1}.\end{aligned} \tag{28}$$

The forcasting MSE is

$$E[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'] = BP_{t|t-1}B' + R. \tag{29}$$

3. Update the inference about S_t given I_t:

Knowing y_t helps to update $S_{t|t-1}$ by the following:

Write

$$\begin{aligned}S_t &= \hat{S}_{t|t-1} + (S_t - \hat{S}_{t|t-1}) \\ y_t &= A + B\hat{S}_{t|t-1} + B(S_t - \hat{S}_{t|t-1}) + v_t.\end{aligned} \tag{30}$$

We have the following joint distribution:

$$\begin{bmatrix} S_t & \\ & |I_{t-1} \\ y_t & \end{bmatrix} \sim N(\begin{bmatrix} \hat{S}_{t|t-1} \\ A + B\hat{S}_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}B' \\ BP_{t|t-1} & BP_{t|t-1}B' + R \end{bmatrix}), \tag{31}$$

Thus,

$$\begin{aligned}\hat{S}_{t|t} &\equiv E[S_t|y_t, I_{t-1}] \\ &= \hat{S}_{t|t-1} + P_{t|t-1}B'(BP_{t|t-1}B' + R)^{-1}(y_t - BS_{t|t-1} - A)\end{aligned} \tag{32}$$

$$\begin{aligned}P_{t|t} &\equiv E[(S_t - \hat{S}_{t|t})(S_t - \hat{S}_{t|t})'] \\ &= P_{t|t-1} - P_{t|t-1}B'(BP_{t|t-1}B' + R)^{-1}BP_{t|t-1}.\end{aligned} \tag{33}$$

4. Forecast S_{t+1} given I_t:

$$\hat{S}_{t+1|t} = E[S_{t+1}|I_t] = F\hat{S}_{t|t}. \tag{34}$$

$$P_{t+1|t} = E[(S_{t+1} - \hat{S}_{t+1|t})(S_{t+1} - \hat{S}_{t+1|t})']$$
$$= FP_{t|t}F' + Q_t. \quad (35)$$

5. Maximum likelihood estimation of parameters:

The likelihood function can be built up recursively

$$\log L(Y_T) = \sum_{t=1}^{T} \log f(y_t|I_{t-1}), \quad (36)$$

where

$$f(y_t|I_{t-1}) = (2\pi)^{-1/2}|H'P_{t|t-1}H + R|^{-1/2}$$
$$* \; exp\{-\frac{1}{2}(y_t - A - B\hat{S}_{t|t-1})'(B'P_{t|t-1}B + R)^{-1}(y_t - A - B\hat{S}_{t|t-1})\}$$
$$for \quad t = 1, 2, \ldots, T. \quad (37)$$

Parameter estimates can then be based on the numerical maximization of the likelihood function.

References

1. Ait-Sahalia, Yacine, 1996, Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies* 9: 427-70.

2. Andersen, T.G., and J. Lund, 1996, Stochastic Volatility and Mean Drift in the Short Rate Diffusion: Sources of Steepness, Level and Curvature in the Yield Curve, unpublished paper, Northwestern University, February.

3. Backus, D., and S. Zin, 1993, Long-Memory Inflation Uncertainty: Evidence from the Term Structure, *Journal of Money, Credit and Banking* 25, 3, part 2 (August): 681-708.

4. Backus, D., and S. Zin, 1994, Reverse Engineering the Yield Curve, unpublished paper, New York University, March.

5. Backus, D., S. Foresi, and S. Zin, 1994, Arbitrage Opportunities in Arbitrage-Free Models of Bond Pricing, unpublished, New York University.

6. Balduzzi, P., S.R. Das, and S. Foresi, 1996, The Central Tendency: A Second Factor in Bond Yields, forthcoming, *Review of Economics and Statistics*.

7. Black, Fischer, E. Derman, and W. Toy, 1990, A One-Factor Model of Interest Rates and Its Application to Treasury Bond Options, *Financial Analysts Journal*, January-February: 33-9.

8. Black, F., and P. Karasinski, 1991, Bond and Option Pricing When Short Rates Are Lognormal, *Financial Analysts Journal*, July-August: 52-9.

9. Brennan, M.J., and E.S. Schwartz, 1982, An Equilibrium Model of Bond Pricing and a Test of Market Efficiency, *Journal of Financial and Quantitative Analysis* 17: 301-29.

10. Campbell, J., A. Lo, and C. MacKinlay, 1994, Models of the Term Structure of Interest Rates, Federal Reserve Bank of Philadelphia Working Paper No. 94-10 (May).

11. Chan, K.C., A. Karolyi, F. Longstaff, and A. Sanders, 1992, An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance* 47: 1209-27.

12. Chen, R.R., 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.

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

14. Duffie, D., 1992. *Dynamic Asset Pricing Theory*, Princeton University Press.

15. Duffie, D., and R. Kan, 1993, A Yield-Factor Model of Interest Rates, Working Paper, Stanford University.

16. Eom, Y.H., 1995, Essays on Empirical Tests of Continuous Time Asset Dynamnics, unpublished, New York University.

17. Engel, E.M.R.A., 1984, A Unified Approach to the Study of Sums, Products, Time-Aggregation and Other Functions of ARMA Processes, *Journal of Time Series Analysis.* 5: 159-71.

18. Fama, E.F., and M.R. Gibbons, 1982, Inflation, real returns, and capital investment, *Journal of Monetary Economics* 9: 297-324.

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

20. Gallant, R., and G. Tauchen, 1996a, Which Moments to Match? *Econometric Theory*, forthcoming.

21. Gallant, R., and G. Tauchen, 1996b, Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions, unpublished Paper, University of North Carolina.

22. Gibbons, Michael R., and Krishna Ramaswamy, 1993, A Test of the Cox, Ingersoll, and Ross Model of the Term Structure, *The Review of Financial Studies* 6: 619-58.

23. Gong, F., and E. Remolona, 1996, Two Factors Along the Yield Curve, *The Manchester School*, forthcoming.

24. Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press.

25. Heath, David, Robert Jarrow, and Andrew Morton, 1992, Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation, *Econometrica* 60: 77-105.

26. Ho, Thomas S.Y., and Sang-Bin Lee, 1986, Term Structure Movements and Pricing Interest Rate Contingent Claims, *Journal of Finance* 41: 1011-29.

27. Hull, J., and A White, 1990, Pricing Interest Rate Derivative Securities, *The Review of Financial Studies* 3: 573-92.

28. Jegadeesh, N., and G.G. Pennacchi, 1996, The Behavior of Interest Rates Implied by the Term Structure of Eurodollar Futures, *Journal of Money, Credit, and Banking* 28: 426-46.

29. Litterman, R., and J. Scheinkman, 1991, Common Factors Affecting Bond Returns, *Journal of Fixed Income*, June: 54-61.

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

31. Lucas, Robert E., 1978, Asset Prices in an Exchange Economy, *Econometrica* 46, no. 6.

32. McCulloch, J. Huston, and Heon-Chul Kwon, 1993, U.S. Term Structure Data, 1947-1991, Ohio State University Working Paper no. 93-6.

33. Pearson, N.D., 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* Vol. XLIX, no. 4: 1279-304.

34. Sargent, Thomas J., 1987. *Dynamic Macroeconomic Theory.* Cambridge, Mass.: Harvard University Press.

35. Schaefer, S.M., and E.S. Schwartz, 1984, A Two-Factor Model of the Term Structure: An Approximate Analytical Solution, *Journal of Financial and Quantitative Analysis*, Vol. 19, no. 4: 413-24.

36. Singleton, Kenneth J., 1990, Specification and Estimation of Intertemporal Asset Pricing Models, in Benjamin M. Friedman and Frank H. Hahn eds. *Handbook of Monetary Economics.* Amsterdam: North Holland.

37. Stanton, Richard, 1996, A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk, unpublished paper, University of California at Berkeley.

38. Vasicek, O.A., 1977, An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics* 5: 177-88.

Table 1
Summary Statistics for End-of-the-Month Zero-Coupon Yields with Maturities of Three Months, Six Months, One Year, Two Years, Five Years, and Ten Years

The sample period is 1984:1-1995:3. Data are obtained from McCulloch and Kwon (1993) for 1984:1 to 1990:12 and from the Federal Reserve Bank of New York for 1991:1 to 1995:3.

Maturity	Monthly Observations (84:1-95:3)		
	Mean	Std Dev	1st-Auto. Corr
3 Months	6.25	2.01	0.98
6 Months	6.45	2.04	0.98
1 Year	6.79	2.07	0.98
2 Year	7.27	2.05	0.97
5 Year	7.99	1.86	0.97
10 Year	8.47	1.67	0.97

Table 2
Correlations Between End-of-the-Month Zero-Coupon Yields with Maturities of Three Months, Six Months, One Year, Two Years, Five Years, and Ten Years

The sample period is 1984:1-1995:3. Data are obtained from McCulloch and Kwon (1993) for 1984:1 to 1990:12 and from the Federal Reserve Bank of New York for 1991:1 to 1995:3.

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
3-month	1.000					
6-month	0.996	1.000				
1-year	0.984	0.996	1.000			
2-year	0.961	0.980	0.992	1.000		
5-year	0.898	0.925	0.947	0.978	1.000	
10-year	0.830	0.863	0.892	0.937	0.987	1.000

Table 3
Parameter Estimates of the Three-Factor Model for Different Sample Periods Using 3-Month, 6-Month, 1-Year, 2-Year, 5-Year, and 10-Year Maturities

The parameters describe AR(1) processes of three factors. The first factor reverts to a fixed mean, while the second factor reverts to the third factor, which in turn reverts to a fixed mean. The parameter $1-\phi_i$ is the rate of mean reversion of the factor x_{it}, θ is the long-run mean of the first factor, and μ is the long-run mean of the third factor. The parameter σ_i is the volatility of the shock $u_{i,t+1}$ to the factor $x_{i,t+1}$, and ρ_{ij} is the correlation between factors x_{it} and x_{jt}. The parameter λ is the price of risk. The measurement errors refer to the three-month, six-month, one-year, two-year, five-year, and ten-year yields, respectively. The parameters are estimated by a Kalman filter maximum likelihood procedure imposing arbitrage conditions as overidentifying restrictions. End-of-the-month zero-coupon yields are obtained from McCulloch and Kwon (1993) for 1984:1 to 1990:12 and from the Federal Reserve Bank of New York for 1991:1 to 1995:3.

Model Parameters	84:1-95:3	84:1-90:12	91:1-95:3
ϕ_1	0.9979 (0.0009)	0.9991 (0.0001)	0.9982 (0.0002)
ϕ_2	0.9117 (0.0167)	0.9525 (0.0015)	0.9307 (0.0007)
ϕ_3	0.9473 (0.0127)	0.9825 (0.0152)	0.9557 (0.0075)
θ	0.3563 (4.7993)	0.5759 (3.6462)	0.3672 (3.7084)
μ	11.2207 (1.1476)	35.1345 (41.4595)	14.1413 (4.5474)
λ	-8.3984 (0.4649)	-7.9805 (0.1032)	-8.1067 (0.0989)
σ_1	0.0013 (0.0005)	0.0001 (0.0000)	0.0015 (0.0000)
σ_2	0.2689 (0.0373)	0.1936 (0.0022)	0.2877 (0.0003)
σ_3	0.1856 (0.0005)	0.1777 (0.0000)	0.2007 (0.0001)
ρ_{12}	0.0048 (0.6583)	0.0046 (9.0595)	0.0049 (1.7748)
ρ_{13}	-0.0366 (0.9311)	-0.0073 (5.7132)	-0.0394 (1.1316)
ρ_{23}	0.8619 (0.0188)	0.9273 (0.0122)	0.9207 (0.0015)

Table 3 Continued: Standard Deviation of Measurement Errors			
e_1	0.2496	0.2836	0.3323
e_2	0.0144	0.0071	0.5041
e_3	0.1152	0.1381	0.2169
e_4	0.2388	0.2986	0.6487
e_5	0.5535	0.6823	0.5606
e_6	0.7785	0.6820	0.3057
Mean Log-Likelihood	2.18	1.79	1.49

Table 4
Comparison of the Three-Factor Model and the Nested Two-Factor Model

The parameters describe AR(1) processes of three factors. The first factor reverts to a fixed mean, while the second factor reverts to the third factor, which in turn reverts to a fixed mean. The parameter $1-\phi_i$ is the rate of mean reversion of the factor x_{it}, θ is the long-run mean of the first factor, and μ is the long-run mean of the third factor. The parameter σ_i is the volatility of the shock $u_{i,t+1}$ to the factor $x_{i,t+1}$, and ρ_{ij} is the correlation between factors x_{it} and x_{jt}. The parameter λ is the price of risk. The measurement errors refer to the three-month, six-month, one-year, two-year, five-year, and ten-year yields, respectively. The parameters are estimated by a Kalman filter maximum likelihood procedure imposing arbitrage conditions as overidentifying restrictions. End-of-the-month zero-coupon yields are obtained from McCulloch and Kwon (1993) for 1984:1 to 1990:12 and from the Federal Reserve Bank of New York for 1991:1 to 1995:3.

Model Parameters	The Three-Factor Model	The Two-Factor Model
ϕ_1	0.9979 (0.0009)	–
ϕ_2	0.9117 (0.0167)	0.9038 (0.0183)
ϕ_3	0.9473 (0.0127)	0.9740 (0.0069)
θ	0.3563 (4.7993)	–
μ	11.2207 (1.1476)	16.9343 (4.6124)
λ	-8.3984 (0.4649)	-12.7464 (1.8370)
σ_1	0.0013 (0.0005)	–
σ_2	0.2689 (0.0373)	0.0948 (0.0363)
σ_3	0.1856 (0.0005)	0.0728 (0.0156)
ρ_{12}	0.0048 (0.6583)	–
ρ_{13}	-0.0366 (0.9311)	–
ρ_{23}	0.8619 (0.0188)	0.4038 (0.1322)

Table 4 Continued: Standard Deviation of Measurement Errors		
e_1	0.2496	0.1416
e_2	0.0144	0.1133
e_3	0.1152	0.2104
e_4	0.2388	0.5370
e_5	0.5535	0.4159
e_6	0.7785	1.0120
Mean Log-Likelihood	2.18	0.817

Table 5

Testing the Arbitrage Restrictions: Restricted and Unrestricted Estimates of the Three-Factor Model

The parameters describe AR(1) processes of three factors. The first factor reverts to a fixed mean, while the second factor reverts to the third factor, which in turn reverts to a fixed mean. The parameter $1-\phi_i$ is the rate of mean reversion of the factor x_{it}, θ is the long-run mean of the first factor, and μ is the long-run mean of the third factor. The parameter σ_i is the volatility of the shock $u_{i,t+1}$ to the factor $x_{i,t+1}$, and ρ_{ij} is the correlation between factors x_{it} and x_{jt}. The parameter λ is the price of risk. The measurement errors refer to the three-month, six-month, one-year, two-year, five-year, and ten-year yields, respectively. The parameters are estimated by a Kalman filter maximum likelihood procedure imposing arbitrage conditions as overidentifying restrictions. End-of-the-month zero-coupon yields are obtained from McCulloch and Kwon (1993) for 1984:1 to 1990:12 and from the Federal Reserve Bank of New York for 1991:1 to 1995:3.

Model Parameters	Restricted by Arbitrage Conditions	Unrestricted Model
ϕ_1	0.9979 (0.0009)	0.9706 (0.0128)
ϕ_2	0.9117 (0.0167)	0.5430 (0.7069)
ϕ_3	0.9473 (0.0127)	0.5359 (0.3588)
θ	0.3563 (4.7993)	3.7799 (1.4519)
μ	11.2207 (1.1476)	0.7784 (1.1302)
λ	-8.3984 (0.4649)	-0.0737 (0.2753)
σ_1	0.0013 (0.0005)	0.0744 (0.0153)
σ_2	0.2689 (0.0373)	0.0110 (0.0228)
σ_3	0.1856 (0.0005)	0.5910 (0.6743)
ρ_{12}	0.0048 (0.6583)	-0.0037 (12.7790)
ρ_{13}	-0.0366 (0.9311)	0.0455 (0.2752)
ρ_{23}	0.8619 (0.0188)	0.9984 (20.1624)

Table 5 Continued: Standard Deviation of Measurement Errors		
e_1	0.2496	1.2413
e_2	0.0144	0.2006
e_3	0.1152	0.0895
e_4	0.2388	0.2618
e_5	0.5535	0.2190
e_6	0.7785	0.0823
Mean Log-Likelihood	2.18	2.76

Table 6

Cross-Section Correlations between the Measurement Errors of Implied Yields

Implied yields are based on the full-sample estimates of the three-factor model with the parameters reported in Table 3.

Maturity	3 m	6 m	1 yr	2 yr	3 yr	4 yr	5 yr	7 yr	10 yr
3 m	1.00								
6 m	0.31	1.00							
1 yr	-0.68	-0.80	1.00						
2 yr	-0.48	-0.69	0.74	1.00					
3 yr	-0.38	-0.51	0.52	0.89	1.00				
4 yr	-0.18	-0.46	0.36	0.81	0.92	1.00			
5 yr	-0.04	-0.44	0.27	0.71	0.77	0.94	1.00		
7 yr	0.01	-0.41	0.21	0.60	0.62	0.84	0.97	1.000	
10 yr	0.01	-0.39	0.20	0.57	0.56	0.80	0.94	0.99	1.000

Table 7
First-Order Autocorrelation of the Measurement Error of Implied Yields
Implied yields are based on the full-sample estimates of the three-factor model with the parameters reported in Table 3.

Maturity	1st. Autocorrelation	t-Statistics	Q-statistic for higher order autocorrelation (Significance Level of Q)
3-month	0.66	7.74	28.45 (0.69)
6-month	0.44	4.38	31.18 (0.56)
1-year	0.70	10.82	26.17 (0.80)
2-year	0.72	10.18	26.14 (0.80)
3-year	0.68	8.73	29.82 (0.63)
4-year	0.75	11.79	17.83 (0.99)
5-year	0.84	15.33	20.39 (0.96)
7-year	0.90	18.88	24.45 (0.86)
10-year	0.91	21.41	23.94 (0.88)

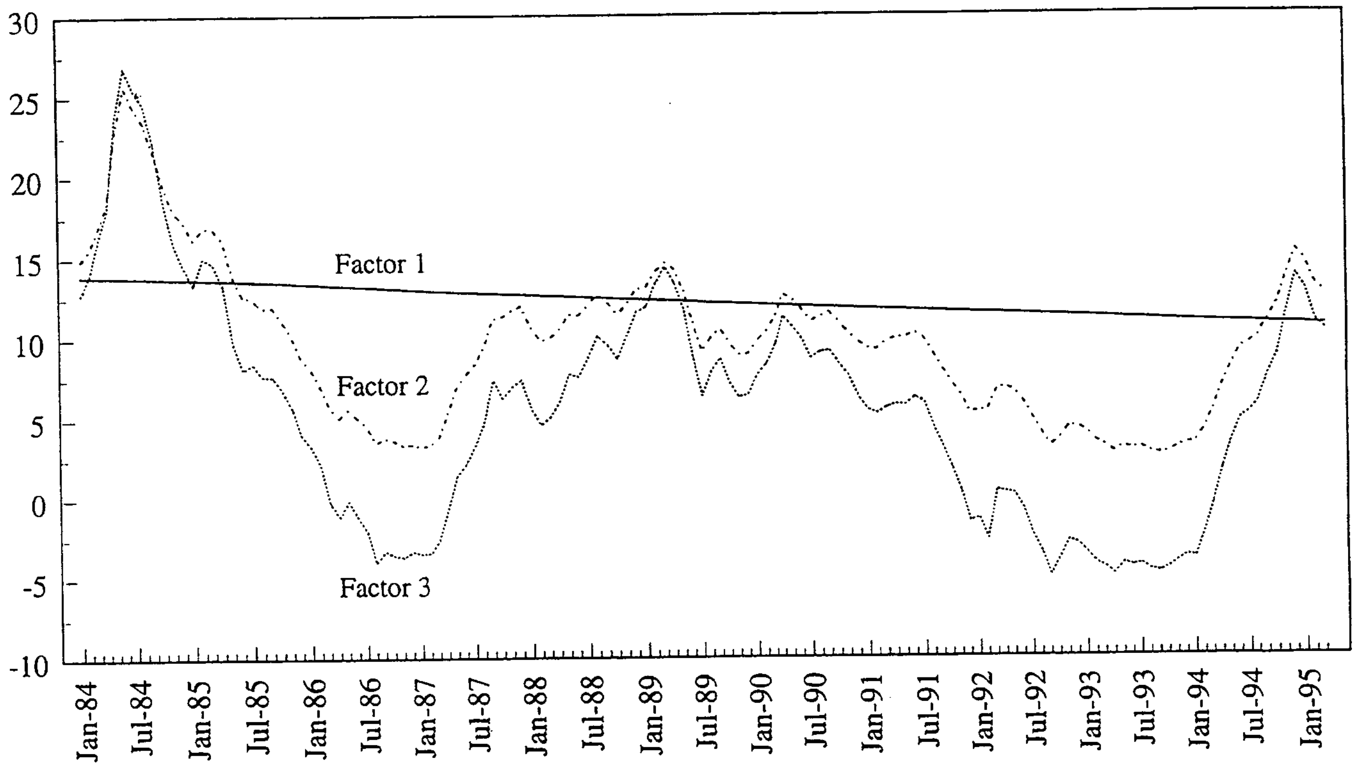

Figure 1. Time Series of the Three Implied Factors. The factors are extracted through the Kalman filter full-sample estimates of the three-factor model with the parameters reported in Table 3. The sample period is 84:1- 95:3.

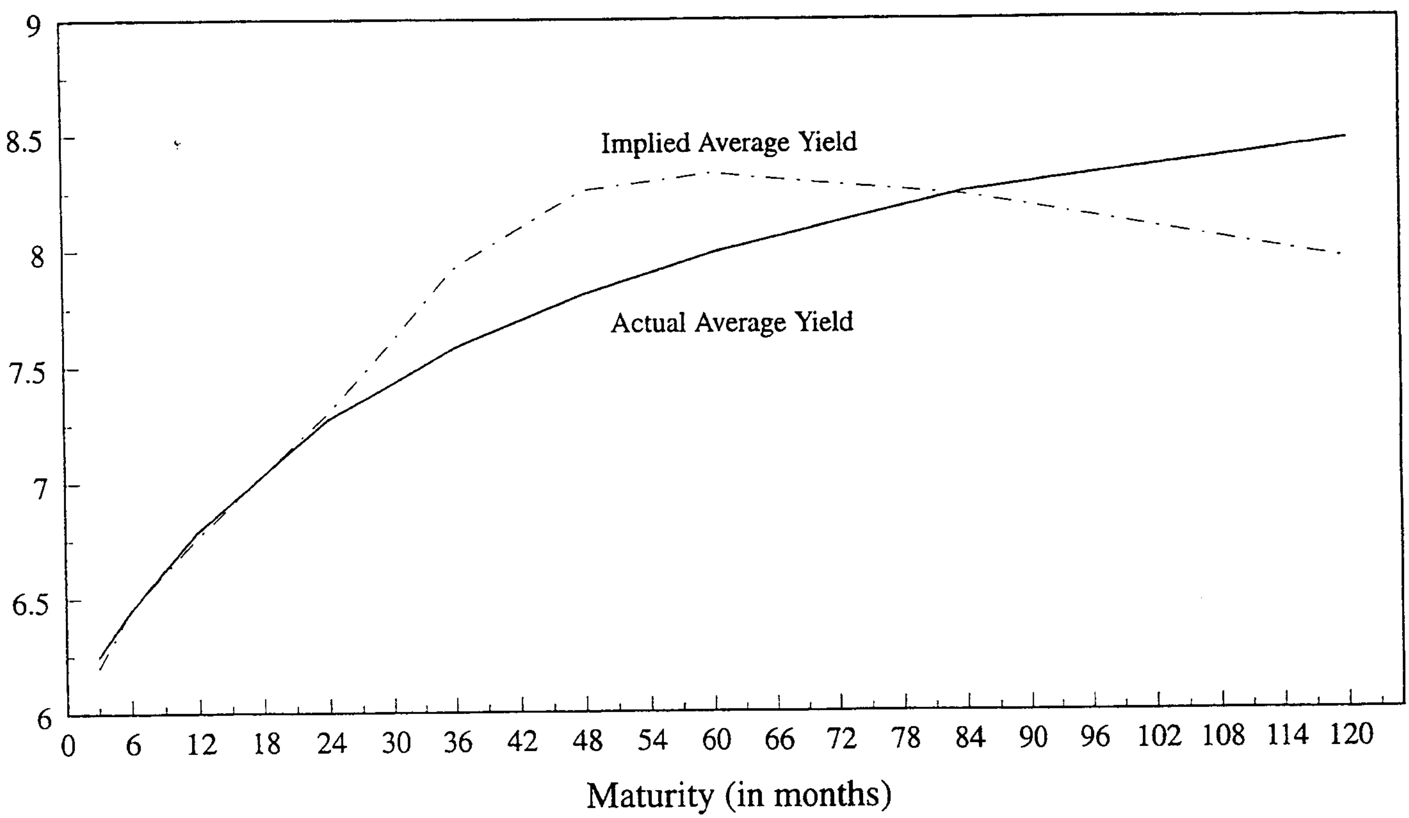

Figure 2. Actual and Implied Average Yield Curves. The implied curve is derived by setting the factors at their unconditional means. The average curve is based on the average for the full sample.

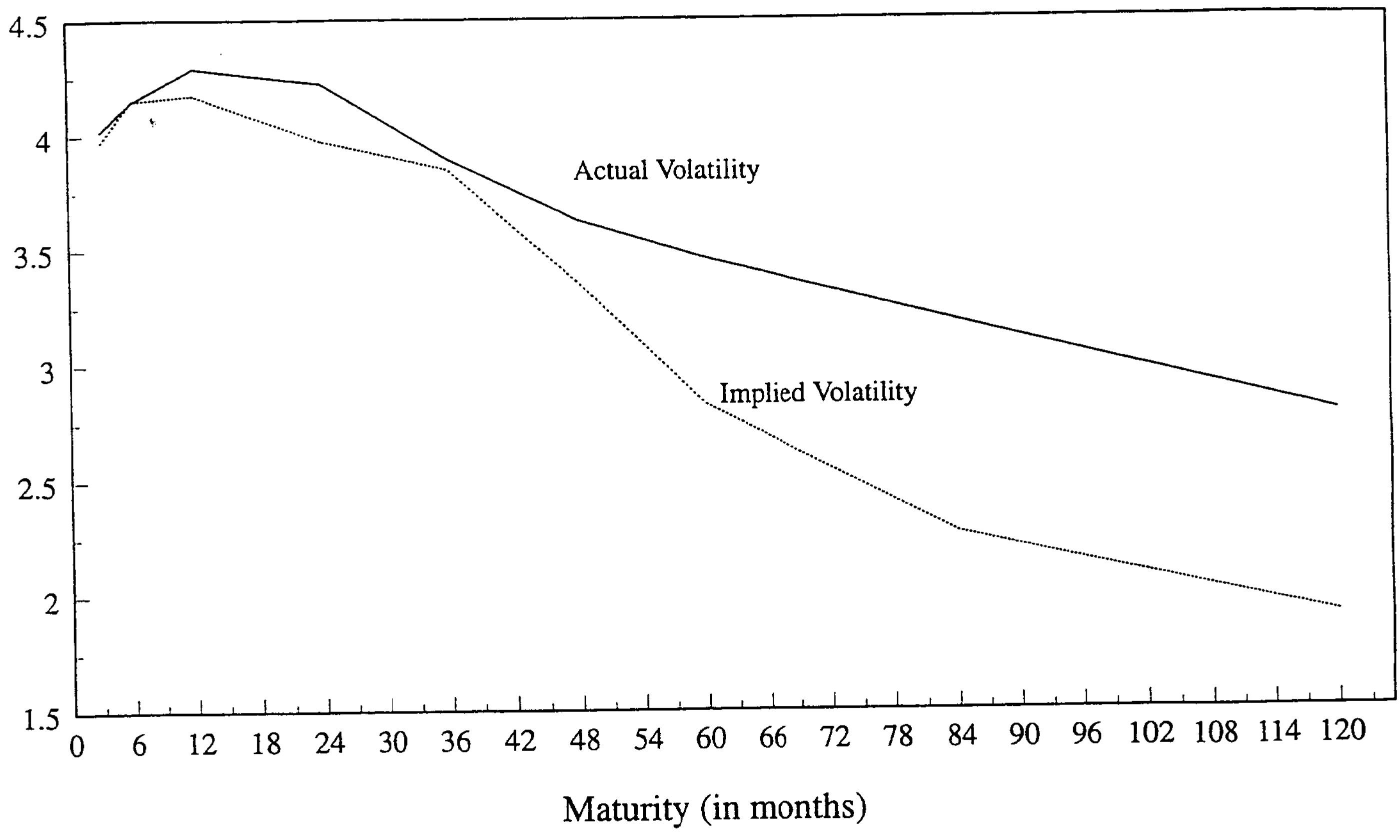

Figure 3. Actual and Implied Volatilities. The implied volatility curve is based on the unconditional variance of the implied yields derived by setting the factors at their unconditional means. The actual volatility curve is the sample variance of the observed yields.

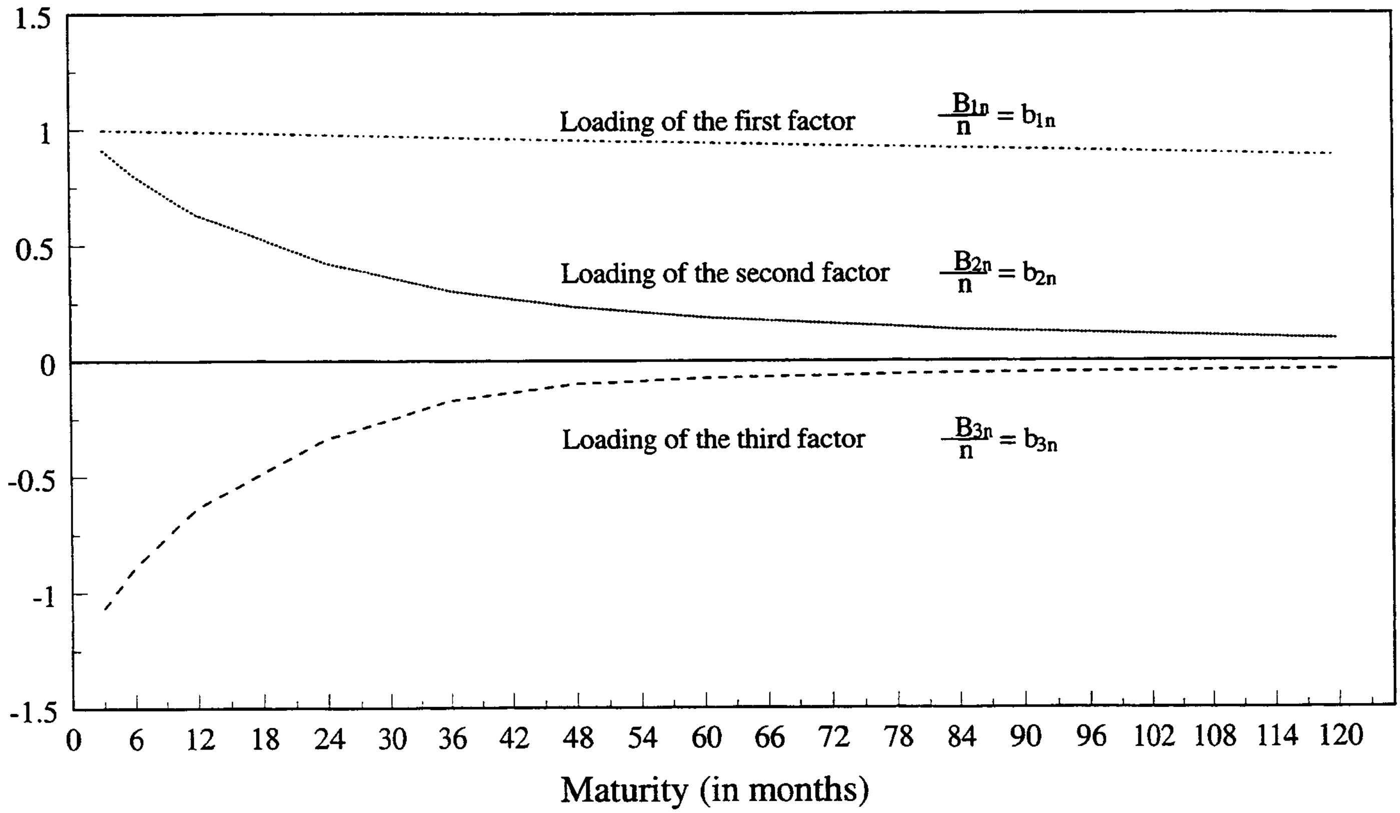

Figure 4. Factor Loadings of the Three-Factor Model. Factor loadings are calculated from equations (8) and (9) based on t full sample estimates of the three-factor model with the parameters reported in Table 3.

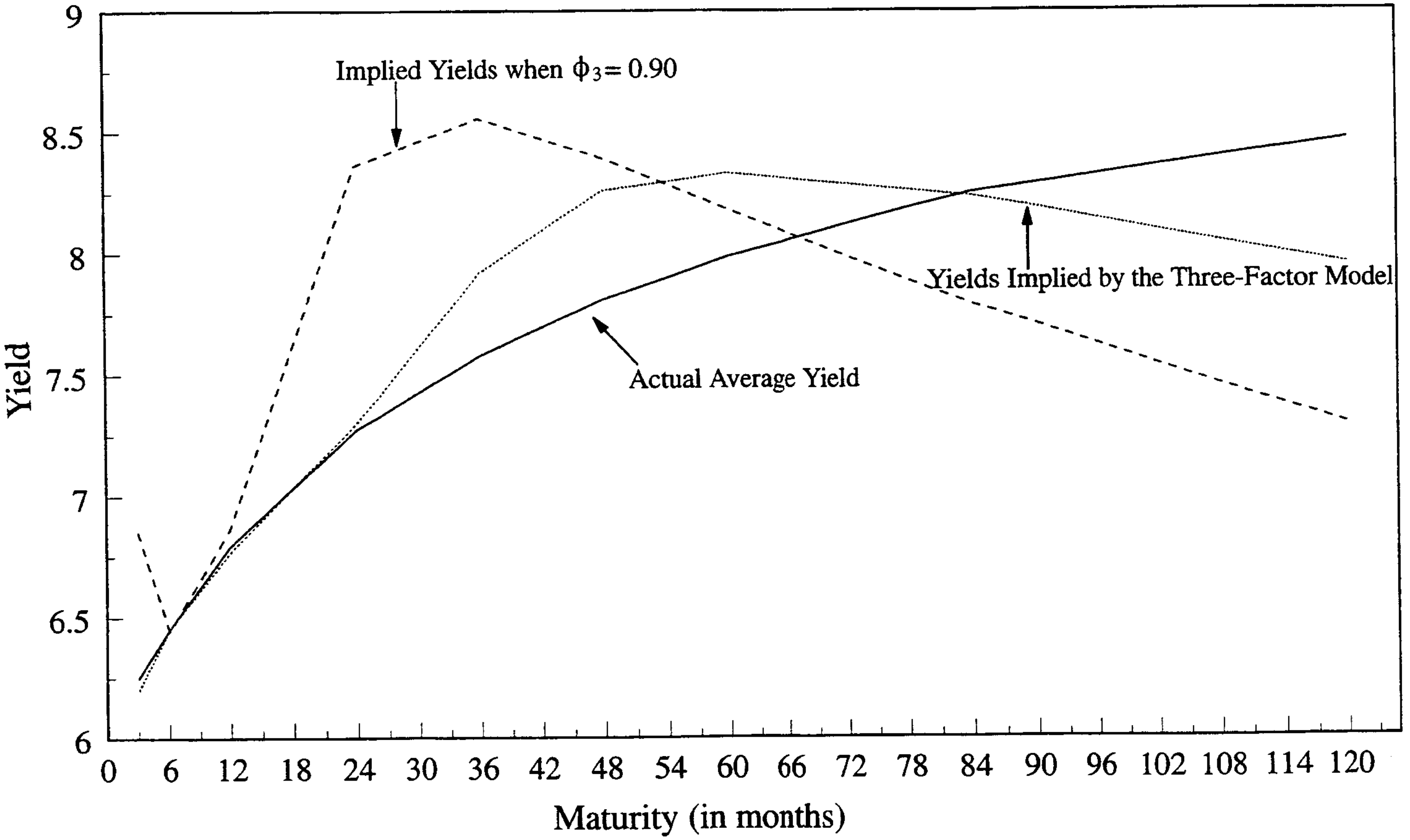

Figure 5. Actual and Implied Yield Curves. The curve implied by the three-factor model is based on the full sample estimate reported in Table 3. The implied curve when $\phi_3 = 0.90$ is calculated from the same estimates except for $\phi_3 = 0.90$.

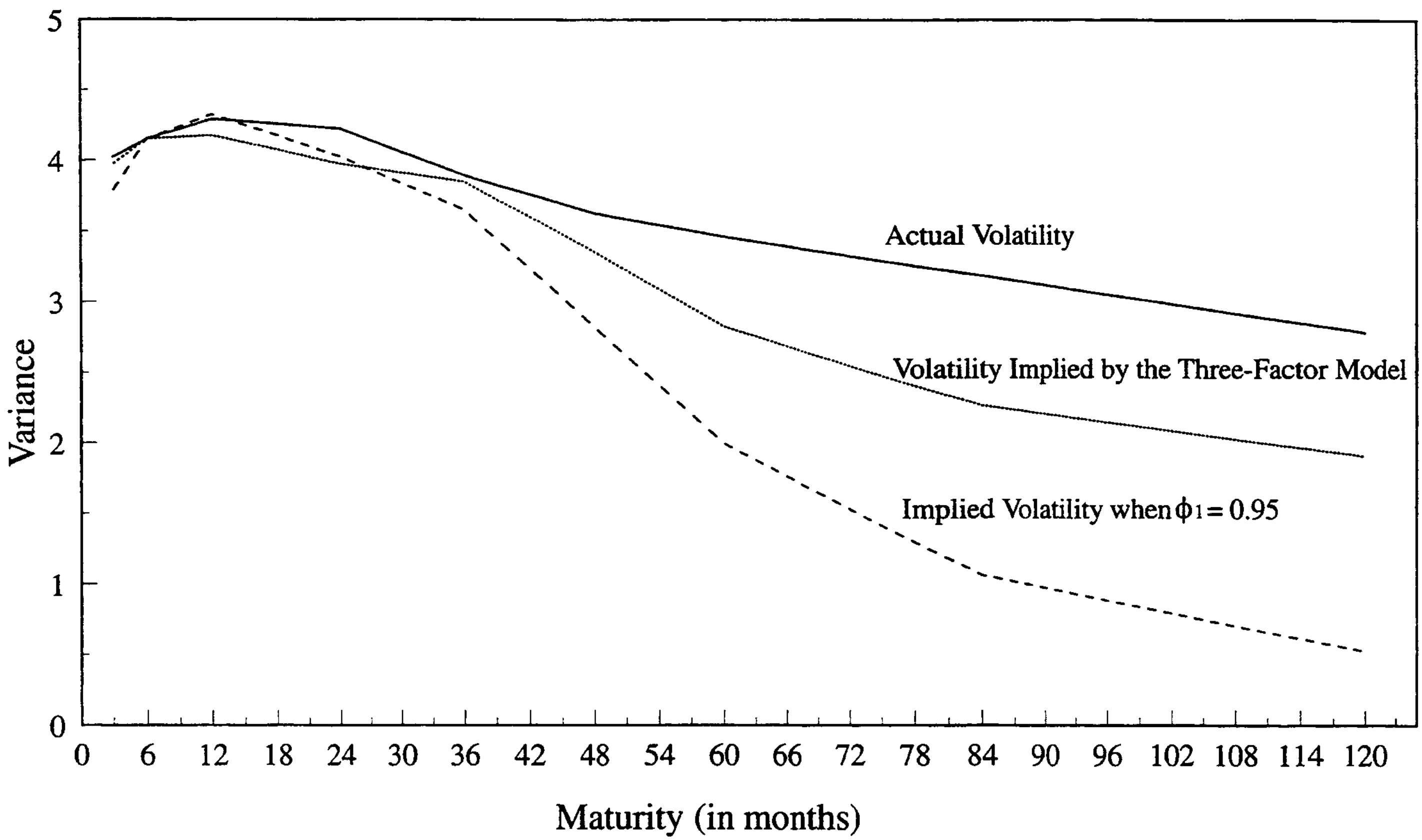

Figure 6. Actual and Implied Volatility Curves. The volatility curve implied by the three-factor model plots the unconditional variance of the implied yields based on the full sample estimates reported inTable 3. The volatility implied when $\phi_1 = 0.95$ is based on the same parameters except for $\phi_1 = 0.95$.

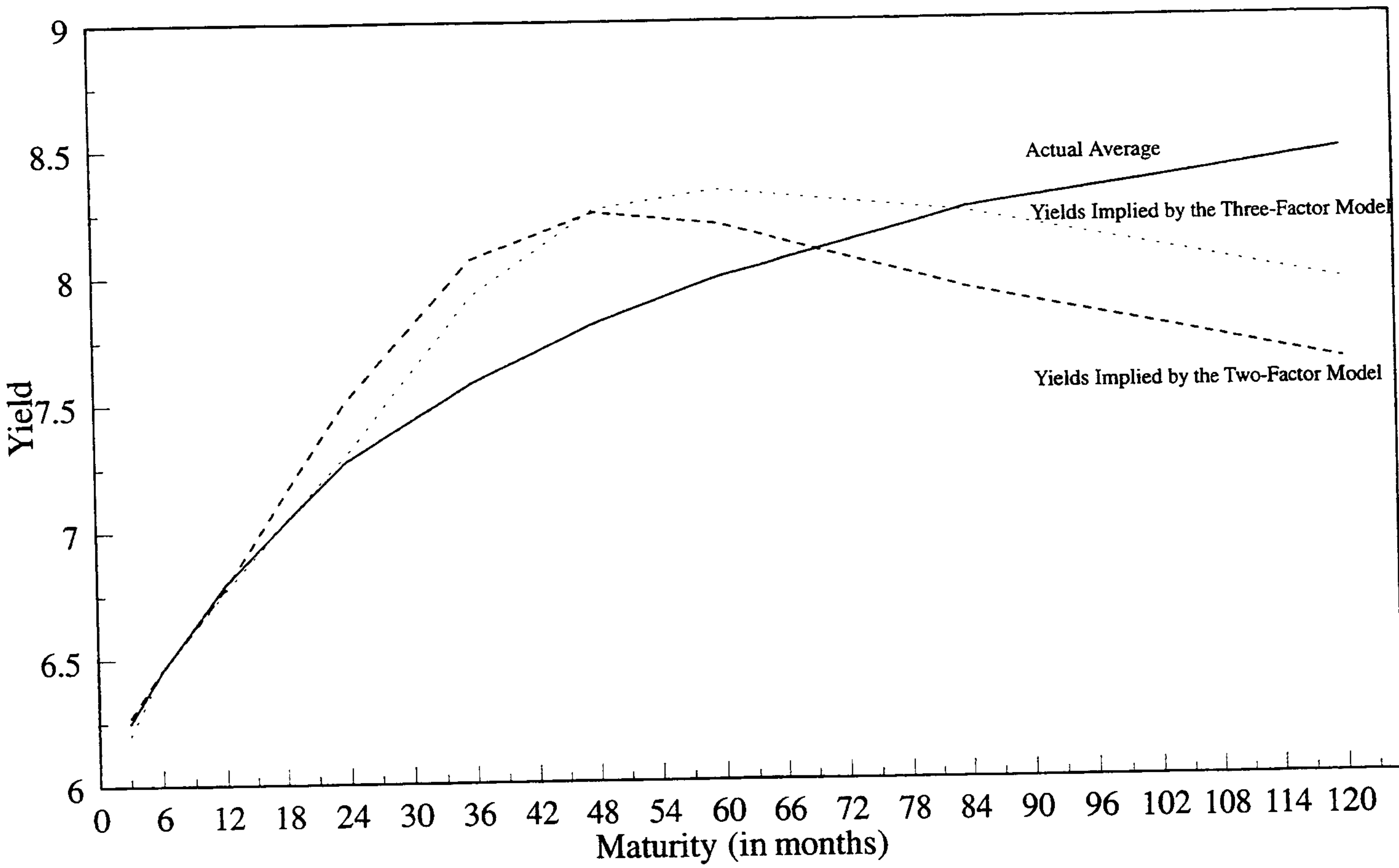

Figure 7. Actual and Implied Yield Curves from the Three-Factor and Two-Factor Models. The curves implied by the three-factor and two-factor models are based on the unconditional mean of the yields derived from the full-sample estimate reported in Table 4.

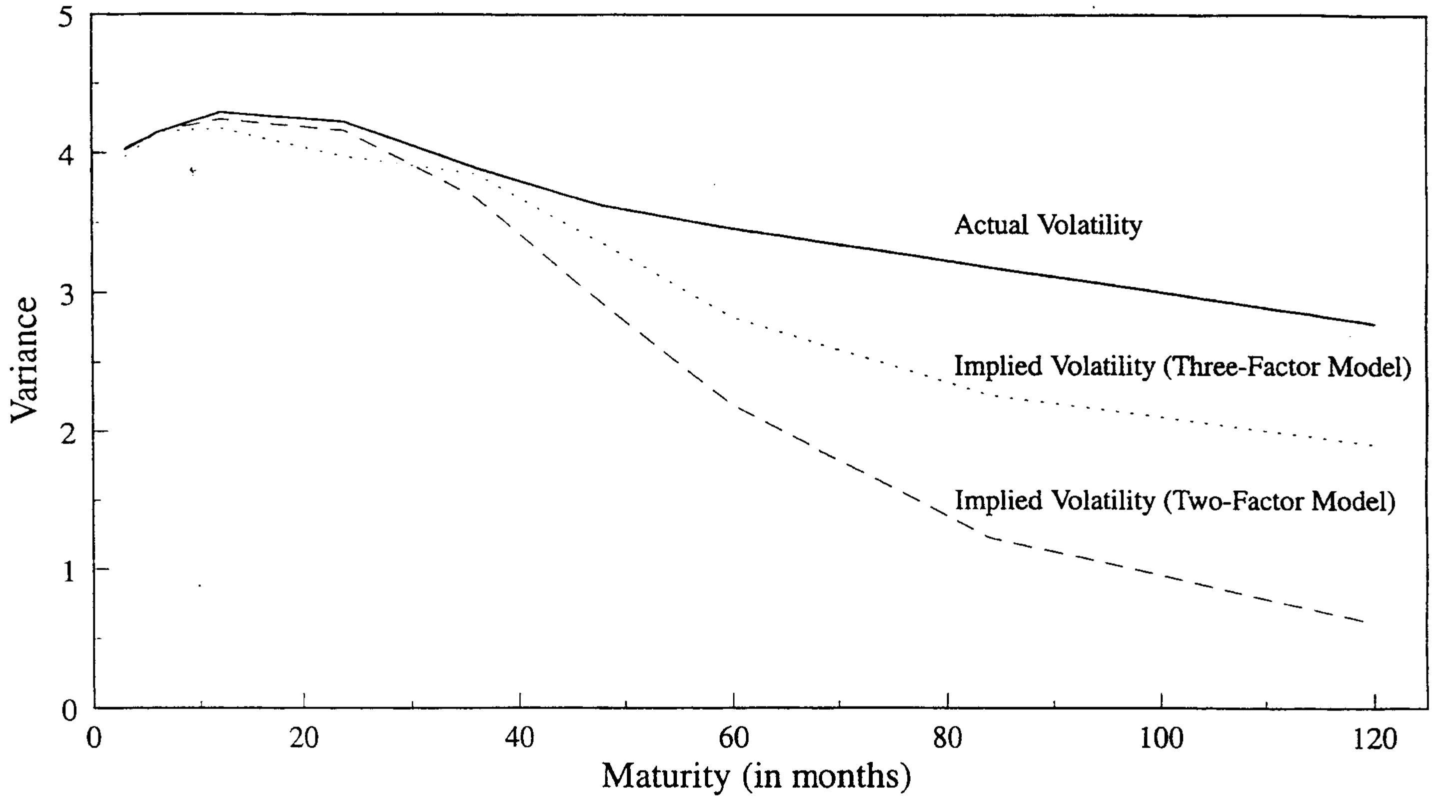

Figure 8. Actual and Implied Volatility Curves. The implied curves are based on the unconditional variance of the implied yields derived from the full-sample estimates reported in Table 4.