Title
Unit Roots and the Estimation of Interest Rate Dynamics

Authors
Ball, Clifford A.
Torous, Walter N.

Publication Date
1992-05-18

Unit Roots and the Estimation of Interest Rate Dynamics

Clifford A. Ball*
Owen Graduate School of Management,
Vanderbilt University,
Nashville, TN 37203 USA.
(615) 322-2909

Walter N. Torous†
Anderson Graduate School of Management
University of California at Los Angeles
Los Angeles, CA 90024 USA.
(310) 825-4059

Revision: May 18, 1995

*Part of this research was supported from a grant from the Dean's Fund and the Financial Markets Research Center at the Owen School of Management, Vanderbilt University.

†We thank Yacine Ait-Sahalia, Eduardo Schwartz, Bruce Tuckman, seminar participants at the University of California, Berkeley, as well as an anonymous referee for their comments and suggestions. We are solely responsible for any remaining errors.

Unit Roots and the Estimation of Interest Rate Dynamics

Abstract

This paper investigates the time series estimation of Cox, Ingersoll, and Ross's square-root, mean-reverting specification for interest rate dynamics. For *a priori* reasonable mean reversion, the stochastic behavior of interest rates is sufficiently close to a non-stationary process with a unit root so that least squares, the generalized method of moments, as well as maximum likelihood estimation provide upward biased estimates of the model's speed of adjustment coefficient. Corresponding bond yields, as a result, exhibit excessive mean reversion. In addition, estimates of the specification's long-term mean interest rate are seen to display erratic behavior when near a unit root. These conclusions are robust to assuming multiple state variable specifications, such as Brennan and Schwartz's two factor model of interest rate dynamics. We also document conditions under which this unit root problem can be alleviated when the cross-sectional restrictions of the Cox, Ingersoll, and Ross single factor term structure model are imposed.

1 Introduction

Single factor models of the term structure of interest rates posit that default-free bond prices depend upon the prevailing instantaneous riskless rate of interest. Resultant default-free bond prices satisfy a second-order partial differential equation, the solution of which varies with the specification of instantaneous riskless interest rate dynamics. Altering this underlying interest rate process then alters the resultant term structure.

Characterizing the dynamics of the instantaneous riskless rate by a square-root, mean-reverting diffusion process, Cox, Ingersoll, and Ross [1985] provide closed-form expressions for default-free bond prices which are amenable to empirical testing. Tests of this model include, among others, Brown and Dybvig [1986], Brown and Schaefer [1994], Chen and Scott [1993], Gibbons and Ramaswamy [1993], and Pearson and Sun [1994].

This paper investigates the estimation of Cox, Ingersoll, and Ross's square-root, mean-reverting specification for interest rate dynamics. Their term structure model is often implemented in practice by estimating the parameters of this process on the basis of a time series of short-term interest rates and then using these estimates in the relevant closed-form expressions for default-free bond prices. Unfortunately, this time series regression is subject to a potentially serious unit root problem (Dickey and Fuller [1979, 1981], Evans and Savin [1981, 1984], and Phillips [1987]).

Intuitively, the instantaneous riskless rate of interest is mean reverting in the Cox, Ingersoll and Ross specification, the degree of mean reversion governed by a speed of adjustment coefficient. The slower this speed of adjustment, the closer the interest rate's stochastic behavior is to a non-stationary process with an exact unit root. In particular, for *a priori* reasonable mean reversion, the stochastic behavior of interest rates is sufficiently close to that of a unit root process so that the sampling distribution of the estimated speed of adjustment coefficient behaves much like its sampling distribution when there is an exact unit root. As a result, we demonstrate that the speed of adjustment coefficient is estimated with a significant upward bias which is not eliminated even in the large sample sizes typically encountered in practice. This unit root problem will consequently impart a systematic bias to single factor term structures: estimated yield curves will converge far too quickly to the corresponding long-term interest rate.

This upward bias is consistent with the empirical results of Gibbons and Ramaswamy [1993] and Pearson and Sun [1994] who find their estimated adjustment speeds to be too quick given the historical behavior of bond prices. For example, using Treasury bill data only, Gibbons and Ramaswamy estimate the speed of adjustment coefficient to be 12.43, while Pearson and Sun's estimate is 9.24. Both of these estimates imply a mean half-life for the estimated interest rate process (that is, the expected time for the process to return halfway to its long-term mean) of less than one month.

Using the generalized method of moments, as opposed to least squares, does not remedy this unit root problem. Furthermore, the problem is not unique to single factor models of the term structure. We confirm similar problems, as well as others, when using time series techniques to estimate Brennan and Schwartz's [1979, 1982] two factor specification for the dynamics of the short and long rates of interest.

We also provide evidence of unstable behavior in the corresponding estimates of the long-term mean interest rate of the Cox, Ingersoll and Ross specification when using shorter time series as evidenced by occasional but substantial outliers. While these alternative estimation procedures will provide consistent estimates of the long-term mean of a stationary interest rate process, larger sample sizes are needed when near a unit root before their asymptotically optimal properties become evident.

The Cox, Ingersoll, and Ross term structure formula imposes a cross-sectional restriction across default-free bonds of different maturities. If we rely exclusively on these cross-sectional properties, as in Brown and Dybvig [1986] or Brown and Schaefer [1994], the speed of adjustment coefficient cannot be identified and, as such, cannot be estimated. But combining this cross-sectional restriction with the model's time series properties allows the speed of adjustment coefficient to be estimated more efficiently since the shape of the term structure provides potentially valuable information about this parameter. For example, Pearson and Sun's estimated adjustment speed falls when they include longer term bonds, while Chen and Scott, who use both short term bills as well as longer term bonds, estimate the speed of adjustment coefficient to be only 0.47, implying a mean half-life of approximately eighteen months. In this paper, we cast this dynamic estimation strategy into a state space framework and use the Kalman filter to obtain maximum likelihood estimates of the parameters of the Cox, Ingersoll, and Ross term structure model. Our analysis indicates that the magnitude of the unit root problem depends on the shape of the term structure and, as a consequence, on whether long-term yields are included. By simply concentrating on the yield curve's short end, say, maturities of less than one year, the informativeness of the yield curve's shape cannot be taken advantage of, and once again the speed of adjustment coefficient will be estimated with significant upward bias. We also find that the inclusion of bonds of varying maturities allows the long-term mean of the Cox, Ingersoll and Ross interest rate process to be estimated more efficiently.

The plan of this paper is as follows. Section 2 investigates the time series regression corresponding to Cox, Ingersoll, and Ross's square-root, mean reverting specification for instantaneous riskless interest rates. We demonstrate the corresponding unit root problem and assess its estimation effects in Section 3. Section 4 shows that these estimation problems also arise if the generalized method of moments is used to estimate the parameters of the square-root, mean-reverting specification. In Section 5 we document unit root problems in the time series regressions corresponding to Brennan and Schwartz's two factor specification for short and long rates of interest. In section 6, we add the cross-sectional restrictions of the Cox, Ingersoll, and Ross single factor term structure model and document the conditions under which this unit root problem can be alleviated. Section 7 provides a summary and conclusions.

2 Single Factor Models

Cox, Ingersoll, and Ross characterize the dynamics of the instantaneous riskless rate r by[1]:

$$dr = \kappa(\theta - r)dt + \sigma\sqrt{r}dz, \tag{1}$$

where z is standard Brownian Motion. For κ, $\theta > 0$, this specification corresponds to a continuous-time first order autoregressive process with the stochastic interest rate r being pulled towards its long-term mean θ at a rate governed by the speed of adjustment coefficient κ.

The resulting process is Markov with a computable transition density so that the joint likelihood of a set of observations $\{r_t\}$ may be calculated. Unfortunately, the instantaneous

[1]Without loss of generality, throughout this paper we interpret interest rates as nominal rates of interest. Sufficient conditions to ensure this are provided by Cox, Ingersoll, and Ross [1985].

riskless rate is not directly observable. However, we assume that observable short rates provide an adequate proxy. [2]

The probability density of the short rate at some future time, s, conditional on its value at the current time, t, is given by:

$$f(r(s), s; r(t), t) = c\, e^{(-u-v)}(v/u)^{q/2} I_q(2\sqrt{uv}), \tag{2}$$

where

$$\begin{aligned} c &= \frac{2\kappa}{\sigma^2(1 - e^{-\kappa(s-t)})}, \\ u &= cr(t)e^{-\kappa(s-t)}, \\ v &= cr(s), \\ q &= \frac{2\kappa\theta}{\sigma^2} - 1, \end{aligned}$$

and $I_q(.)$ is the modified Bessel function of the first kind of order q. The conditional distribution function of the short rate $r(s)$ is a constant multiple times a non-central chi-squared distribution. Specifically,

$$r(s) \mid r(t) \sim \chi^2(\nu, \lambda)/(2c)$$

where ν is the degrees of freedom $\frac{4\kappa\theta}{\sigma^2}$ and λ is the non-centrality parameter $2u$. Noting that the mean and variance of a non-central chi-squared distribution are $\nu + \lambda$ and $2(\nu + 2\lambda)$, respectively, Cox, Ingersoll, and Ross also show that:

$$\begin{aligned} E[r(s) \mid r(t)] &= r(t)e^{-\kappa(s-t)} + \theta(1 - e^{-\kappa(s-t)}), \\ Var[r(s) \mid r(t)] &= r(t)(\sigma^2/\kappa)(e^{-\kappa(s-t)} - e^{-2\kappa(s-t)}) + \theta(\sigma^2/2\kappa)(1 - e^{-\kappa(s-t)})^2./22 \end{aligned}$$

Diffusions represented by stochastic differential equations behave locally like Brownian Motions. Therefore, for a small increment in time Δ, the conditional distribution of $r(t+\Delta)$

[2]We relax this assumption later.

given $r(t)$ will be approximately normal with computable moments. Adopting subscripts to indicate the corresponding discrete model, we have that

$$
\begin{aligned}
E[r_{t+\Delta} \mid r_t] &= r_t (1-\kappa\Delta) + \theta(\kappa\Delta) + O(\Delta^2) \\
Var[r_{t+\Delta} \mid r_t] &= r_t \sigma^2 \Delta + O(\Delta^2).
\end{aligned}
$$

Given a sequence of observations $\{r_t\}$, the discrete-time model implies the following generalized least squares regression:

$$
r_{t+\Delta} = \beta_0 + \beta_1 r_t + \sigma \epsilon_t \sqrt{\Delta} \tag{3}
$$

with the $\{\epsilon_t\}$ assumed distributed independently normal with mean zero and variance proportional to r_t.

Alternatively,

$$
\frac{r_{t+\Delta}}{\sqrt{r_t}} = \frac{\beta_0}{\sqrt{r_t}} + \beta_1 \sqrt{r_t} + \xi_t, \tag{4}
$$

where the error ξ_t satisfies the standard regression assumptions with constant variance ω^2. Estimation follows by least squares regression with $\hat{\kappa} = (1-\hat{\beta}_1)/\Delta$, $\theta = \hat{\beta}_0/(1-\hat{\beta}_1)$, and $\hat{\sigma}^2 = \hat{\omega}^2\Delta$. However, as we demonstrate next, for κ near zero and, particularly, for small Δ, there exists a unit root problem which may have significant effects on this estimation.

3 The Unit Root Problem

We review the unit root problem by considering the time series

$$
\begin{aligned}
X_t &= \rho X_{t-1} + e_t \qquad t = 1, 2, \ldots, \\
X_0 &= 0,
\end{aligned}
$$

where $\{e_t\}$ are assumed distributed independently normal with mean 0 and variance σ^2. The corresponding sample first order autocorrelation coefficient $\hat{\rho}_n$ is given by

$$\hat{\rho}_n = \sum_{t=1}^{n} X_{t+1}X_t / \sum_{t=1}^{n} X_t^2$$

where n is the number of observations.[3]

Fuller [1976], Dickey and Fuller [1979, 1981] and others have examined the statistical properties of this estimator in some detail. Briefly summarizing their results, for values of ρ less than one, $\sqrt{n}(\hat{\rho} - \rho)$ is distributed asymptotically normal with mean zero and variance $\sigma_\rho^2 = (1 - \rho^2)$. For $\rho = 1$, when the underlying process is no longer stationary, the estimator converges more rapidly but its asymptotic normality breaks down. In fact, in the presence of a unit root, $n(\hat{\rho} - \rho)$ has a nonnormal limiting distribution, tabulated, for example, in Fuller [1976], page 371, which is markedly skewed to the left. Estimation of the process's mean also becomes problematical in this non-stationary case.

Furthermore, for values of ρ close to but less than one, the near-integrated case, Phillips [1987] demonstrates that the sample first order autocorrelation coefficient behaves very much as in the nonstationary case, $\rho = 1$.[4] Fuller's [1976] simulation evidence also confirms that for values of ρ close to but less than one, the sampling distribution of $\sqrt{n}(\hat{\rho} - \rho)$ is far from normal, even for large samples, being skewed to the left with the mean of the empirical distribution reflecting the downward biasedness of $\hat{\rho}$. See also Evans and Savin [1981, 1984][5].

The discrete generalized least squares regression model outlined in Section 2 should have

[3] When there is no ambiguity, we will drop the n subscript on $\hat{\rho}_n$.

[4] Phillips [1988] develops asymptotics relying on functionals of the Ohrnstein-Uhlenbeck process. This is to be contrasted with the standard Dickey-Fuller asymptotics for the exact unit root case which involve functionals of Brownian Motion.

[5] Similar results are evident when a constant is included in the regression specification. We would expect the long-term mean's estimation to also be troublesome in the near unit root case.

similar problems. Assuming annualized interest rate data available on a monthly basis, $\Delta = 1/12$, we select the following plausible parameter values: $\theta = 0.05$, $\sigma^2 = 0.0025$, and $\kappa = 1$. Therefore,

$$E[r_{t+\Delta} \mid r_t] = r_t e^{-\kappa\Delta} + \theta(1 - e^{-\kappa\Delta}) = 0.9200 r_t + 0.0800\theta.$$

Under these assumptions, there is a reversion to the mean at 8% per month. Since $\beta_1 \equiv (1 - \kappa\Delta) = 0.92 \sim 1$, we have a potential unit root problem. Least squares regression, as well as other time series estimation procedures, will provide downward biased estimates of β_1.[6]

We perform a simulation study to investigate the statistical properties of the generalized least squares estimators. For these assumed parameter values, we generate five and twenty year's of monthly data, n = 60 and 240 observations, respectively.[7] Initial spot rates of interest vary from .03 to .07, in increments of .01. We repeat each experiment 1,000 times and summarize the sampling results in Tables 1a and 1b.

Notice that the sampling distribution of $\hat{\kappa}$ is biased upwards throughout, with the bias being most pronounced for initial spot rates close to θ. For example, using five years of simulated monthly data and an initial spot rate of 0.05 (the assumed long-term mean), we observe a 94.0% upward bias in $\hat{\kappa}$. However, all else equal, setting the initial spot rate at 0.07, the upward bias in $\hat{\kappa}$ is reduced to 67.9%. Also, as the sample size increases, this bias

[6]The assumed κ value lies between the estimates of Gibbons and Ramaswamy (12.43) and Pearson and Sun (9.24), and that of Chen and Scott (0.47). All else equal, for κ values smaller than one, holding Δ fixed, it follows that $\beta_1 \equiv (1 - \kappa\Delta)$ is closer to one, thereby magnifying the unit root problem and increasing the resultant percentage bias. A similar conclusion follows if, alternatively, κ is held fixed but Δ is decreased to reflect more frequent data sampling.

[7]Interest rate data are simulated using a first order discrete-time approximation to expression (1) assuming 360 time steps per year. We then sample once every 30 of the resultant observations to obtain our simulated monthly data.

diminishes. For example, twenty years of monthly data gives an upward bias of 16.6% in $\hat{\kappa}$ for an initial spot rate of 0.07.

Intuitively, we obtain more information on the speed of adjustment coefficient κ when interest rates are distant from their long-term mean, as well as when we have more observations. For the larger sample size, the other estimators are essentially unbiased and appear to be well-behaved. However, for the smaller sample size we see significant instability in $\hat{\theta}$, particularly for low initial spot rates.[8]

To investigate the economic significance of this bias in $\hat{\kappa}$, Table 2 compares Cox, Ingersoll, and Ross bond yields across a number of terms to maturity for the assumed value of $\kappa = 1$ versus mean $\hat{\kappa}$, holding all else constant. We set the initial spot rate at 0.05 and use mean estimated speed of adjustment coefficients assuming n = 60 as well as n = 240 months of data. To complete the specification of the Cox, Ingersoll, and Ross term structure model, we must additionally specify the market price of interest rate risk, μ. We alternatively set $\mu = 0$, and $\mu = -.5$, the latter a value consistent with that estimated by Brennan and Schwartz [1982]. For $\mu = 0$ (Panel A) an approximately flat term structure obtains, while an upward sloping term structure results for $\mu = -.5$ (Panel B). In both Panels we see minimal bias at very short maturities as both the assumed and estimated term structures are near the initial spot rate. Since the flat term structure is insensitive to the speed of adjustment coefficient, there is also little, if any, bias evident in Panel A for longer maturities. However, in the case of the upward sloping term structure, where yields are more sensitive to the speed of adjustment coefficient, the resultant bias is pronounced at longer maturities. For example, from Panel B, the percentage difference in ten-year yields is on the order of 25% for n = 60

[8] For example, for n=60 and an initial spot rate of.03, the minimum value of $\hat{\theta}$ is -9.21, while its maximum value is 38.93.

months, while a percentage difference in ten-year yields of approximately 17% obtains for n = 240 months.

The use of weekly ($\Delta = 1/52$), rather than monthly ($\Delta = 1/12$), data is explored in Table 3. In particular, rather than using 5 years of monthly data (n = 60), we assume 5 years of weekly data (n = 260). As can be seen from Table 3, the sampling properties of the estimators, especially the upward biasedness of $\hat{\kappa}$, are not dramatically altered. Also, estimates of θ are unstable. Intuitively, the benefits of using more data are offset by sampling that data more frequently and thereby magnifying the unit root problem.

4 Method of Moments Estimation

In this section, we determine to what extent, if any, our previous results are due to least squares estimation. To do so, we use the method of moments to estimate the parameters of the Cox, Ingersoll, and Ross's square-root, mean-reverting specification of interest rate dynamics. Unlike maximum likelihood estimation, method of moments estimation relies on only a few moments of the posited statistical distribution of interest rate changes. We also investigate the generalized method of moments. This procedure requires that the distribution of interest rates be stationary and ergodic and that the relevant moments exist.

4.1 Method of Moments

Cox, Ingersoll, and Ross's specification for interest rate dynamics requires that $N = 3$ parameters be estimated, $\varphi = (\kappa, \theta, \sigma^2)$. Suppose that $f_t(\varphi)$ denotes a vector of 3 functionally independent theoretical moments evaluated at time t, and that m_t denotes the corresponding set of sample moments, $t = 1, 2, \cdots, n$.

Letting

$$f(\varphi) = \frac{1}{n}\sum_{t=1}^{n} f_t(\varphi) \qquad \text{and} \qquad m = \frac{1}{n}\sum_{t=1}^{n} m_t,$$

the method of moments estimator is that $\hat{\varphi}$ which satisfies

$$f(\hat{\varphi}) = m$$

or equivalently

$$(f(\hat{\varphi}) - m)'(f(\hat{\varphi}) - m) = 0. \tag{5}$$

In general, there are a number of potential problems with method of moments estimators. For example, while $\hat{\varphi}$ is a consistent estimator, it may not be asymptotically efficient. Furthermore, depending upon the choice of moment conditions, the solution of (5) may not be unique or may not lie in a feasible parameter region.

As an example, suppose that for our problem we choose the following 3 moment conditions:

$$f_t(\varphi) = \begin{bmatrix} E(r_{t+1}|r_t) \\ E(r_{t+1}^2|r_t) \\ E(r_{t+1}^3|r_t) \end{bmatrix} = \begin{bmatrix} (\nu + \lambda r_t)/2c \\ ((\nu + \lambda r_t)^2 + (2\nu + 4\lambda r_t))/(2c)^2 \\ (8((\eta + 3\lambda r_t) + 6(\nu + 2\lambda r_t)(\nu + \lambda r_t) + (\nu + \lambda r_t)^3)/(2c)^3 \end{bmatrix},$$

$$m_t = \begin{bmatrix} r_{t+1} \\ r_{t+1}^2 \\ r_{t+1}^3 \end{bmatrix}$$

where, without loss of generality, we assume $\Delta = 1$.

Even though these moments are functionally independent, for practical parameter values, the first moment is dominant. In particular, when we are near a unit root, that is, for small

$\kappa > 0$,

$$v \sim 0, \ \ \lambda/(2c) \sim 1, \ \ \text{and} \ \ \lambda/(2c)^2 \sim 0.$$

Substituting into the above moment conditions gives

$$E(r_{t+1}^i | r_t) \sim r_t^i \ \ \text{for all integers } i.$$

In other words, for κ very small but positive, all the sample moments can be approximately matched by their population counterparts. For the sample sizes and the parameter values previously considered, simulation evidence (not reported here) confirms that the method of moments based on these moment conditions consistently provides κ estimates indistinguishable from zero.

Of course, with the appropriate choice of moment conditions, the method of moments reduces to generalized least squares. In particular, the following 3 moments:

$$f_t(\varphi) = \begin{bmatrix} \xi_{t+1} \\ \xi_{t+1} r_t \\ \xi_{t+1}^2 - \omega^2 \end{bmatrix}$$

where $\xi_{t+1} = r_{t+1}/\sqrt{r_t} - (\theta\kappa\Delta)/\sqrt{r_t} - (1 - \kappa\Delta)\sqrt{r_t}$ characterize generalized least squares.[9] Under these conditions, the method of moments will exhibit the previously discussed effects of the unit root problem. To see whether additional moment restrictions alleviate the unit root problem, we next turn our attention to the generalized method of moments.

4.2 Generalized Method of Moments

The generalized method of moments allows the number of moment conditions to exceed the number of parameters to be estimated. A test of the posited model's goodness of fit is

[9]The first moment condition ensures the residuals sum to zero, the second gives the normal equation for generalized least squares, while the third moment condition estimates the variance by using the sum of squared standardized residuals.

provided by these overidentifying restrictions.

We consider the following four moment conditions suggested by Chan, Karolyi, Longstaff, and Sanders [1992] in their empirical analysis of short term interest rate dynamics

$$f_t(\varphi) = \begin{bmatrix} \epsilon_{t+1} \\ \epsilon_{t+1} r_t \\ \epsilon_{t+1}^2 - \sigma^2 r_t \\ (\epsilon_{t+1}^2 - \sigma^2 r_t) r_t \end{bmatrix}$$

where, $\epsilon_{t+1} = r_{t+1} - (\beta_0 + \beta_1 r_t)$. The corresponding generalized method of moments estimator, $\hat{\varphi}$, is given by

$$\underset{\varphi}{argmin} \quad (f(\varphi) - m)' W_n (f(\varphi) - m)$$

where W_n is a symmetric nonsingular weighting matrix. We use the Newey-West [1987] weighting matrix. Since the number of moments now exceeds the number of parameters to be estimated, the minimum value of this objective function will not in general equal zero for any φ.

Simulations are also carried out to assess the performance of these generalized method of moments estimators. For comparison purposes, we select the same parameter values, initial spot rates, sample sizes, and number of replications as used in our investigation of generalized least squares. Summary statistics are reported in Tables 4a and 4b.

The two estimation procedures are in broad agreement with the resultant bias in $\hat{\kappa}$ being evident using either procedure. However, the bias appears to be more pronounced for the generalized method of moments estimators and, furthermore, these estimators tend to have larger standard deviations. As before, for the smaller sample size, the sampling distribution of $\hat{\theta}$ is contaminated by a number of extreme values resulting in excessive skewness and kurtosis.[10] However, these effects are not evident in the larger sample size.[11]

5 Two State Models

We have so far investigated estimation problems associated with a univariate specification of interest rate dynamics. Many authors, including Brennan and Schwartz [1979, 1982], have introduced multiple factor models of the term structure.[12] In particular, Brennan and Schwartz assume a bivariate specification of interest rate dynamics where, as before, r represents the short rate of interest, l now represents the rate of return on a long term consol bond, with their dynamics given by

$$\begin{aligned} dr &= \beta_1(r,l,t)dt + \eta_1(r,l,t)dz_1 \\ dl &= \beta_2(r,l,t)dt + \eta_2(r,l,t)dz_2 \end{aligned}$$

where z_1 and z_2 are standardized Brownian Motions such that $E[dz_1 dz_2] = \rho dt$. This section inquires into the effects of the unit root problem on the estimation of this bivariate specification.

For estimation purposes, Brennan and Schwartz specialize their bivariate specification to

$$\begin{aligned} dr &= (a_1 + b_1(l - r))dt + r\sigma_1 dz_1 \\ dl &= l(a_2 + b_2 r + c_2 l)dt + l\sigma_2 dz_2 \end{aligned}$$

[10]For example, in Table 4a, for an initial spot rate of.05, the minimum value of $\hat{\theta}$ is -0.36 while its maximum is 748.18. By contrast, for an initial spot rate of .06, the minimum value of $\hat{\theta}$ is -8.27 while its maximum is 15.88. As a result, the standard deviation of the sampling distribution of $\hat{\theta}$ is significantly smaller in the latter case.

[11]As expected, the sampling properties of the resultant χ^2 goodness-of-fit statistics, not reported here, do not indicate excessive rejection of the posited model.

[12]Alternative specifications of the state variables have also been put forward. For example, see Schaefer and Schwartz [1984] and Longstaff and Schwartz [1992].

and discretize it as follows

$$
\begin{aligned}
r_{t+\Delta} - r_t &= (a_1 + b_1(l_t - r_t))\Delta + r_t\sigma_1\sqrt{\Delta\epsilon_1} \\
l_{t+\Delta} - l_t &= l_t(a_2 + b_2 r_t + c_2 l_t)\Delta + l_t\sigma_2\sqrt{\Delta\epsilon_2}
\end{aligned}
$$

where ϵ_1 and ϵ_2 are assumed bivariate normal with zero means, unit variances, and correlation ρ, and where Δ is again a small increment in time. Notice that now the short rate does not revert towards a constant but rather towards the long rate which itself varies through time.

The above system represents two linear equations with contemporaneously correlated error components and can be estimated in a variety of ways. One approach, maximum likelihood estimation, optimizes the corresponding likelihood function using nonlinear methods. Alternatively, as pointed out by Brennan and Schwartz [1982], this linear system can be estimated with less computational effort using generalized multivariate regression with a consistent estimate of the error covariance obtained by applying ordinary least squares equation by equation (Seemingly Unrelated Regression Estimation (SURE) as introduced by Zellner [1962, 1963] and Zellner and Huang [1962]).

Rather than this two-stage procedure, we implement the following more efficient three-stage procedure. In the first stage, we apply ordinary least squares to the two equations separately and estimate ρ from the resultant residuals. In the second stage, using the estimated error covariance, we perform a generalized multivariate regression. The residuals from this multivariate regression are used to update the estimated error covariance which is then used in the third stage to perform a subsequent generalized multivariate regression. We use this three-stage procedure since it gives estimation results in very close agreement to maximum likelihood estimation, without the extensive computational effort.

To examine this bivariate specification as an extension of the previously investigated univariate model, we set $a_1 = a_2 = 0$:

$$\begin{aligned} r_{t+\Delta} - r_t &= b_1(l_t - r_t)\Delta + r_t\sigma_1\sqrt{\Delta}\epsilon_1 \\ l_{t+\Delta} - l_t &= l_t[b_2 r_t + c_2 l_t]\Delta + l_t\sigma_2\sqrt{\Delta}\epsilon_2. \end{aligned}$$

This implies that for small σ_2 as well as small b_2 and c_2, the short rate reverts to an essentially constant long rate.[13]

We calibrate our simulation experiments to allow us to compare and contrast the statistical properties of this two factor specification relative to its one factor counterpart. To that end, we choose $b_1 = 1$ since this parameter corresponds directly to the univariate model's speed of adjustment coefficient κ. We also set $\sigma_1 = 0.2236$ and initialize the term structure with $r_0 = l_0 = 0.05$, since, under these assumptions, the bivariate model's short rate process has approximately the same incremental variance as that previously assumed for Cox, Ingersoll, and Ross's square root specification. We parameterize σ_2, the volatility of the long rate, to range from 0.001 to 0.01 to 0.05, allowing us to assess the estimation effects of an increasingly important second factor. The correlation between increments in the short rate and the long rate process is assumed to be $\rho = .5$.[14] Finally, we simulate the posited bivariate system with $b_2 = c_2 = 0$, though we estimate these parameters and comment upon their sampling properties below.

As in our univariate results, the simulations assume five and twenty year's of monthly data ($n = 60$ and 240 observations, respectively). We repeat the experiments 1000 times

[13]We stress, however, that the limiting one factor model does not correspond exactly to the square root model studied earlier. In the present case, the infinitesimal volatility is proportional to the short rate rather than its square root. As such, we would not expect to obtain identical estimation effects.

[14]We also ran these simulations assuming $\rho = 0$ with similar results.

and summarize the sampling results in Tables 5a and 5b.

Notice that the sampling distribution of $\hat{b}_1$ is biased upwards throughout, the bias being more pronounced the shorter the sample period. For example, using five years of simulated monthly data, we observe an upward bias of approximately 45% in $\hat{b}_1$. However, all else equal, with twenty years of data, this upward bias is reduced to approximately 5%. Increasing σ_2 has a minimal effect on the upward biasedness of $\hat{b}_1$.

Recall that our simulations assume that $b_2 = c_2 = 0$. Nevertheless, it is interesting to examine the sampling properties of $\hat{b}_2$ and $\hat{c}_2$. Especially for larger values of σ_2, these estimators are extremely volatile and significantly negatively correlated with one another. In fact, it appears that the estimators cancel each other's effects, yet their actual values are, on average, quite large. Based on these results, one might conclude, quite spuriously, that a spread variable, $r_t - l_t$, plays a statistically significant role in explaining long rate dynamics. In fact, these exact effects are evident in the empirical results of Brennan and Schwartz [1982]. Notwithstanding various economic reasons why the spread between short and long rates may play an important role in a model of the term structure of interest rates (Schaefer and Schwartz [1984]), our sampling results suggest that the unit root problem exaggerates the spread's statistical importance.

6 Imposing Cross-Sectional Term Structure Restrictions

Returning to the single factor model, we now extend our estimation efforts beyond simply the time series properties of the square-root, mean-reverting specification for interest rate dynamics. In this case, Cox, Ingersoll, and Ross's closed-form term structure model imposes a cross-sectional restriction on default-free bond prices prevailing at any point in time. Since the shape of this term structure depends on the model's speed of adjustment coefficient, we investigate whether incorporating this cross-sectional restriction minimizes the previously documented unit root problem.

However, unlike Brown and Dybvig [1986] and Brown and Schaefer [1994], we do not rely exclusively on the term structure restrictions of the Cox, Ingersoll, and Ross model since these cross-sectional estimation techniques exclude information on how the term structure evolves over time and, as such, cannot separately identify and estimate the speed of adjustment coefficient. Rather, we follow, among others, Chen and Scott [1993], Gibbons and Ramaswamy [1993], and Pearson and Sun [1994], who estimate the parameters of the Cox, Ingersoll, and Ross term structure model by combining both the model's time series and cross-sectional properties.

We cast this dynamic estimation strategy into a linear state-space framework, similar to Pennacchi [1991], which recognizes that the underlying state variable, $r(t)$, is unobserved. However, bond prices, or equivalently, bond yields, which we assume are observed with error, are linear functions of $r(t)$. By treating the state variable as an unobservable, we ensure that our estimation results are not due to proxying $r(t)$ by a particular short-term interest rate. Measurement errors in the observed bond yields reflect noise arising from, for example, the averaging of bid and ask quotes or possible quotation errors.

From Cox, Ingersoll and Ross, the state variable $r(t)$ evolves according to a transition equation which can be written in discrete-time form as

$$r_{t+\Delta} = \theta\kappa\Delta + (1 - \kappa\Delta)r_t + \sigma\epsilon_t\sqrt{\Delta}$$

where the ϵ_t are independent normal with zero mean and variance proportional to r_t. Unlike Gibbons and Ramaswamy [1993], notice that we rely on the conditional density of the state variable rather than its unconditional or steady-state density.

Suppose that at each date t we observe the yields of discount bonds with M maturities, τ_1, τ_2 $\cdots$, τ_M. Each observed yield, y_{obs}, is given by the corresponding Cox, Ingersoll, and Ross yield plus an independent normally distributed measurement error:

$$y_{obs}(r, \tau_i; t) = -lnA(\tau_i)/\tau_i + (B(\tau_i)/\tau_i)r + error(\tau_i, t) \tag{6}$$

where

$$\begin{aligned} A(\tau_i) &= \left[\frac{2\gamma e^{[(\kappa+\mu+\gamma)\tau_i]/2}}{(\gamma+\kappa+\mu)(e^{\gamma\tau_i}-1)+2\gamma}\right]^{2\kappa\theta/\sigma^2}, \\ B(\tau_i) &= \frac{2(e^{\gamma\tau_i}-1)}{(\gamma+\kappa+\mu)(e^{\gamma\tau_i}-1)+2\gamma}, \\ \gamma &= ((\kappa+\mu)^2+2\sigma^2)^{1/2}, \end{aligned}$$

$i = 1, 2, \cdots, M$, $t = 1, 2, \cdots, T$. As before, μ represents the market price of interest rate risk. We assume that each measurement equation is perturbed by an error; that is, each yield is observed with noise. This is distinct from Pearson and Sun [1994] who assume that both the 13- and 26-week T-bill rates are observed without measurement errors or Chen and Scott [1993] who assume that the 13-week T-bill rate is observed without error. By including both the model's time series and cross-sectional properties, and noting that bond yields are assumed observed with error, all parameters of the Cox, Ingersoll, and Ross term structure model now become individually identified.

Given these assumptions, the likelihood of the observed yields $Y_T = \{y_{obs}(r, \tau_i, t), i =$

$1, 2, \cdots, M,\ t = 1, 2, \cdots, T\}$ can be written as[15]

$$L(Y_T; \psi) = \prod_{t=1}^{T} p(y_t \mid Y_{t-1}),$$

where $y_t = \{y_{obs}(r, \tau_i, t), i = 1, 2, \cdots, M\}$. Notice that $p(y_t \mid Y_{t-1})$ is the conditional density of the t^{th} set of observations given all observations through $t - 1$.

Our state-space framework assumes that the measurement errors $error(\tau_i, t)$ are normally distributed and independent over time, that the measurement equation (6) is linear in the state variable $r(t)$, and that $r(t)$ follows an autoregressive process with normal disturbances. Under these conditions we may use the Kalman filter[16] to optimally predict the underlying unobservable state variable, $r(t)$, as well as to efficiently evaluate the likelihood function[17].

Employing the Kalman filter to evaluate the likelihood, numerical optimization of L over $\psi = (\kappa, \theta, \sigma^2, \mu)$, generates the maximum likelihood estimator $\hat{\psi}$ of the parameters of the Cox, Ingersoll, and Ross term structure model.

To assess the statistical properties of this maximum likelihood estimation procedure, we conduct a number of sampling experiments. The cross-sectional restrictions of Cox, Ingersoll, and Ross's term structure model are incorporated by considering $M = 3$ yields, each perturbed by independently normally distributed error terms with standard deviations

[15]We use y_t to denote the t^{th} set of observations while Y_t denotes the set of these observations through t.

[16]Strictly speaking, since we model the dynamics of the underlying state variable by a square-root process, we are introducing heteroscedasticity and so we actually use the extended Kalman filter (see, Harvey [1989] for an excellent description of the Kalman and extended Kalman filters) which nevertheless provides satisfactory asymptotic results.

[17]In fact, even if the measurement errors are not normally distributed, the Kalman filter may still be used. Provided we correctly specify the first two moments of the measurement error distribution, Harvey, Ruiz, and Shephard [1992] demonstrate that using the Kalman filter and assuming normal measurement errors provides parameter estimators that are consistent and asymptotically normally distributed. Therefore, the Kalman filter approach is quite general in its applicability to any term structure model which generates a linear relationship between yields and underlying state variables when bond prices are observed with error.

of 10 basis points. As before, we assume $\kappa = 1$, $\theta = 0.05$, and $\sigma = 0.05$. To investigate the effects of the term structure's shape, we assume μ = -0.50 and, alternatively, $\mu = 0$. Since we fix the initial rate, r_0, at 0.05 throughout, upward sloping term structures are generated, on average, by assuming μ = -0.50, while flatter term structures, on average, obtain for μ = 0. The informativeness of the term structure's short end is examined by assuming that τ_1 = 1/12 (or 1 month), τ_2 = 1/2 (or 6 months), and τ_3 = 1 (or 1 year), while the effects of incorporating the term structure's longer end are investigated by assuming that τ_1 = 1/12, τ_2 = 1, and τ_3 = 5 (or 5 years). We restrict our attention to yields having a maximum maturity of 5 years as this is the maximum maturity of default-free bonds in the CRSP (Center for Research in Security Prices) bond file used by many researchers to empirically investigate term structure models. Each sampling experiment is repeated 500 times.

Table 6a summarizes the results when $n = 60$ months of data are used, while the results assuming $n = 240$ months of data are summarized in Table 6b. The sampling properties and, in particular, the magnitude of the unit root problem, now depend on the shape of the term structure and, as a consequence, on whether or not longer term yields are included in the empirical analysis. In particular, incorporating cross-sectional information does not necessarily alleviate the upward biasedness in estimating the speed of adjustment coefficient.

From Table 6a we see that when the term structure is, on average, upward sloping (μ = -0.5) and we include longer term yields (τ_1 = 1/12, τ_2 =1, and τ_3 = 5), the bias in $\hat{\kappa}$ is minimal, 2.7%. Intuitively, in this case the shape of the term structure provides information on the speed of adjustment coefficient and longer term yields allow us to use this information to accurately estimate κ. To emphasize this latter point, note from Table 6a that when we only include yields from the short-end of the term structure (τ_1 = 1/12, τ_2 = 1/2, and τ_3 =

1), the bias in $\hat{\kappa}$ is once again substantial, 19.0%, even though the term structure is upward sloping (μ = -0.5). Therefore, to minimize the effects of the unit root problem, dynamic estimation strategies should not concentrate solely on the short-end of the term structure, like Gibbons and Ramaswamy [1993]. However, if the term structure is, on average, flat ($\mu = 0$), then even incorporating information from its longer end will not alleviate the unit root problem. Intuitively, a flat term structure provides little information on the speed of adjustment coefficient and, as such, incorporating longer term yields will not eliminate the upward bias in estimating κ. For example, when $\mu = 0$ in Table 6a, the upward bias in $\hat{\kappa}$ is 22.7% for $\tau_1 = 1/12$, $\tau_2 = 1/2$, and $\tau_3 =1$, yet is still 14.9% for $\tau_1 = 1/12$, $\tau_2 =1$, and $\tau_3 =5$. Comparing Tables 6a and 6b, the sampling results are similar but we see that using 240 months of data (Table 6b), rather than 60 months (Table 6a), reduces the upward bias in $\hat{\kappa}$. In both Tables 6a and 6b we see that θ is estimated accurately, while σ^2 is estimated with upward bias throughout. Since bond prices are affected significantly by the value of θ, it is not surprising that the inclusion of cross-sectional information markedly stabilizes the estimation of θ. The market price of interest rate risk appears to be estimated unbiasedly, though not precisely.

7 Summary and Conclusions

Models of interest rate dynamics typically incorporate mean reversion. For example, Cox, Ingersoll, and Ross [1985] characterize interest rate dynamics by a square-root, mean reverting diffusion process and provide closed-form expressions for default-free bond prices. However, the time series estimation of the corresponding speed of adjustment coefficient is subject to a potentially serious unit root problem. As we demonstrate, both least squares and the generalized method of moments give significantly upward biased estimates of this parameter, resulting in estimated yield curves which converge far too quickly. Observed bond prices will not be consistent with such excessive mean reversion. Significant instability in estimating the specification's long-term mean interest rate may also be expected in the near unit root case.

Incorporating the term structure model's cross-sectional restriction across bonds of different maturities can potentially alleviate this unit root problem. This follows since to the extent that the yield curve is not flat, it provides valuable information about the speed of adjustment coefficient. However, if the yield curve's short end only is sampled, then this information cannot be taken advantage of, and once again the speed of adjustment coefficient will be estimated with an upward bias. Also, estimation of the long-term mean interest rate is markedly improved when this cross-sectional information on bond prices is included.

Our conclusions are reminiscent of Merton's [1980] observation that it is difficult to precisely estimate expected returns from a time series of realized stock returns. This may not be a significant consideration in pricing equity derivatives since the drift of the stock price process is irrelevant and the variance of returns can be estimated far more accurately. Unfortunately, this is not the case with interest rate derivatives where it is important to accurately estimate the mean reversion inherent in the drift of the underlying interest rate process.

8 References

1. M. Brennan and E. Schwartz, 'A Continuous Time Approach to the Pricing of Bonds,' *Journal of Banking and Finance*, 3, 133-155, [1979].

2. ——————, 'An Equilibrium Model of Bond Pricing and a Test of Market Efficiency,' *Journal of Financial and Quantitative Analysis*, 17, 75-100, [1982].

3. S. J. Brown and P. H. Dybvig, 'The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates', *Journal of Finance*, pp. 617-632, [1986].

4. R. Brown and S. Schaefer, 'The Term Structure of Real Interest Rates and the Cox, Ingersoll, and Ross Model', *Journal of Financial Economics*, Vol. 35, No. 1, pp. 3-42, [1994]

5. K. Chan, G. Karolyi, F. Longstaff, and A. Sanders, 'An Empirical Comparison of Alternative Models of the Short-Tern Interest Rate,' *Journal of Finance*, 47, pp. 1209-1227, [1992].

6. R-R. Chen and L. Scott, 'Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates,' *Journal of Fixed Income*, 3, pp. 14-31, [1993].

7. J. C. Cox, J. E. Ingersoll, and S. A. Ross, 'A Theory of the Term Structure of Interest Rates', *Econometrica*, 53 pp.385-408, [1985].

8. D. A. Dickey and W. A. Fuller, 'Distribution of the Estimators for Autoregressive Time Series', *Journal of the American Statistical Association*, 74 pp. 427-431, [1979].

9. ——————, 'Likelihood Ratio Statistics for Autoregressive Time Series with Unit Roots', *Econometrica*, 49 pp. 1057-1072, [1981].

10. G. B. A. Evans and N. E. Savin, 'Testing for Unit Roots: 1', *Econometrica*, 49, pp. 753-779, [1981].

11. ——————, 'Testing for Unit Roots: 2', *Econometrica*, 52, pp. 1241-1269, [1984].

12. W. A. Fuller, Introduction to Statistical Time Series, Wiley Incorporated, New York, [1976].

13. M. R. Gibbons and K. Ramaswamy, 'A Test of the Cox, Ingersoll, and Ross Model of the Term Structure', *The Review of Financial Studies*, 6, pp. 619-658, [1993].

14. A. C. Harvey, Forecasting, Structural Times Series Models and the Kalman Filter, Cambridge University Press, Cambridge, [1989].

15. A. C. Harvey, E. Ruiz, and N. Shephard, 'Multivariate Stochastic Variance Models', *Working Paper*, London School of Economics, [1992].

16. F. Longstaff and E. S. Schwartz, 'Interest Rate Volatility and the Term Structure: A Two Factor General Equilibrium Model,' *Journal of Finance*, Vol. 47, pp. 1259-1282, [1992].

17. R. C. Merton, 'On Estimating the Expected Return on the Market - An Exploratory Investigation,' *Journal of Financial Economics*, Vol. 8, pp. 323-361, [1980].

18. W. K. Newey and K. D. West, 'A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica*, Vol. 55, pp. 703-708, [1987].

19. N. D. Pearson and T. Sun, 'Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model', *Journal of Finance*, Vol. 49, pp. 1279-1304, [1994].

20. G. G. Pennacchi, 'Identifying the Dynamics of Real Interest Rates and Inflation: Evidence Using Survey Data', *Review of Financial Studies*, Vol. 4, Number 1, pp. 53-89, [1991].

21. P. C. Phillips, 'Time Series Regression with a Unit Root', *Econometrica*, 55 pp. 277-301, [1987].

22. ——————, 'Regression theory for Near-Integrated Time Series', *Econometrica*, 56 pp. 1021-1043, [1988].

23. S. M. Schaefer and E. S. Schwartz, 'A Two Factor Model of the Term Structure: An Approximate Analytic Solution,' *Journal of Financial and Quantitative Analysis*, vol. 19, pp. 413-424, [1984].

24. A. Zellner, 'An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias,' *Journal of the American Statistical Association*, vol. 57, pp. 348-368, [1962].

25. A. Zellner, 'Estimators for Seemingly Unrelated Regressions: Some Exact Finite Sample Results,' *Journal of the American Statistical Association*, vol. 58, pp. 977-992, [1963].

26. A. Zellner and D. S. Huang, 'Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations,' *International Economic Review*, vol. 3, pp. 300-313, [1962].

Table 1a

Sampling properties of $\hat{\kappa}, \hat{\theta}, \hat{\sigma^2}$ using Generalized Least Squares and assuming $n = 60$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Cox, Ingersoll, and Ross mean-reverting, square-root specification for interest rate dynamics. A sample size of $n = 60$ months of data is assumed. The experiment is based on 1,000 replications.

$r_0 = 0.03$	$n = 60$	$\kappa = 1.00$	$\theta = 0.05$	$\sigma^2 = 0.0025$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.5049	0.7033	0.7254	0.7589
$\hat{\theta} * 10^2$	5.0858	1.6790	13.2000	265.7335
$\hat{\sigma^2} * 10^3$	2.4952	0.4686	0.3551	-0.0464
$r_0 = 0.04$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.78323	0.8983	0.7188	0.4729
$\hat{\theta} * 10^2$	5.0370	1.1001	-0.6769	97.4124
$\hat{\sigma^2} * 10^3$	2.4877	0.4679	0.3543	-0.0632
$r_0 = 0.05$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.9397	1.0407	1.0019	1.3055
$\hat{\theta} * 10^2$	5.1147	1.2943	15.6818	317.1570
$\hat{\sigma^2} * 10^3$	2.4808	0.4697	0.3560	-0.0810
$r_0 = 0.06$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.8588	0.9819	1.0919	2.2738
$\hat{\theta} * 10^2$	5.0856	0.5480	-0.6040	5.9130
$\hat{\sigma^2} * 10^3$	2.4814	0.4700	0.3673	-0.0688
$r_0 = 0.07$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.6789	0.8433	1.0201	1.6311
$\hat{\theta} * 10^2$	5.0529	1.8897	-27.3917	822.6653
$\hat{\sigma^2} * 10^3$	2.4850	0.4690	0.3868	-0.02675

Table 1b

Sampling properties of $\hat{\kappa}, \hat{\theta}, \hat{\sigma^2}$ using Generalized Least Squares and assuming $n = 240$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Cox, Ingersoll, and Ross mean-reverting, square-root specification for interest rate dynamics. A sample size of $n = 240$ months of data is assumed. The experiment is based on $1,000$ replications.

$r_0 = 0.03$	$n = 240$	$\kappa = 1.00$	$\theta = 0.05$	$\sigma^2 = 0.0025$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1604	0.3345	0.7903	1.2640
$\hat{\theta} * 10^2$	5.0141	0.2489	0.1495	0.01221
$\hat{\sigma^2} * 10^3$	2.5049	0.2309	0.2286	0.1466
$r_0 = 0.04$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1846	0.3642	0.8123	1.0835
$\hat{\theta} * 10^2$	5.0170	0.2469	0.1544	0.0217
$\hat{\sigma^2} * 10^3$	2.5047	0.2307	0.2297	0.1457
$r_0 = 0.05$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1905	0.3725	0.8204	1.0698
$\hat{\theta} * 10^2$	5.0206	0.2460	0.1567	0.0233
$\hat{\sigma^2} * 10^3$	2.5047	0.2306	0.2325	0.1430
$r_0 = 0.06$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1822	0.3640	0.7985	0.9535
$\hat{\theta} * 10^2$	5.0242	0.2462	0.1549	0.0183
$\hat{\sigma^2} * 10^3$	2.5047	0.2305	0.2354	0.1421
$r_0 = 0.07$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1664	0.3463	0.7551	0.7278
$\hat{\theta} * 10^2$	5.0272	0.2470	0.1505	0.0119
$\hat{\sigma^2} * 10^3$	2.5047	0.2304	0.2377	0.1425

Table 2

Yield Bias induced by Estimated $\hat{\kappa}$.

This table provides percentage differences in Cox, Ingersoll, and Ross zero-coupon yields assuming $\kappa = mean(\hat{\kappa})$ as opposed to $\kappa = 1$ for alternative terms to maturity. The speed of adjustment coefficient is estimated by generalized least squares assuming $n = 60$ as well as $n = 240$ months of data. In Panel A, the market price of interest rate risk, μ, equals 0, giving an approximately flat term structure, while in Panel B, μ equals -0.50, giving an upward sloping term structure.

Panel A	$\mu = 0$	
Term to Maturity (yrs)	Yield Bias % (n=60)	Yield Bias % (n=240)
1	-0.008	-0.002
2	-0.025	-0.009
3	-0.040	-0.015
5	-0.057	-0.023
10	-0.073	-0.030
20	-0.081	-0.034
30	-0.083	-0.036
Panel B	$\mu = -0.50$	
Term to Maturity (yrs)	Yield Bias % (n=60)	Yield Bias % (n=240)
1	3.806	1.920
2	9.196	4.949
3	13.611	7.812
5	19.401	11.930
10	25.350	16.673
20	28.456	19.287
30	29.437	20.117

Table 3

Sampling properties of $\hat{\kappa}, \hat{\theta}, \hat{\sigma^2}$ using Generalized Least Squares and assuming $n = 260$ weeks of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Cox, Ingersoll, and Ross mean-reverting, square-root specification for interest rate dynamics. A sample size of $n = 260$ weeks of data is assumed. The experiment is based on 1,000 replications.

$r_0 = 0.03$	$n = 260$	$\kappa = 1.0$	$\theta = 0.05$	$\sigma^2 = 0.0025$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.5922	0.8152	1.1642	1.7794
$\hat{\theta} * 10^2$	4.9808	0.6541	-1.0418	17.9331
$\hat{\sigma^2} * 10^3$	2.4903	0.2197	0.0759	-0.0613
$r_0 = 0.04$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.8736	1.0203	1.4405	4.4161
$\hat{\theta} * 10^2$	4.9943	0.5462	0.1250	3.4888
$\hat{\sigma^2} * 10^3$	2.4889	0.2203	0.0726	-0.0678
$r_0 = 0.05$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.9994	1.1111	1.4522	4.0991
$\hat{\theta} * 10^2$	5.0195	0.5748	-4.4314	74.0034
$\hat{\sigma^2} * 10^3$	2.4880	0.2205	0.0743	-0.0677
$r_0 = 0.06$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.8989	1.0524	1.4468	3.9041
$\hat{\theta} * 10^2$	5.0612	0.5238	-0.5444	3.0684
$\hat{\sigma^2} * 10^3$	2.4883	0.2198	0.0816	-0.0638
$r_0 = 0.07$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.6895	0.8607	1.0336	1.2446
$\hat{\theta} * 10^2$	5.1784	3.2093	29.8649	924.3690
$\hat{\sigma^2} * 10^3$	2.4896	0.2192	0.0862	-0.0571

Table 4a

Sampling properties of $\hat{\kappa}, \hat{\theta}, \hat{\sigma^2}$ using Generalized Method of Moments and assuming $n = 60$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Cox, Ingersoll, and Ross mean-reverting, square-root specification for interest rate dynamics. A sample size of $n = 60$ months of data is assumed. The experiment is based on 1,000 replications.

$r_0 = 0.03$	$n = 60$	$\kappa = 1.0$	$\theta = 0.05$	$\sigma^2 = 0.0025$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.5624	0.7242	0.6691	0.6713
$\hat{\theta} * 10^2$	6.1098	25.1463	27.1560	785.3395
$\hat{\sigma^2} * 10^3$	2.3536	0.4596	0.3499	-0.0262
$r_0 = 0.04$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.8286	0.9257	0.6915	0.4638
$\hat{\theta} * 10^2$	6.0197	19.7535	21.9303	469.5128
$\hat{\sigma^2} * 10^3$	2.3424	0.4571	0.3376	-0.0921
$r_0 = 0.05$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.9842	1.0702	0.9476	1.2331
$\hat{\theta} * 10^2$	6.9556	34.7982	18.9392	364.8025
$\hat{\sigma^2} * 10^3$	2.3304	0.4591	0.3438	-0.1087
$r_0 = 0.06$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.8930	1.0220	1.0912	2.4978
$\hat{\theta} * 10^2$	5.0839	0.8646	-0.7876	94.8303
$\hat{\sigma^2} * 10^3$	2.3277	0.4609	0.3452	-0.0570
$r_0 = 0.07$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.6932	0.8786	0.9950	1.4981
$\hat{\theta} * 10^2$	4.6129	15.1467	-31.4901	991.4160
$\hat{\sigma^2} * 10^3$	2.3340	0.4596	0.3709	0.0589

Table 4b

Sampling properties of $\hat{\kappa}, \hat{\theta}, \hat{\sigma^2}$ using Generalized Method of Moments and assuming $n = 240$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Cox, Ingersoll, and Ross mean-reverting, square-root specification for interest rate dynamics. A sample size of $n = 240$ months of data is assumed. The experiment is based on 1,000 replications.

$r_0 = 0.03$	$n = 240$	$\kappa = 1.0$	$\theta = 0.05$	$\sigma^2 = 0.0025$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1816	0.3476	0.7208	1.1048
$\hat{\theta} * 10^2$	5.0077	0.2529	0.1746	-0.0245
$\hat{\sigma^2} * 10^3$	2.4510	0.2332	0.1958	0.1101
$r_0 = 0.04$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.2043	0.3782	0.7605	0.9364
$\hat{\theta} * 10^2$	5.0013	0.2508	0.1784	-0.0208
$\hat{\sigma^2} * 10^3$	2.4502	0.2332	0.1951	0.1088
$r_0 = 0.05$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.2099	0.3877	0.7789	0.9168
$\hat{\theta} * 10^2$	5.0156	0.2497	0.1760	-0.0298
$\hat{\sigma^2} * 10^3$	2.4502	0.2332	0.1952	0.1064
$r_0 = 0.06$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1986	0.3785	0.7796	0.9589
$\hat{\theta} * 10^2$	5.0201	0.2495	0.1665	-0.0323
$\hat{\sigma^2} * 10^3$	2.4509	0.2328	0.2019	0.0796
$r_0 = 0.07$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1783	0.3785	0.7468	0.8058
$\hat{\theta} * 10^2$	5.0235	0.2498	0.1533	-0.0244
$\hat{\sigma^2} * 10^3$	2.4521	0.2324	0.2157	0.0796

Table 5a

Simulation of $\hat{b}_1, \hat{b}_2, \hat{c}_2, \hat{\sigma}_1, \hat{\sigma}_2$, and $\hat{\rho}$ using a three-stage Seemingly Unrelated Regression Estimation (SURE) procedure assuming $n = 60$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Brennan and Schwartz two factor specification of interest rate dynamics. A sample size of $n = 60$ months of data is assumed. The experiment is based on $1,000$ replications.

$b_1 = 1.00$	$b_2 = 0.00$	$c_2 = 0.00$	$\sigma_1 = 0.2236$	$\rho = 0.50$
$r_0 = 0.05$ $l_0 = 0.05$ $n = 60$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{b}_1$	1.4318	0.8898	1.2712	1.3925
$\hat{b}_2$	-0.0428	0.0823	-0.2341	-0.0380
$\hat{c}_2$	0.0410	0.0818	0.3309	-0.1002
$\hat{\sigma}_1 * 10^1$	2.2043	0.1843	1.0102	-1.9728
$(\sigma_2 = 0.0010)$ $\hat{\sigma}_2 * 10^3$	0.9727	0.1083	1.0107	-1.9713
$\hat{\rho}$	0.5008	0.1139	1.0630	-1.8384
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{b}_1$	1.4446	0.9211	1.3150	1.5441
$\hat{b}_2$	-0.4094	0.8431	-0.2240	0.0292
$\hat{c}_2$	0.3914	0.8386	0.3278	-0.0447
$\hat{\sigma}_1 * 10^1$	2.2040	0.1835	1.0102	-1.9729
$(\sigma_2 = 0.010)$ $\hat{\sigma}_2 * 10^2$	0.9730	0.0826	1.0107	-1.9712
$\hat{\rho}$	0.5009	0.1139	1.0630	-1.8385
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{b}_1$	1.4906	1.0263	1.2502	0.9380
$\hat{b}_2$	-1.4664	4.5245	-0.1183	0.2476
$\hat{c}_2$	1.3657	4.5008	0.2405	0.0856
$\hat{\sigma}_1 * 10^1$	2.2031	0.1823	1.0100	-1.973
$(\sigma_2 = 0.050)$ $\hat{\sigma}_2 * 10^2$	4.8720	0.4163	1.0109	-1.9707
$\hat{\rho}$	0.5015	0.1141	1.0631	-1.8383

Table 5b

Simulation of $\hat{b}_1, \hat{b}_2, \hat{c}_2, \hat{\sigma}_1, \hat{\sigma}_2$, and $\hat{\rho}$ using a three-stage Seemingly Unrelated Regression Estimation (SURE) procedure assuming $n = 240$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Brennan and Schwartz two factor specification of interest rate dynamics. A sample size of $n = 240$ months of data is assumed. The experiment is based on 1,000 replications.

$b_1 = 1.00$	$b_2 = 0.00$	$c_2 = 0.00$	$\sigma_1 = 0.2236$	$\rho = 0.50$
$r_0 = 0.05$ $l_0 = 0.05$ $n = 240$				
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{b}_1$	1.0382	0.3029	1.1476	2.7899
$\hat{b}_2$	-0.0106	0.0312	-0.5362	0.7995
$\hat{c}_2$	0.0106	0.0310	0.4796	0.6885
$\hat{\sigma}_1 * 10$	2.2310	0.0943	1.0027	-1.9929
$(\sigma_2 = 0.0010)$ $\hat{\sigma}_2 * 10^3$	0.9926	0.0994	1.0037	-1.9901
$\hat{\rho}$	0.4940	0.0486	1.0138	-1.9638
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{b}_1$	1.0394	0.3095	1.1540	2.8104
$\hat{b}_2$	-0.1003	0.3160	-0.5318	0.8383
$\hat{c}_2$	0.0994	0.3143	0.4721	0.7214
$\hat{\sigma}_1 * 10$	2.2310	0.0944	1.0027	-1.9929
$(\sigma_2 = 0.010)$ $\hat{\sigma}_2 * 10^2$	0.9927	0.0532	1.0037	-1.9901
$\hat{\rho}$	0.4940	0.0486	1.0138	-1.9638
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{b}_1$	1.0468	0.3366	1.0296	2.0700
$\hat{b}_2$	-0.3503	1.6669	-0.4903	1.3369
$\hat{c}_2$	0.3204	1.6565	0.4240	1.1769
$\hat{\sigma}_1 * 10$	2.2307	0.0941	1.0027	-1.9929
$(\sigma_2 = 0.050)$ $\hat{\sigma}_2 * 10^2$	4.9645	0.2538	1.0037	-1.9901
$\hat{\rho}$	0.4940	0.0486	1.0138	-1.9637

Table 6a

Sampling properties of $\hat{\kappa}, \hat{\theta}, \hat{\sigma^2}, \hat{\mu}$ using Maximum Likelihood Estimation and assuming $n = 60$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Cox, Ingersoll, and Ross term structure model. A variety of term to maturity zero-coupon bonds are used. The market price of interest rate risk is $\mu = 0$ or $\mu = -0.5$. A sample size of $n = 60$ months of data is assumed. The experiment is based on 500 replications.

	$\kappa = 1.0$	$\theta = 0.05$	$\sigma^2 = 0.0025$	
$n = 60$		$\tau_1 = 1/12$	$\tau_2 = 1$	$\tau_3 = 5$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.0274	0.1142	0.3169	0.5816
$\hat{\theta} * 10^2$	5.0157	0.5008	0.4295	0.4927
$\hat{\sigma^2} * 10^3$	3.3061	0.7030	0.2644	-0.1806
$(\mu = -0.5)$ $\hat{\mu} * 10$	-5.1514	1.0591	-0.3160	0.4683
$n = 60$		$\tau_1 = 1/12$	$\tau_2 = 1/2$	$\tau_3 = 1$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1902	0.1501	0.2094	0.1366
$\hat{\theta} * 10^2$	5.0162	0.4917	0.4170	0.4501
$\hat{\sigma^2} * 10^3$	3.2886	0.6919	0.2265	-0.2097
$(\mu = -0.5)$ $\hat{\mu} * 10$	-5.3337	1.1227	-0.2575	0.3055
$n = 60$		$\tau_1 = 1/12$	$\tau_2 = 1$	$\tau_3 = 5$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1489	0.1480	0.3081	-0.0110
$\hat{\theta} * 10^2$	5.0197	0.4959	0.4175	0.4905
$\hat{\sigma^2} * 10^3$	3.6506	0.8065	0.3214	-0.1375
$(\mu = 0.0)$ $\hat{\mu} * 10$	-0.0702	1.0993	-0.2765	0.3175
$n = 60$		$\tau_1 = 1/12$	$\tau_2 = 1/2$	$\tau_3 = 1$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.2268	0.1725	0.5998	1.0478
$\hat{\theta} * 10^2$	5.0208	0.4942	0.2446	-0.1254
$\hat{\sigma^2} * 10^3$	3.4104	0.6823	0.2938	-0.0704
$(\mu = 0.0)$ $\hat{\mu} * 10$	-0.0701	1.1294	-0.3239	-0.1840

Table 6b

Sampling properties of $\hat{\kappa}, \hat{\theta}, \hat{\sigma^2}, \hat{\mu}$ using Maximum Likelihood Estimation and assuming $n = 240$ months of data.

This table provides sample mean, standard deviation, skewness, and kurtosis of the estimated parameters of the Cox, Ingersoll, and Ross term structure model. A variety of term to maturity zero-coupon bonds are used. The market price of interest rate risk is $\mu = 0$ or $\mu = -0.50$. A sample size of $n = 240$ months of data is assumed. The experiment is based on 500 replications.

	$\kappa = 1.0$	$\theta = 0.05$	$\sigma^2 = 0.0025$	
$n = 240$		$\tau_1 = 1/12$	$\tau_2 = 1$	$\tau_3 = 5$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.0242	0.0567	0.1042	0.1794
$\hat{\theta} * 10^2$	5.0001	0.2533	0.2308	0.3180
$\hat{\sigma^2} * 10^3$	3.3596	0.3259	0.3400	0.1594
$(\mu = -0.5)\ \hat{\mu} * 10$	-5.1091	0.5331	-0.1194	0.1651
$n = 240$		$\tau_1 = 1/12$	$\tau_2 = 1/2$	$\tau_3 = 1$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1499	0.0738	0.1137	0.2765
$\hat{\theta} * 10^2$	5.0000	0.2529	0.2287	0.3128
$\hat{\sigma^2} * 10^3$	3.3103	0.3206	0.3598	0.1629
$(\mu = -0.5)\ \hat{\mu} * 10$	-5.2374	0.5720	-0.1300	0.1851
$n = 240$		$\tau_1 = 1/12$	$\tau_2 = 1$	$\tau_3 = 5$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1169	0.0703	0.3566	-0.2198
$\hat{\theta} * 10^2$	4.9954	0.2409	-0.0261	-0.2003
$\hat{\sigma^2} * 10^3$	3.6516	0.3615	0.4224	0.5794
$(\mu = 0.0)\ \hat{\mu} * 10$	-0.0417	0.5341	-0.3051	-0.0245
$n = 240$		$\tau_1 = 1/12$	$\tau_2 = 1/2$	$\tau_3 = 1$
Statistic	Mean	Std. Deviation	Skew	Kurtosis
$\hat{\kappa}$	1.1740	0.0788	0.4363	-0.0042
$\hat{\theta} * 10^2$	4.9956	0.2398	-0.0259	-0.1983
$\hat{\sigma^2} * 10^3$	3.4189	0.3278	0.4236	0.6519
$(\mu = 0.0)\ \hat{\mu} * 10$	-0.0419	0.5404	-0.3102	0.0208