Maximum Likelihood Estimation of the Cox-Ingersoll-Ross Model Using Particle Filters*

Giuliano De Rossi
Faculty of Economics and Politics
Cambridge University

October 24, 2003

Abstract

I show that the QML procedure, used in many papers in the current literature to estimate the CIR model from time series data, is based on an approximation of the latent factors' density that becomes very inaccurate for typical parameter values. I also argue that this issue is not addressed by the Monte Carlo experiments carried out to support the conclusion that the QML bias is negligible.
The second part of the paper describes a computationally efficient maximum likelihood estimator based on particle filters. The advantage of this estimator is to take into account the exact likelihood function while avoiding the huge computational burden associated with MCMC methods. The proposed methodology is implemented and tested on a sample of simulated data.

Introduction

The aim of this paper is to derive a maximum likelihood procedure to estimate the Cox, Ingersoll and Ross (1985) (henceforth CIR) model from multivariate time series data using Monte Carlo integration.

The CIR model and its generalised versions[1] have proved very popular both in the academic literature and among practitioners. The key to the success of the model is arguably the fact that it can replicate three features that are commonly observed in the data: nonnegativity of the interest rates, conditional heteroscedasticity and time-varying market prices of risk.

*These notes are part of the work for my Ph.D. dissertation. I am grateful to Andrew Harvey for continuous advice and comments. I would also like to thank Bent Jesper Christensen, Arnaud Doucet, William Fitzgerald, Simon Godsill, Siem Jan Koopman, Eric Zivot and seminar participants at the 2003 Econometric Study Group Annual Meeting, Bristol, Department of Engineering, Cambridge University, and Aarhus University for helpful comments on earlier versions of the paper. All remaining errors are my own.

[1] To name a few: the extended CIR model of Hull and White (1990), the two-factor model of Longstaff and Schwartz (1992), the Bakshi and Chen (1997) model, the multi-factor CIR models of Dai and Singleton (2000).

Estimates of the parameters can be obtained either from a cross section of data (typically zero coupon rates or bond prices) or from time series. The former solution is usually chosen in the financial literature when the purpose of the econometric analysis is to price derivative assets. The disadvantage of this approach is that the risk premium parameters cannot be identified because they are subsumed in the drift term. In addition, if the estimation is carried out sequentially at different points in time with different cross sections of rates, the estimated parameters can vary with sudden jumps when temporary shocks affect the observations.

There are several approaches to the estimation of the CIR model from time series data. Some authors use a univariate time series approach, often fitting the model to a short term observed interest rate.[2] To fully exploit the information contained in the observed interest rates, however, it is necessary to use rates for a wide range of maturities. This raises an issue of identifiability because if multivariate time series data are used, then the CIR model as such is clearly underidentified. It would be invariably rejected by the data because it implies that any cross section of rates observed at time t is a function of the parametrs (which are constant over time) and the value of the risk factors at time t. This is not true in general when one uses more interest rates than risk factors. One solution is to allow for discrepancies between the observed and the theoretical rates, i.e. to introduce measurement errors in the relation between the observed rates and the latent factors. This is typically done by assuming that the observed rates are affected by temporary shocks, usually Gaussian white noise errors with 'small' variance (relative to the variance of the innovations to the theoretical rates). When this assumption is made, it becomes impossible to write down explicitly the likelihood function for the observed rates, so that maximum likelihood estimation becomes a difficult problem. The most common solution, adopted among others by Lund (1997), de Jong and Santa Clara (1999), Duan and Simonato (1999), Duffee (1999), Geyer and Pichler (1999) and de Jong (2000), is to use a quasi maximum likelihood (QML) estimator based on the linear Kalman filter. The resulting estimator is biased and inconsistent, but the authors claim that the size of the bias is negligible for typical parameter values. I argue in Section II that the evidence provided to support the claim is not conclusive and that this approach can be misleading, especially when the Gaussian densities are used for filtering and smoothing.

MCMC estimation is an alternative to the QML approach and has recently been proposed by Lamoureaux and Witte (2002). They use a Gibbs sampler to draw from the conditional distributions of the state variables and parameter values, so as to find the marginal predictive densities of all the hyperparameters. The main drawback of this approach is that it turns out to be extremely time consuming because the state variables evolve very slowly. Lamoureaux and Witte (2002) report that it takes more than five days on a powerful machine to obtain a sufficient number of iterations for the two-factor model. They work with a sample of five instruments and around 350 weekly observations; presumably, the amount of time required to carry out the same procedure using more instruments would be significantly larger.

[2] Durham and Gallant (2002) give useful references on the subject.

The main purpose of the analysis proposed in this paper is to obtain an estimator that maximises the *exact* likelihood (and therefore is consistent and asymptotically efficient) avoiding the computational burden associated with the MCMC approach.

I The CIR model in state space form

The Cox-Ingersoll-Ross model is characterised by one factor, call it r, that evolves in continuous time according to

$$dr_t = \kappa\left(\theta - r_t\right)dt + \sigma\sqrt{r_t}dB_t \tag{1}$$

where B_t is Brownian motion, θ is the long term mean the factor reverts to and κ and σ are constants. Consider a sequence of points in time t_0, $t_0+\Delta$, $t_0+2\Delta$, ... for an arbitrary time interval Δ. Define the sequence $\{\alpha_s\}_{s=1}^{\infty} \equiv \left\{r_{t_0+(s-1)\Delta}\right\}_{s=1}^{\infty}$. Let

$$q \equiv \frac{2\kappa\theta}{\sigma^2} - 1, \qquad c \equiv \frac{2\kappa}{\sigma^2\left(1-e^{-\kappa\Delta}\right)}.$$

It can be shown that the density of $\boldsymbol{\alpha}$ can be written as:

$$\begin{aligned} p\left(\alpha_1\right) &= \text{gamma}\left(\alpha_1; c\left(1-e^{-\kappa\Delta}\right), q+1\right) \\ &= \frac{\left[c\left(1-e^{-\kappa\Delta}\right)\right]^{q+1}}{\Gamma\left(q+1\right)}\alpha_1^q e^{-\alpha_1 c\left(1-e^{-\kappa\Delta}\right)} \end{aligned} \tag{2}$$

and

$$\begin{aligned} p\left(\alpha_t|\alpha_{t-1}\right) &= 2c\ \text{noncentral}\chi^2\left(2c\alpha_t; 2q+2, 2ce^{-\kappa\Delta}\alpha_{t-1}\right) \\ &= c\exp\left[-c\left(\alpha_t + e^{-\kappa\Delta}\alpha_{t-1}\right)\right]\left(\frac{\alpha_t}{e^{-\kappa\Delta}\alpha_{t-1}}\right)^{q/2} \\ &\quad I_q\left(2c\sqrt{\alpha_t e^{-\kappa\Delta}\alpha_{t-1}}\right) \end{aligned} \tag{3}$$

where $I_q\left(\right)$ is the modified Bessel function of the first kind of order q. The observed interest rates $y_t\left(\tau\right)$, characterised by a *time to maturity* τ, are then obtained as:

$$y_t\left(\tau\right) = -A\left(\tau\right) + B\left(\tau\right)\alpha_t + \varepsilon_t$$

where

$$\begin{aligned} B\left(\tau\right) &= \frac{1}{\tau}\frac{2\left(e^{\gamma\tau}-1\right)}{2\gamma + \left(\kappa+\lambda+\gamma\right)\left(e^{\gamma\tau}-1\right)} \\ \gamma &= \sqrt{\left(\kappa+\lambda\right)^2 + 2\sigma^2} \end{aligned}$$

and

$$A(\tau) = \frac{q+1}{\tau} \log \left\{ \frac{2\gamma \exp\left(\tau\left(\kappa+\lambda+\gamma\right)/2\right)}{2\gamma + \left(\kappa+\lambda+\gamma\right)\left(e^{\gamma\tau}-1\right)} \right\}.$$

The quantities A and B are known functions of τ, the risk premium parameter λ and the CIR parameters κ, θ and σ. We can now gather N yields in the observable vector $\mathbf{y}_t$

$$\begin{aligned} \mathbf{y}_t &= \begin{bmatrix} y_t(\tau_1) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} B(\tau_1) \\ \vdots \\ B(\tau_N) \end{bmatrix} \alpha_t - \begin{bmatrix} A(\tau_1) \\ \vdots \\ A(\tau_N) \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{N,t} \end{bmatrix} \\ &= \mathbf{B}\alpha_t - \mathbf{A} + \boldsymbol{\varepsilon}. \end{aligned}$$

The measurement error ϵ_t is assumed to be IID Gaussian with covariance matrix hI_N. Therefore, the density of $y_t(\tau)$ conditional on α_t is

$$\begin{aligned} p\left(y_t(\tau)\,|\alpha_t\right) &= \mathcal{N}\left(y_t(\tau);\ -A(\tau) + B(\tau)\,\alpha_t,\ h\right) \\ &= \frac{1}{\sqrt{2\pi h}} \exp\left[-\frac{\left[y_t(\tau) + A(\tau) - B(\tau)\,\alpha_t\right]^2}{2h} \right]. \end{aligned}$$

Multi-factor CIR models typically assume that the short rate is a sum of factors evolving as (1). It can be shown that zero coupon interest rates are still linear functions of the factors.

II Quasi maximum likelihood in the existing literature

To obtain the QML estimator one needs to

1. replace expressions (2) and (3) with Gaussian densities;
2. find the parameter value that maximises the likelihood $p(\boldsymbol{y})$.

The Gaussian densities are chosen by matching the first two moments of $\alpha_t|\alpha_{t-1}$:

$$\begin{aligned} E\left(\alpha_t|\alpha_{t-1}\right) &= e^{-\kappa\Delta}\alpha_{t-1} + \left(1-e^{-\kappa\Delta}\right)\theta \\ var\left(\alpha_t|\alpha_{t-1}\right) &= \left(e^{-\kappa\Delta} - e^{-2\kappa\Delta}\right)\frac{\sigma^2}{\kappa}\alpha_{t-1} + \left(1-e^{-\kappa\Delta}\right)^2 \frac{\sigma^2}{2\kappa}\theta \end{aligned}$$

so that in the approximating density

$$\alpha_t = e^{-\kappa\Delta}\alpha_{t-1} + \left(1-e^{-\kappa\Delta}\right)\theta + \eta_t$$

where

$$\eta_t|\alpha_{t-1} \sim \mathcal{N}\left(0,\ \left(e^{-\kappa\Delta} - e^{-2\kappa\Delta}\right)\frac{\sigma^2}{\kappa}\alpha_{t-1} + \left(1-e^{-\kappa\Delta}\right)^2\frac{\sigma^2}{2\kappa}\theta\right).$$

The resulting transition density is still nonlinear for two reasons:

1. α_t must be prevented from assuming negative values;

Author	q_1	q_2
Lamoureaux and Witte (2002)	37.15	-0.79
de Jong and Santa Clara (1999)	-0.015	7.29
Duan and Simonato (1999)	13.48	-1.00
Geyer and Pichler (1999)	12.24	0.73

Table 1: Estimated parameters q in the literature on two-factor CIR models. The value of q_2 in Duan and Simonato, which has been rounded to minus one to be displayed in the table, is obviously less than one in absolute value.

2. the conditional variance of the state at time t is a function of its position at time $t-1$.

To overcome these difficulties one can run a modified Kalman filter so as to set the state value equal to zero when the filter gives a negative value and make the covariance matrix of the innovations in the transition equation, $\mathbf{Q}_t$, linear in the *filtered* value of α at time $t-1$.[3]

Under which conditions is the Gaussian density a good approximation of the non-central χ^2? To answer this question recall that the noncentral χ^2 density with ν degrees of freedom and parameter of noncentrality ξ can be thought of as a mixture of central χ^2 densities with Poisson probabilities as weights[4]

$$p_{\chi'^2_{\nu,\xi}}(x) = \sum_{j=0}^{\infty} \left(\frac{(\xi/2)^j}{j!} e^{-\xi/2} \right) p_{\chi^2_{\nu+2j}}(x) .$$

It is clear from (3) that, since $\xi/2 = ce^{-\kappa\Delta}\alpha_{t-1}$, the parameter of the Poisson distribution is a linear function of α_{t-1}, the previous position of the state. This implies that if α_{t-1} is small enough, then the Poisson probabilities will tend to concentrate on the value zero. As a result, the noncentral χ^2 density will approach a central χ^2 with $2q+2$ degrees of freedom. Therefore, for small values of α_{t-1} only if q is large does the Gaussian density give a good approximation. In particular, if $q < 0$ then $2q + 2 < 2$ and the density becomes a monotonically decreasing function that diverges when the argument α_t tends to zero. Clearly, such a function should not be approximated by a bell-shaped normal density function.

The condition $q > 0$ ensures that the factor cannot reach the origin. If we compute the values of q implied by the estimated parameters obtained in the existing literature (as shown in Table 1) we can see immediately that the case $q < 0$ is indeed relevant: one of the factors can typically reach the origin. The only exception is the paper by Geyer and Pichler (1999) but it is worth noting that their analysis of the two-factor CIR model cannot be compared directly to the others because the estimation is carried out on rescaled interest rates.

[3] Lund (1997) gives details of the procedure.

[4] See Johnson and Kotz (1970; ch. 28).

The Monte Carlo experiments carried out in Lund (1997) and Duan and Simonato (1999) do not take this feature of the multi-factor CIR models into account. In fact, Duan and Simonato (1999) *only simulate from one-factor models.* Clearly, in a one-factor model it is not realistic to assume that the single factor, which is also the instantaneous interest rate, can reach the origin. But in a multi-factor model the factors typically have very different statistical properties. Lund (1997) uses a two-factor model which cannot be compared to the ones whose estimates are shown in Table 1 because the specifications are different. Lund's model is a 'double decay' model, i.e. a model in which one factor reverts to a stochastic mean level that evolves as a CIR process, whereas in the model discussed here (like in all the articles quoted in the table) the short rate is the sum of two CIR processes. The reason why neither factor can attain the origin in the 'double decay' case is that both factors can be interpreted as interest rates (a short rate and a long term mean rate) and therefore the estimated parameters reflect the fact that the observed interest rates do not reach the level zero. Despite its useful interpretation, Lund's model is less tractable than the one discussed in this paper as bond prices are not available in closed form.

III Computing the likelihood

A Monte Carlo integration

In the state space form described above it is impossible to write the density of the observations $\mathbf{y}_1, \ldots, \mathbf{y}_T$ directly, because we only make assumptions about the conditional density. Theoretically, one could write down the joint density of $\mathbf{y}_1, \ldots, \mathbf{y}_T$ and $\alpha_1, \ldots, \alpha_T$ and then integrate out the state values:

$$p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\alpha})\, d\boldsymbol{\alpha} \tag{4}$$

where

$$p(\mathbf{y}, \boldsymbol{\alpha}) = p(\mathbf{y}|\boldsymbol{\alpha})\, p(\boldsymbol{\alpha}) = \underbrace{\prod_{t=1}^{T} p(\mathbf{y}_t|\alpha_t)}_{\text{Gaussian}} \quad \underbrace{\prod_{t=2}^{T} p(\alpha_t|\alpha_{t-1})}_{\text{Noncentral } \chi^2} \quad \cdot \underbrace{p(\alpha_1)}_{\text{Gamma}}. \tag{5}$$

In this paper I propose an efficient procedure to compute the above high-dimensional integral.

The first product in (5) can be thought of as a density for α_t, $t = 1, \ldots, T$ multiplied

by a constant. Call

$$\begin{aligned}\phi_t(x) &= p(\mathbf{y}_t|\alpha_t = x) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi h}} \exp\left[-\frac{[y_t(\tau_i) + A(\tau_i) - B(\tau_i)x]^2}{2h}\right] \\ &= \frac{1}{(2\pi h)^{N/2}} \exp\left(-ax^2 - b_t x - c_t\right)\end{aligned}$$

where

$$a \equiv \frac{1}{2h}\sum_{i=1}^{N} B^2(\tau_i), \qquad b_t \equiv -\frac{1}{h}\sum_{i=1}^{N}(y_t(\tau_i) - A(\tau_i))B(\tau_i)$$

and

$$c_t \equiv \frac{1}{2h}\sum_{i=1}^{N}(y_t(\tau_i) - A(\tau_i))^2$$

Then the importance density can be chosen to be

$$f(\alpha_t) = \phi_t(\alpha_t) / \int_0^\infty \phi_t(x)\,dx.$$

Since a is a sum of squares, the density $f()$ exists and is symmetric around the mean $-b_t/(2a)$, i.e. the OLS estimator of α_t computed from the cross section of observations at time t. Note that the random draws $\alpha_t^{(i)}$ are mutually independent by construction. The choice of this importance density is motivated by the informativeness of the observation distribution $p(y|\alpha)$, which is typically very concentrated around the mean. The parameter h, i.e. the measurement error variance, plays an important role in determining the shape of $f()$ as it appears in the denominator of a. Small values of h result in very concentrated densities. For example, de Jong and Santa Clara (1999) obtain an estimate of 4 basis points for $\sqrt{h}$ in their two-factor CIR model.

Thus the integral (4) becomes:

$$\begin{aligned}p(\mathbf{y}) &= \int_0^\infty \prod_{t=1}^{T} \phi_t(\alpha_t) \cdot p(\alpha_1) \prod_{t=2}^{T} p(\alpha_t|\alpha_{t-1})\,d\boldsymbol{\alpha} \\ &\propto E_f\left(p(\alpha_1)\prod_{t=2}^{T} p(\alpha_t|\alpha_{t-1})\right) \qquad (6)\end{aligned}$$

where E_f denotes expectation with respect to the joint distribution $\prod f(\alpha_t)$. The constant of proportionality is given by the product of integrals

$$\prod_{t=1}^{T}\int_0^\infty \phi_t(x)\,dx.$$

The expectation (6) can be estimated as

$$\frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}}\left[p\left(\alpha_1^{(i)}\right)\prod_{t=2}^{T}p\left(\alpha_t^{(i)}|\alpha_{t-1}^{(i)}\right)\right]\equiv\frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}}w^{(i)} \tag{7}$$

where $\alpha_t^{(i)}$ is a draw from $f(\alpha_t)$.

B Particle filtering

Unfortunately, for realistic sample sizes, the procedure described above suffers from a problematic degeneracy: one of the weights $w^{(i)}$ typically dominates the sum (7), so that the others have no effect on the estimate. This problem can be solved by using importance sampling with sequential resampling. A good introduction to sequential Monte Carlo methods can be found in Doucet, de Freitas and Gordon (2001).

Rewrite the joint likelihood $p(\mathbf{y})$ as

$$p(\mathbf{y})=p(\mathbf{y}_1)\prod_{t=2}^{T}p(\mathbf{y}_t|\mathbf{Y}_{t-1}).$$

Each density will be estimated by Monte Carlo. Firstly, $p(\mathbf{y}_1)$ can be estimated as

$$\frac{K_1}{\overline{n}}\sum_{i=1}^{\overline{n}}p\left(\alpha_1^{(i)}\right)$$

where $\alpha_1^{(i)}$ is drawn from $f(\alpha_1)$ and

$$K_t\equiv\int_0^{\infty}\phi_t(x)\,dx.$$

Let

$$w_1^{(i)}\equiv K_1\ p\left(\alpha_1^{(i)}\right)$$

and

$$w_t^{(i)}\equiv K_1\ p\left(\alpha_1^{(i)}\right)\prod_{s=2}^{t}\frac{K_s\ p\left(\alpha_s^{(i)}|\alpha_{s-1}^{(i)}\right)}{\sum_{j=1}^{\overline{n}}w_{s-1}^{(j)}}=K_t\ p\left(\alpha_t^{(i)}|\alpha_{t-1}^{(i)}\right)\frac{w_{t-1}^{(i)}}{\sum_{j=1}^{\overline{n}}w_{t-1}^{(j)}},\quad t=2,\ldots,T.$$

Thus we have that the densities $p\left(\mathbf{y}_t|\mathbf{Y}_{t-1}\right)$, for $t = 2, \ldots, T$, can be estimated as[5]

$$\sum_{i=1}^{\overline{n}} w_t^{(i)}.$$

The sequential nature of the procedure allows us to check the accuracy of the approximation step by step, i.e. at each point in time t. The statistic

$$\widetilde{n} = \frac{\left(\sum_{i=1}^{\overline{n}} w_t^{(i)}\right)^2}{\sum_{i=1}^{\overline{n}} \left(w_t^{(i)}\right)^2}$$

can be used to detect the presence of degeneracies. When all weights are equal, and therefore we are not wasting any draw, $\widetilde{n} = \overline{n}$. When the degeneracy is extreme, i.e. all weights are equal to zero except one, $\widetilde{n} = 1$. At each step I compute the statistic and resample if it falls below the threshold set at $\overline{n}/2$. The resampling algorithm used here

Importance sampling/sequential resampling Monte Carlo filter

1. Sample $\alpha_1^{(i)}$ from $f\left(\alpha_1\right)$ for $i = 1, \ldots, \bar{n}$. Compute the weights $w_1^{(i)}$.
2. For $t = 2, \ldots, T$ sample $\alpha_t^{(i)}$ from $f\left(\alpha_t\right)$ for $i = 1, \ldots, \bar{n}$. Compute the weights $w_t^{(i)}$ recursively from $w_{t-1}^{(i)}$.
3. Given all the weights for time t, compute the statistic $\widetilde{n}$. If $\widetilde{n} < \bar{n}/2$ then resample.
4. Go back to point 2.

is the systematic resampling described by Kitagawa (1996). After the resampling stage the weight of each particle becomes

$$\widehat{w}_t^{(i)} = \frac{\sum_{i=1}^{\overline{n}} w_t^{(i)}}{\overline{n}}.$$

The advantage of this residual resampling scheme on, for example, multinomial resampling is that it minimises the variance of the number of replicates of each particle in the old sample that will appear in the new one.

[5]Doucet, Godsill and Andrieu (2000) and Hürzeler and Künsch (2001) describe alternative estimators that are more efficient, especially when the resampling procedure is used many times. In this case their procedures would involve taking draws from a noncentral χ^2 distribution to generate a *prediction sample* at each iteration.

Resampling algorithm

First, compute the cumulative distribution function of the normalised weights $w_t^{(i)} / \sum_{i=1}^{\overline{n}} w_t^{(i)}$:

$$cdf_i = \begin{cases} 0 & i = 0 \\ cdf_{i-1} + w_t^{(i)} / \sum_{i=1}^{\overline{n}} w_t^{(i)} & i = 1, \ldots, \overline{n} \end{cases}$$

Then the cumulative distribution function for the new sample of particles, $\widehat{cdf}_i$, is obtained as follows. First, draw $\widehat{cdf}_1$ from a uniform distribution with support $(0, 1/\overline{n})$ and set $\widehat{cdf}_i = \widehat{cdf}_{i-1} + 1/\overline{n},\ i = 2, \ldots, \overline{n}$ Furthermore, let $j = 1$. Then, starting with $i = 1$

- If $cdf_j \geq \widehat{cdf}_i$ then set the i-th particle in the new sample equal to the j-th particle in the old one, i.e.

  $$\left(\widehat{\alpha}_1^{(i)}, \ldots, \widehat{\alpha}_t^{(i)}\right) = \left(\alpha_1^{(j)}, \ldots, \alpha_t^{(j)}\right)$$

  Increase i and check the inequality again.

- If $cdf_j < \widehat{cdf}_i$ then increase j and check the inequality again.

The two steps are repeated until all the new particles $\left(\widehat{\alpha}_1^{(i)}, \ldots, \widehat{\alpha}_t^{(i)}\right)$ have been obtained.

IV Filtering and smoothing

Monte Carlo integration can also be used to obtain filtered and smoothed values of the state. Firstly, suppose that we are interested in the expectation $E(\alpha_t|\mathbf{Y}_t)$, i.e. the expectation of the position of the state at time t given the observations up to time t. Then we have

$$\begin{aligned} E(\alpha_t|\mathbf{Y}_t) &= \int \alpha_t\ p(\alpha_t|\mathbf{Y}_t)\, d\alpha_t \\ &= \int \alpha_t\ p(\alpha_t, \mathbf{y}_1, \ldots, \mathbf{y}_t) / p(\mathbf{y}_1, \ldots, \mathbf{y}_t)\, d\alpha_t \\ &= \int \alpha_t\ \frac{\int p(\alpha_1, \ldots, \alpha_t, \mathbf{y}_1, \ldots, \mathbf{y}_t)\, d\alpha_{t-1} \ldots d\alpha_1}{\int p(\alpha_1, \ldots, \alpha_t, \mathbf{y}_1, \ldots, \mathbf{y}_t)\, d\alpha_t \ldots d\alpha_1} d\alpha_t \\ &= \frac{\int \alpha_t\ p(\alpha_1, \ldots, \alpha_t, \mathbf{y}_1, \ldots, \mathbf{y}_t)\, d\alpha_t \ldots d\alpha_1}{\int p(\alpha_1, \ldots, \alpha_t, \mathbf{y}_1, \ldots, \mathbf{y}_t)\, d\alpha_t \ldots d\alpha_1}. \end{aligned} \tag{8}$$

The last expression in (8) can be estimated using

$$\frac{\sum_{i=1}^{\overline{n}}\left[\alpha_t^{(i)}\ p\left(\alpha_1^{(i)}\right)\prod_{s=2}^{t}p\left(\alpha_s^{(i)}|\alpha_{s-1}^{(i)}\right)\right]}{\sum_{i=1}^{\overline{n}}\left[p\left(\alpha_1^{(i)}\right)\prod_{s=2}^{t}p\left(\alpha_s^{(i)}|\alpha_{s-1}^{(i)}\right)\right]}=\frac{\sum_{i=1}^{\overline{n}}\alpha_t^{(i)}\ w_t^{(i)}}{\sum_{i=1}^{\overline{n}}w_t^{(i)}},\quad t=2,\ldots,T \tag{9}$$

where again $\alpha_t^{(i)}$ is a draw from $f(\alpha_t)$.[6] Note that the sequential procedure can be carried out in a computationally efficient way if we store at each step the values $w_t^{(i)}$. This allows us to draw only one additional set of $\overline{n}$ values for the state, i.e. $\alpha_{t+1}^{(i)}$, $i=1,\ldots,\overline{n}$, from $f(\alpha_{t+1})$ and set

$$w_{t+1}^{(i)}=p\left(\alpha_{t+1}^{(i)}|\alpha_t^{(i)}\right)w_t^{(i)}.$$

A similar device can be used to obtain the smoothed estimates of α. Let

$$\begin{aligned}
E(\alpha_t|\mathbf{Y}_T) &= \int \alpha_t\ p(\alpha_t|\mathbf{Y}_T)\,d\alpha_t \\
&= \int \alpha_t\ p(\alpha_t,\mathbf{y}_1,\ldots,\mathbf{y}_T)/p(\mathbf{y}_1,\ldots,\mathbf{y}_T)\,d\alpha_t \\
&= \int \alpha_t\ \frac{\int p(\alpha_1,\ldots,\alpha_T,\mathbf{y}_1,\ldots,\mathbf{y}_T)\,d\alpha_T\ldots d\alpha_{t+1}d\alpha_{t-1}\ldots d\alpha_1}{\int p(\alpha_1,\ldots,\alpha_T,\mathbf{y}_1,\ldots,\mathbf{y}_t)\,d\alpha_T\ldots d\alpha_1}d\alpha_t \\
&= \frac{\int \alpha_t\ p(\alpha_1,\ldots,\alpha_T,\mathbf{y}_1,\ldots,\mathbf{y}_T)\,d\alpha_T\ldots d\alpha_1}{\int p(\alpha_1,\ldots,\alpha_T,\mathbf{y}_1,\ldots,\mathbf{y}_t)\,d\alpha_T\ldots d\alpha_1}.
\end{aligned} \tag{10}$$

The last expression in (10) can be estimated using

$$\frac{\sum_{i=1}^{\overline{n}}\left[\alpha_t^{(i)}\ p\left(\alpha_1^{(i)}\right)\prod_{s=2}^{T}p\left(\alpha_s^{(i)}|\alpha_{s-1}^{(i)}\right)\right]}{\sum_{i=1}^{\overline{n}}\left[p\left(\alpha_1^{(i)}\right)\prod_{s=2}^{T}p\left(\alpha_s^{(i)}|\alpha_{s-1}^{(i)}\right)\right]}=\frac{\sum_{i=1}^{\overline{n}}\alpha_t^{(i)}\ w_T^{(i)}}{\sum_{i=1}^{\overline{n}}w_T^{(i)}},\quad t=2,\ldots,T \tag{11}$$

where again $\alpha_t^{(i)}$ is a draw from $f(\alpha_t)$. Note that the filter and the smoother can be applied together in a computationally efficient way using the weights $w_T^{(i)}$ that have been computed in the last step of the filter to compute the smoothed estimates backward. To implement this procedure, one only needs to store all the draws $\alpha_t^{(i)}$ while computing the filtered estimates and use them to compute the quantity (11) for $t=T,\ldots,2$.

[6] Several refinements of the estimators (9) and (11) have been proposed in the literature, for example by Doucet, Godsill and Andrieu (2000).

V Generation of pseudo-random numbers

This section describes the pseudo-random number generator used to draw from $f(\alpha_t)$. Recall that

$$f(\alpha_t) = \phi_t(\alpha_t) / \int_0^\infty \phi_t(x)\, dx \tag{12}$$

where

$$\phi_t(x) = (2\pi h)^{-N/2} \exp\left(-ax^2 - b_t x - c_t\right)$$

and the values a, b_t and c_t are functions of the model's parameters. A draw from $f(\alpha_t)$ can be obtained in two stages using the algorithm described below. The Newton method

Algorithm to draw from the importance density

1. Draw a quantity δ from the uniform distribution.
2. Take $\widetilde{\alpha}_t = F^{-1}(\delta)$ as the draw from $f(\alpha_t)$. $F(x)$ is the cumulative density function associated with $f(\alpha_t)$, i.e.

$$F(x) = \int_0^x f(u)\, du$$

can be used to invert F, as if we were to solve the problem

$$\min_x \left(F(x) - \delta\right)^2$$

Start with an initial value $\widetilde{\alpha}_t^0 = 0.5$. Then iterate

$$\begin{aligned}
\widetilde{\alpha}_t^{n+1} &= \widetilde{\alpha}_t^n - \left(\frac{d^2}{dx^2}\left(F(x) - \delta\right)^2\right)^{-1} \frac{d}{dx}\left(F(x) - \delta\right)^2 \\
&= \widetilde{\alpha}_t^n - \frac{\int_0^{\widetilde{\alpha}_t^n} \phi_t(u)\, du - \delta^*}{\left(1 + (2\pi h)^{N/2}\left(\int_0^{\widetilde{\alpha}_t^n} \phi_t(u)\, du - \delta^*\right)\right)\phi\left(\widetilde{\alpha}_t^n\right)}
\end{aligned}$$

where $\delta^* \equiv \delta K_t = \delta \int_0^\infty \phi_t(x)\, dx$. If regularity conditions apply, the sequence $\widetilde{\alpha}_t^n$ converges to a draw from $f(\alpha_t)$. In practice, I have iterated (13) until the second term in the difference reached $10^{-6}/(2c)$. Convergence usually occurs after a few iterations.

The integrals of $\phi_t()$ are computed numerically using the Simpson method. This can be done efficiently by storing the value of the integral from zero to the maximum point $-b_t/(2a)$ after the computation of K_t and then adding the integral from the maximum point to x each time a new integral of the form (12) has to be calculated.

It is worth noting that the procedure lends itself naturally to a bivariate generalisation. In a two-factor CIR model the exponentials in the Gaussian densities are linear in

κ	θ	σ	λ/σ^2
0.1862	0.0654	0.0481	-32.03

Table 2: CIR parameters as estimated by de Jong and Santa Clara (1999).

both factors, so that the function ϕ becomes:

$$\begin{aligned}\phi_t(x,y) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi h}} \exp\left[-\frac{\left[y_t(\tau_i) + A(\tau_i) - B_x(\tau_i)\, x - B_y(\tau_i)\, y\right]^2}{2h}\right] \\ &= \frac{1}{(2\pi h)^{N/2}} \exp\left(-a_x x^2 - a_y y^2 - a_{t,xy} xy - b_{t,x} x - b_{t,y} y - c_t\right)\end{aligned}$$

where the quantities A, B_x and B_y are known functions of the parameters. If we had to draw from $f(\alpha_{1,t}, \alpha_{2,t})$ we could first fix $\alpha_{1,t}$ at the level $\widehat{\alpha}_{1,t}$ given by the OLS estimator of $\alpha_{1,t}$ and draw $\alpha_{2,t}^{(i)}$ from the univariate density $f(\alpha_{2,t}|\alpha_{1,t} = \widehat{\alpha}_{1,t})$. Then $\alpha_{1,t}^{(i)}$ could be drawn from the univariate conditional density $f\left(\alpha_{1,t}, \alpha_{2,t}|\alpha_{2,t} = \alpha_{2,t}^{(i)}\right)$ which can be written as a univariate density of the form of $f()$. As a result, the procedure would involve taking two draws from univariate densities.

VI Results for simulated data

A sample of $T = 250$ weekly observations for the CIR model has been simulated using the parameters estimated by de Jong and Santa Clara (1999) (shown in Table 2) and a two step procedure described by Duan and Simonato (1999). Sequential draws from the Poisson distribution simulate the degrees of freedom of the central chi squared from which the state variables are drawn. Measurement errors with a standard deviation of 10 basis points are then added to form the observations for eight ($N = 8$) maturities: 0.25, 0.5, 1, 2, 3, 5, 7 and 10 years. The methodology described above has been used to compute the likelihood function from the simulated data at the true parameter value. The computer code was written in Ox 3.0 (see Doornik (2002)) and C. To assess the accuracy of the estimator, the procedure has been repeated 100 times on the same simulated sample with different sets of pseudo-random numbers in the Monte Carlo integration part . Furthermore, the whole exercise has been carried out for three different values of $\bar{n}$: 100, 200 and 500.

The results are presented in Table 3 and Figure 1. The number of resampling stages used in the procedure is relatively small (below 13% of the sample size T), not significantly larger, for example, than the corresponding figure reported in Doucet, Godsill and Andrieu (2000) who use a less efficient resampling scheme and a lower threshold ($\bar{n}/3$). It turns out that the parameter h affects the number of times resampling is performed. In particular, large values of h make the observation density $p(y|\alpha)$ less informative and tend to increase the number of resampling stages. As for the computation times, it takes 2.2 seconds to compute the likelihood with $\bar{n} = 100$ on a Pentium 4 PC, which is

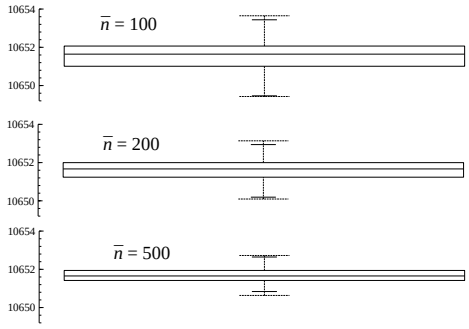

Figure 1: Each box plot represents the distribution of 100 estimates of the log-likelihood of the same simulated sample with $T = 250$ and $N = 8$ different interest rates. The number of particles $\bar{n}$ used in each exercise is respectively 100, 200 and 500.

$\bar{n}$	RS	Time	S.d.
100	31	2.2	0.744
200	32	4.2	0.474
500	32	10.8	0.359

Table 3: Descriptive statistics for 100 estimates (each obtained for a different seed of the random generator) of the log-likelihood of the same simulated sample for three different choices of $\bar{n}$, the number of particles. RS is the average number of resamplings, time is the average computation time in seconds and S.d. is the standard deviation of the 100 estimates.

a satisfactory result. Using more particles, the computation time becomes 4.2 seconds for $\bar{n} = 200$ and 10.8 seconds for $\bar{n} = 500$. Both the standard deviations in Table 3 and the box plots in Figure 1 show that increasing the simulation size $\bar{n}$ results in a more accurate estimation. The distribution of the values of the statistic clearly tends to gradually concentrate around the true value as $\bar{n}$ increases.

To assess the accuracy of the maximum likelihood estimator proposed in the paper I have estimated the CIR parameters κ, λ, σ, θ and h from each of the 100 simulated samples. The numerical maximisation is a difficult task in this case because the usual methods that require the calculation of numerical first derivatives fail. This is due to the fact that the errors introduced by the simulation dominate when one tries to measure the effect of small changes in the value of a parameter and as a result any iterative optimisation procedure becomes unstable. Fixing the set of pseudo-random numbers used in the Monte Carlo integration (so that the same draws are used each time the calculation is performed) does not solve the problem, because resampling introduces discontinuities in the approximate likelihood function. The solution adopted here is the use of a grid search. Starting with a suitable initial value, I have calculated the likelihood function in the neighborhood of the initial point. The procedure is repeated, iteration

by iteration, for the point giving the largest value of the statistic, until a maximum point is reached. I have tried several initial points to ensure convergence to a global maximum.[7] It should be noted that, in practice, I used the logarithm of the likelihood function, thereby introducing a bias in the estimation.

The results are summarised in Figure 2. Each panel reports two histograms that represent the empirical distribution of the estimated values of four parameters: κ (a), λ/σ^2 (b), σ (c) and θ (d). The histogram at the top of each panel shows the estimates obtained with 100 particles, the one at the bottom illustrates the results for $\bar{n} = 200$. As for the parameter h, using a grid with increments of one basis point I obtained an estimate equal to the true value (0.001 i.e. 10 basis points) in all cases. For all four parameters the distributions are bell-shaped and peak at the grid point that is nearest to the true value. This result indicates that even using a moderate simulation size ($\bar{n} = 100$) it is possible to obtain a relatively accurate estimate of the parameter value. If we compare for each parameter the two sets of estimates we can conclude that doubling the simulation size does not produce significant changes in the empirical distribution of the estimates. The only effect seems to be that the values of $\hat{\theta}$ are slightly more concentrated around the true value. We can conclude that the errors in the estimation of the likelihood that are caused by simulation do not seem to undermine the estimation procedure.

VII Conclusion

The contribution of this paper is twofold.

Firstly, I show that the QML procedure, used in many papers in the current literature to estimate the CIR model from time series data, is based on an approximation of the latent factors' density that becomes very inaccurate for typical parameter values. I also argue that this issue is not addressed by the Monte Carlo experiments carried out to support the conclusion that the QML bias is negligible.

Secondly, I show how to build in a computationally efficient way a maximum likelihood estimator based on simulation. The advantage of this estimator is to take into account the exact likelihood function while avoiding the huge computational burden associated with MCMC methods. A simulation based filter and smoother is also derived to compute conditional moments of the latent factors. The proposed methodology is implemented and tested for the single factor CIR model on a sample of simulated data. The Monte Carlo procedure is stable and convergence to the true value of the likelihood is relatively fast. Moreover, I use a grid search to maximise the likelihood function with respect to the CIR parameters. The estimator seems to have good properties even using moderate simulation sizes (100 particles for a sample of 250 observations of eight interest rates). In particular, the effect of simulation errors does not seem to undermine the estimation procedure.

Future work will be focussed on the estimation of the two-factor version of the CIR

[7] A more sophisticated procedure is proposed by Hürzeler and Künsch (2001) who use a small number of sets of particles to compute local approximations of the likelihood functions and then optimise.

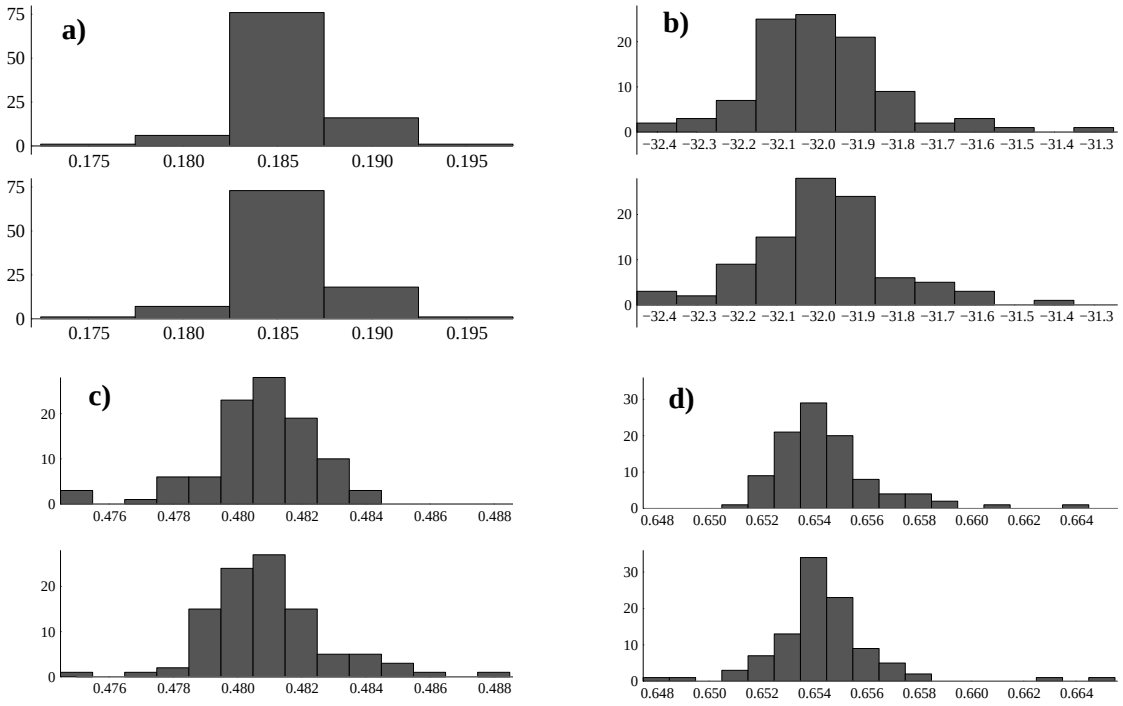

Figure 2: Distribution of the estimators calculated using a grid search from 100 simulated samples for a number of particles $\bar{n} = 100$ and 200. Panel a) plots the empirical distribution of $\hat{\kappa}$. Panels b) to d) plot the distributions of $\hat{\lambda/\sigma^2}$, $\hat{\sigma}$ and $10\hat{\theta}$ respectively. The histogram at the top of each panel shows the results for $\bar{n} = 100$, the one at the bottom shows the results for $\bar{n} = 200$. The true parameter values are respectively 0.1862, -32.03, 0.481, 0.654.

model, using a natural generalisation of the univariate procedure. It is also possible to explore ways to improve the computational efficiency of the procedure, in particular with more efficient numerical integration algorithms.

References

Bakshi, S. B., and Chen, Z. (1997): "An alternative valuation model for contingent claims", *Journal of Financial Economics*, 44, 123–165.

Cox, J. C., Ingersoll, J. E., and Ross, S. (1985): "A theory of the term structure of interest rates", *Econometrica*, 53, 385–407.

Dai, Q., and Singleton, K. J. (2000): "Specification Analysis of Affine Term Structure Models", *Journal of Finance*, 55, 1943–1978.

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

de Jong, F., and Santa-Clara, P. (1999): "The dynamics of the forward interest rate curve: A formulation with state variables", *Journal of Financial and Quantitative Analysis*, 34, 131–157.

Doornik, J. A. (2002): *Object-Oriented Matrix Programming Using Ox.* London: Timberlake Consultants Press.

Doucet, A., de Freitas, N., and Gordon, N. (2001): "An introduction to sequential Monte Carlo methods", in: Doucet, A., de Freitas, N., and Gordon, N. (eds.), *Sequential Monte Carlo Methods in Practice.* New York: Springer.

Doucet, A., Godsill, S., and Andrieu, C. (2000): "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing*, 10, 197–208.

Duan, J. C., and Simonato, J. G. (1999): "Estimating and testing exponential-affine term structure models by Kalman filter", *Review of Quantitative Finance and Accounting*, 13, 102–127.

Duffee, G. R. (1999): "Estimating the price of default risk", *Review of Financial Studies*, 12, 197–226.

Durham, G. B., and Gallant, A. R. (2002): "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes", *Journal of Business and Economic Statistics*, 20, 325–327.

Geyer, A. L. J., and Pichler, S. (1999): "A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure", *Journal of Financial Research*, 23, 107–130.

Hull, J., and White, A. (1990): "Pricing interest rate derivative securities", *Review of Financial Studies*, 3, 573–92.

Hürzeler, M., and Künsch, H. R. (2001): "Approximating and maximising the likelihood for a general state-space model", in: Doucet, A., de Freitas, N., and Gordon, N. (eds.), *Sequential Monte Carlo Methods in Practice.* New York: Springer.

Johnson, N. L., and Kotz, S. (1970): *Distributions in statistics: continuous univariate distributions.* Boston: Houghton Miffin Company.

Kitagawa, G. (1996): "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", *Journal of Computational and Graphical Statistics*, 5, 1–25.

Lamoureaux, C., and Witte, H. D. (2002): "Empirical analysis of the yield curve: the information in the data viewed through the window of Cox, Ingersoll and Ross", *Journal of Finance*, 57, 1479–1520.

Longstaff, F. A., Schwartz, S. (1992): "Interest rate volatility and the term structure: a two factor general equilibrium model", *Journal of Finance*, 47, 1259–1282.

Lund, J. (1997): "Econometric analysis of continuous-time arbitrage-free models of the term structure of interest rates", Working Paper, Aarhus School of Business.