MAXIMUM LIKELIHOOD ESTIMATION OF THE COX-INGERSOLL-ROSS PROCESS: THE MATLAB IMPLEMENTATION

Kamil Kladívko[1]

Department of Statistics and Probability Calculus, University of Economics, Prague
and Debt Management Department, Ministry of Finance of the Czech Republic
`kladivk@vse.cz` or `kamil.kladivko@mfcr.cz`

Abstract

The square root diffusion process is widely used for modeling interest rates behaviour. It is an underlying process of the well-known Cox-Ingersoll-Ross term structure model (1985). We investigate maximum likelihood estimation of the square root process (CIR process) for interest rate time series. The MATLAB implementation of the estimation routine is provided and tested on the PRIBOR 3M time series.

1 CIR Process for Interest Rate Modeling

A continuous-time model in finance typically rest on one or more stationary diffusion processes $\{X_t, t \geq 0\}$, with dynamics represented by stochastic differential equations:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{1}$$

where $\{W_t, t \geq 0\}$ is a standard Brownian motion. The functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ are, respectively, the drift and the diffusion functions of the process.

The fundamental process in interest rate modeling is the square root process given by the following stochastic differential equation:

$$dr_t = \alpha(\mu - r_t)dt + \sqrt{r_t}\sigma dW_t, \tag{2}$$

where r_t is the interest rate and $\theta \equiv (\alpha, \mu, \sigma)$ are model parameters. The drift function $\mu(r_t, \theta) = \alpha(\mu - r_t)$ is linear and possess a mean reverting property, i.e. interest rate r_t moves in the direction of its mean μ at speed α. The diffusion function $\sigma^2(r_t, \theta) = r_t\sigma^2$ is proportional to the interest rate r_t and ensures that the process stays on a positive domain.

The square root process (2) is the basis for the Cox, Ingersoll, and Ross short-term interest rate model [1] and therefore often denoted as the CIR process in the financial literature.

1.1 CIR process densities

If α, μ, σ are all positive and $2\alpha\mu \geq \sigma^2$ holds, the CIR process is well-defined and has a steady state (marginal) distribution. The marginal density is gamma distributed.

For maximum likelihood estimation of the parameter vector $\theta \equiv (\alpha, \mu, \sigma)$ transition densities are required. The CIR process is one of few cases, among the diffusion processes, where the transition density has a closed form expression. We follow the notation given in [1] on page 391. Given r_t at time t the density of $r_{t+\Delta t}$ at time $t + \Delta t$ is

$$p(r_{t+\Delta t}|r_t; \theta, \Delta t) = ce^{-u-v}(\frac{v}{u})^{\frac{q}{2}} I_q(2\sqrt{uv}), \tag{3}$$

[1] This work was supported by grant of IGA VSE nr. IG410046

where

$$\begin{aligned} c &= \frac{2\alpha}{\sigma^2(1-e^{-\alpha\Delta t})}, \\ u &= cr_t e^{-\alpha\Delta t}, \\ v &= cr_{t+\Delta t}, \\ q &= \frac{2\alpha\mu}{\sigma^2} - 1, \end{aligned}$$

and $I_q(2\sqrt{uv})$ is modified Bessel function of the first kind and of order q. The transition density (3) has been originally derived in [2].

Sometimes, it is particularly useful to work with a transformation $s_{t+\Delta t} = 2cr_{t+\Delta t}$. We can easily derive that the transition density of $s_{t+\Delta t}$ is

$$g(s_{t+\Delta t}|s_t;\theta,\Delta t) = g(2cr_{t+\Delta t}|2cr_t;\theta,\Delta t) = \frac{1}{2c}p(r_{t+\Delta t}|r_t;\theta,\Delta t), \tag{4}$$

which is the the noncentral χ^2 distribution with $2q+2$ degrees of freedom and non-centrality parameter $2u$.

2 Maximum Likelihood Implementation in MATLAB

Parameter estimation is carried out on interest rate time series with N observations $\{r_{t_i}, i = 1 \ldots N\}$. We consider equally spaced observations with Δt time step.

2.1 Likelihood function

The likelihood function for interest rate time series with N observations is

$$L(\theta) = \prod_{i=1}^{N-1} p(r_{t_{i+1}}|r_{t_i};\theta,\Delta t). \tag{5}$$

It is computationally convenient to work with the log-likelihood function

$$\ln L(\theta) = \sum_{i=1}^{N-1} \ln p(r_{t_{i+1}}|r_{t_i};\theta,\Delta t), \tag{6}$$

from which we easily derive the log-likelihood function of the CIR process

$$\ln L(\theta) = (N-1)\ln c + \sum_{i=1}^{N-1}\left\{-u_{t_i} - v_{t_{i+1}} + 0.5q\ln\left(\frac{v_{t_{i+1}}}{u_{t_i}}\right) + \ln\{I_q(2\sqrt{u_{t_i}v_{t_{i+1}}})\}\right\}, \tag{7}$$

where $u_{t_i} = cr_{t_i}e^{-\alpha\Delta t}$ and $v_{t_{i+1}} = cr_{t_{i+1}}$. We find maximum likelihood estimates $\hat{\theta}$ of parameter vector θ by maximizing the log-likelihood function (7) over its parameter space:

$$\hat{\theta} \equiv (\hat{\alpha}, \hat{\mu}, \hat{\sigma}) = \arg\max_{\theta} \ln L(\theta). \tag{8}$$

Since the logarithmic function is monotonically increasing, maximizing the log-likelihood function also maximizes the likelihood function.

For solving optimization problem (8) we have to rely on numerical solution. The function `fminsearch`, which is a standard part of MATLAB, makes the job. The function `fminsearch` is an implementation of Nelder-Mead simplex method, see MATLAB help for details.

2.2 Initial estimates

For convergence to the global optimum initial (starting) points of optimization are crucial. We suggest to use Ordinary Least Squares (OLS) on discretized version of (2):

$$r_{t+\Delta t} - r_t = \alpha(\mu - r_t)\Delta t + \sigma\sqrt{r_t}\varepsilon_t, \tag{9}$$

where ε_t is normally distributed with zero mean and variance Δt, more precisely ε_t is a white noise process. For performing OLS we transform (9) into:

$$\frac{r_{t+\Delta t} - r_t}{\sqrt{r_t}} = \frac{\alpha\mu\Delta t}{\sqrt{r_t}} - \alpha\sqrt{r_t}\Delta t + \sigma\varepsilon_t. \tag{10}$$

The drift initial estimates are found by minimizing the OLS objective function

$$(\widehat{\alpha}, \widehat{\mu}) = \arg\min_{\alpha,\mu} \sum_{i=1}^{N-1} \left(\frac{r_{t_{i+1}} - r_{t_i}}{\sqrt{r_{t_i}}} - \frac{\alpha\mu\Delta t}{\sqrt{r_{t_i}}} + \alpha\sqrt{r_{t_i}}\Delta t \right)^2, \tag{11}$$

which is solved by

$$\widehat{\alpha} = \frac{N^2 - 2N + 1 + \sum\limits_{i=1}^{N-1} r_{t_{i+1}} \sum\limits_{i=1}^{N-1} \frac{1}{r_{t_i}} - \sum\limits_{i=1}^{N-1} r_{t_i} \sum\limits_{i=1}^{N-1} \frac{1}{r_{t_i}} - (N-1) \sum\limits_{i=1}^{N-1} \frac{r_{t_{i+1}}}{r_{t_i}}}{\left(N^2 - 2N + 1 - \sum\limits_{i=1}^{N-1} r_{t_i} \sum\limits_{i=1}^{N-1} \frac{1}{r_{t_i}} \right) \Delta t}, \tag{12}$$

$$\widehat{\mu} = \frac{(N-1) \sum\limits_{i=1}^{N-1} r_{t_{i+1}} - \sum\limits_{i=1}^{N-1} \frac{r_{t_{i+1}}}{r_{t_i}} \sum\limits_{i=1}^{N-1} r_{t_i}}{N^2 - 2N + 1 + \sum\limits_{i=1}^{N-1} r_{t_{i+1}} \sum\limits_{i=1}^{N-1} \frac{1}{r_{t_i}} - \sum\limits_{i=1}^{N-1} r_{t_i} \sum\limits_{i=1}^{N-1} \frac{1}{r_{t_i}} - (N-1) \sum\limits_{i=1}^{N-1} \frac{r_{t_{i+1}}}{r_{t_i}}}. \tag{13}$$

The diffusion parameter initial estimate $\widehat{\sigma}$ is found as a standard deviation of residuals.

The practical MATLAB implementation can, of course, rely on a back slash operator (\) which automatically performs OLS and yields the same results as (12) and (13):

```
% CIR initial parameters estimation
x = Model.Data(1:end-1); % Time series of interest rates observations
dx = diff(Model.Data);
dx = dx./x.^0.5;
regressors = [Model.TimeStep./x.^0.5, Model.TimeStep*x.^0.5];
drift = regressors\dx; % OLS regressors coefficients estimates
res = regressors*drift - dx;
alpha = -drift(2);
mu = -drift(1)/drift(2);
sigma = sqrt(var(res, 1)/Model.TimeStep);
InitialParams = [alpha mu sigma]; % Vector of initial parameters
```

The initial estimates $(\widehat{\alpha}, \widehat{\mu}, \widehat{\sigma})$ are starting points for maximizing the log-likelihood function (7)

2.3 Implementing the log-likelihood function using the command `besseli`

For implementing the objective function (7) we need to evaluate the modified Bessel function of the first kind $I_q(2\sqrt{uv})$. The Bessel function $I_q(2\sqrt{uv})$ is available under the command `besseli(q, 2*sqrt(u.*v))` in MATLAB but direct usage of this function results in an estimation failure. The reason is that the Bessel function $I_q(2\sqrt{uv})$ approaches rapidly to the plus infinity and optimization routines (e.g. fminsearch, fmincon) are not able to handle this problem.[2]

Fortunately there is a scaled version of the Bessel function in MATLAB which we denote $I_q^1(2\sqrt{uv})$ and is available under the command besseli(q,2*sqrt(u.*v),1). The $I_q^1(2\sqrt{uv})$ equals $I_q(2\sqrt{uv})\exp(-2\sqrt{uv})$ and therefore solves the problem of the rapid divergence. We accordingly adjust the objective function (7) into

$$\ln L = (N-1)\ln c + \sum_{i=1}^{N-1}\left\{-u_{t_i} - v_{t_{i+1}} + 0.5q\ln\left(\frac{v_{t_{i+1}}}{u_{t_i}}\right) + \ln\{I_q^1(2\sqrt{u_{t_i}v_{t_{i+1}}})\} + 2\sqrt{u_{t_i}v_{t_{i+1}}}\right\}. \tag{14}$$

The MATLAB code of the log-likelihood function can look as follows:

```
function lnL = CIRobjective1(Params, Model)
% =========================================================================
% PURPOSE : Log-likelihood objective function (multiplied by -1) for the
%           CIR process using the MATLAB besseli function
% =========================================================================
% USAGE   : Model.TimeStep  = Delta t
%           Model.Data      = Time series of interest rates observations
%           Params          = Model parameters (alpha, mu, sigma)
% =========================================================================
% RETURNS : lnL             = Objective function value
% =========================================================================
    Data = Model.Data;
    DataF = Data(2:end);
    DataL = Data(1:end-1);
    Nobs = length(Data);
    TimeStep = Model.TimeStep;
    alpha = Params(1);
    mu = Params(2);
    sigma = Params(3);

    c = 2*alpha/(sigma^2*(1-exp(-alpha*TimeStep)));
    q = 2*alpha*mu/sigma^2-1;
    u = c*exp(-alpha*TimeStep)*DataL;
    v = c*DataF;
    z = 2*sqrt(u.*v);
    bf = besseli(q,z,1);
    lnL= -(Nobs-1)*log(c) + sum(u + v - 0.5*q*log(v./u) - log(bf) - z);
end
```

2.4 Implementing the log-likelihood function using the command ncx2pdf

There is an alternative way to implement the log-likelihood function in MATLAB and that is by using directly the noncentral χ^2 probability density function (pdf) which is available in the Statistics Toolbox under the command ncx2pdf. The function ncx2pdf relyes on an alternative definition of the noncentral χ^2 pdf. The alternative definition of the noncentral χ^2 pdf does not use the modified Bessel function of the first kind but is based on the central χ^2 distribution

[2] We already can encounter this problem when calculating the transition density (3). On one hand the Bessel function $I_q(2\sqrt{uv})$ approaches rapidly to the plus infinity, on the other hand the term $\exp(-u-v)$ approaches rapidly to the minus infinity and that causes calculation crash.

weighted by Poisson distribution[3].

In this alternative implementation we use the random variable transformation given by (4). We call `ncx2pdf(s,2*q+2,2.*u)`, where $s = 2cr_{t+\Delta t}$ are the transformed interest rate values. The MATLAB code of the log-likelihood function can look as follows:

```
function lnL = CIRobjective2(Params, Model)
% =====================================================================
% PURPOSE : Log-likelihood objective function (multiplied by -1) for the
%           CIR process using MATLAB ncx2pdf function.
% =====================================================================
% USAGE   : Model.TimeStep  = Delta t
%           Model.Data      = Time series of interest rates observations
%           Params          = Model parameters (alpha, mu, sigma)
% =====================================================================
% RETURNS : lnL             = Objective function value
% =====================================================================
    Data = Model.Data;
    DataF = Data(2:end);
    DataL = Data(1:end-1);
    TimeStep = Model.TimeStep;
    alpha = Params(1);
    mu = Params(2);
    sigma = Params(3);

    c = 2*alpha/(sigma^2*(1-exp(-alpha*TimeStep)));
    q = 2*alpha*mu/sigma^2-1;
    u = c*exp(-alpha*TimeStep)*DataL;
    v = c*DataF;
    s = 2*c*DataF;
    nc = 2*u; % noncentrality parameter
    df = 2*q+2; % degrees of freedom

    gpdf = ncx2pdf(s, df, nc);
    ppdf = 2*c*gpdf;
    lnL = sum(-log(ppdf));
end
```

2.5 Optimization

The minimization of the negative of the log-likelihood function can be achieved by the `fminsearch` function. Both suggested implementation of the log-likelihood function work equally well with only slight differences in results. The `ncx2pdf` implementation is generally more time consuming.

Both implementations of the log-likelihood are available under `CIRobjective1.m`, respectively, `CIRobjective2.m`. The optimization routine (initial estimates and optimization itself) is implemented under `CIRestimation.m`.

[3] Details can be found at `http://en.wikipedia.org/wiki/Noncentral_chi_distribution` or by reading the `ncx2pdf` source code.

3 Estimating the CIR Process for PRIBOR 3M

We present the results of estimating the CIR process for times series of daily (business days) observations of PRIBOR 3M from 3 January 1994 to 4 October 2007. As our data are sampled each business day, we set the time step $\Delta t = 1/250$. Time series of PRIBOR 3M is plotted in Figure 1.

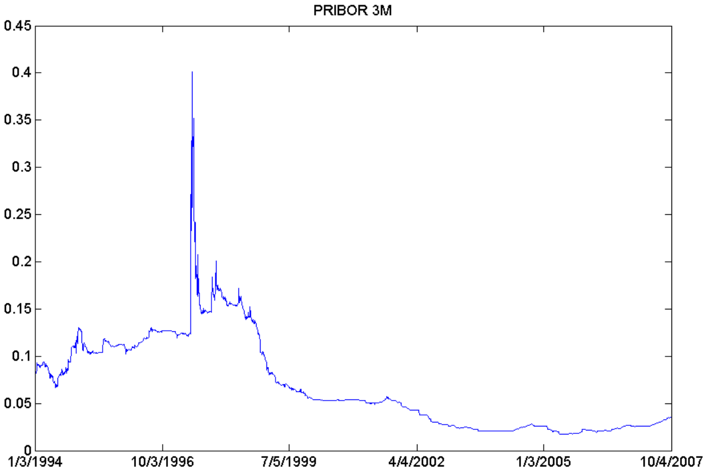

Figure 1: Time series of PRIBOR 3M from 1/3/1994 through 10/4/2007. Interest rate in percents.

3.1 Estimation results

The estimation results reports Table 1. In this particular case the OLS results are noticeably different from the maximum likelihood (ML) results which is not a common outcome for other interest rates data sets. It is caused by jumps of PRIBOR 3M in 1997. It is, of course, arguable whether the CIR process is a right choice for the given PRIBOR 3M data set but this is not our concern in this paper. Using either implementation of the log-likelihood function returns results which differ on the fourth decimal number.

	$\hat{\alpha}$	$\hat{\mu}$	$\hat{\sigma}$	$\ln L$
OLS (initial)	0.1209	0.0423	0.1642	
ML `besseli`	0.4363	0.0613	0.1491	4.7140
ML `ncx2pdf`	0.4360	0.0612	0.1492	4.7080

Table 1: This table reports the OLS and maximum likelihood (ML) results for daily (business day) PRIBOR 3M data set from 3 January 1994 to 4 October 2007. Time step $\Delta t = 1/250$. The $\ln L$ refers to the maximized value of the log-likelihood divided by the number of observations.

3.2 The optimization propagation and the log-likelihood

The optimization has been carried out using the function `fminsearch`. The optimization propagation is recorded in Figure 2. The `fminsearch` algorithm has terminated after 270 steps.

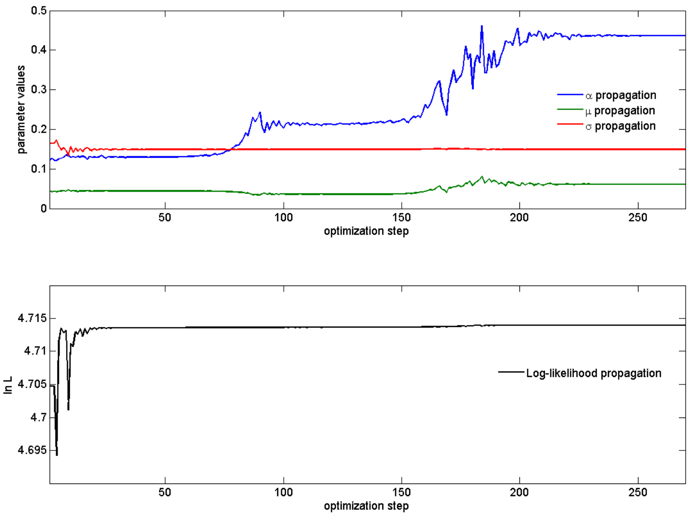

Figure 2: Optimization propagation for the PRIBOR 3M data set. The upper graph plots parameters propagation, the lower graph log-likelihood propagation.

The Figure 3 shows the log-likelihood space of α and μ for optimal σ. The Figure 4 shows the log-likelihood space of α and σ for optimal μ. Finally, the Figure 5 shows the log-likelihood space of μ and σ for optimal α.

We can observe that the log-likelihood function is very flat in the drift parameters α and μ and it is not an easy task to determine the optimal values of this parameters. Thus the first parameter to be optimized is σ and subsequently the values of α and μ are fine tuned. Compare the Figure 2 (optimization propagation) to the Figures 3, 4 and 5 (parameters space).

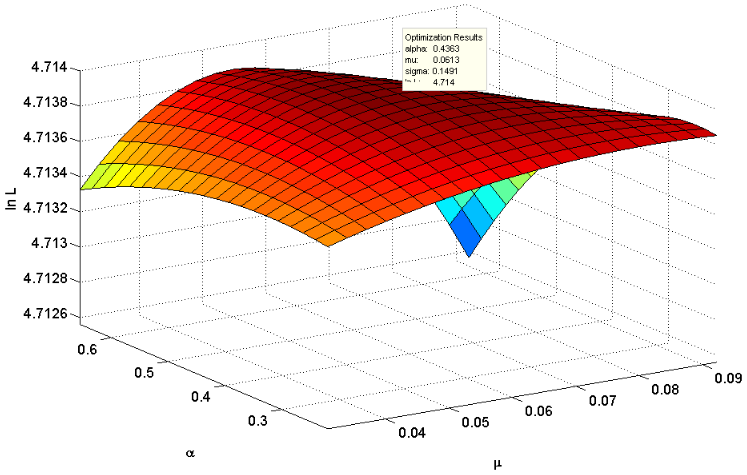

Figure 3: The log-likelihood space of α and μ for optimal σ ($\sigma = 0.1491$). Calculated using `besseli` implementation.

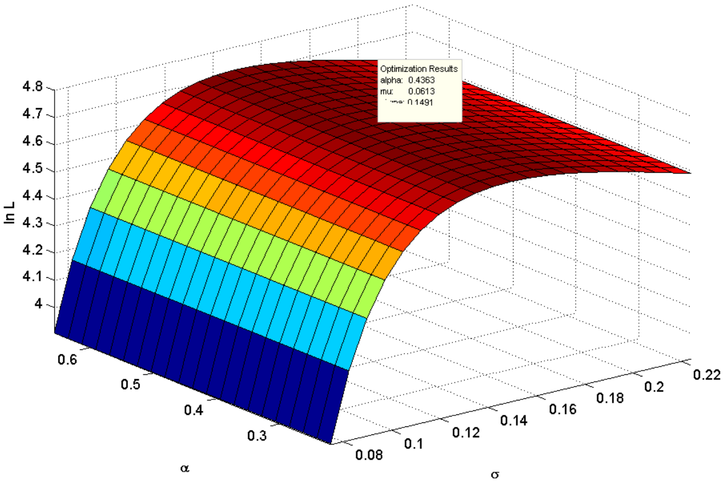

Figure 4: The log-likelihood space of α and σ for optimal μ ($\mu = 0.0613$). Calculated using `besseli` implementation.

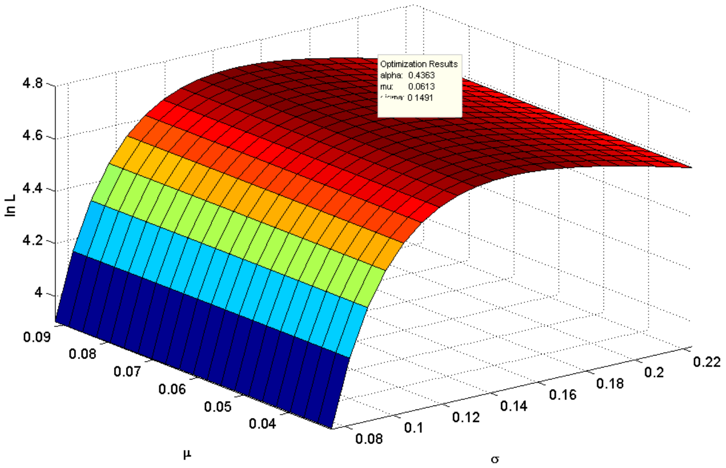

Figure 5: The log-likelihood space of μ and σ for optimal α ($\alpha = 0.4363$). Calculated using `besseli` implementation.

Conclusion

This paper has discussed the MATLAB implementation of the maximum likelihood estimation of the Cox-Ingersoll-Ross process. Two alternative approaches to evaluate the log-likelihood function have been provided. OLS regression has been suggested for initial parameters estimates. The implementation has been demonstrated for the PRIBOR 3M time series.

References

[1] Cox, J.C., Ingersoll, J.E and Ross, S.A. *A Theory of the Term Structure of Interest Rates.* Econometrica, Vol. 53, No. 2 (March, 1985)

[2] Feller, W. *Two Singular Diffusion Problems.* Annals of Mathematics, Vol. 54, No. 1 (July, 1951)