QML Estimation of the Parameters of SDEs

Stan Hurn[1] Ken Lindsay[2]

[1]Queensland University of Technology

[2]University of Glasgow

Fields Institute, Toronto, 24 April 2010

Table of contents

Statement of the problem

Suppose the N-dimensional process $\mathbf{X}(t)$ satisfies the stochastic differential equation

$$d\mathbf{X} = \boldsymbol{\mu}(\mathbf{X};\boldsymbol{\theta})\,dt + \sigma(\mathbf{X};\boldsymbol{\theta})\,d\mathbf{W}$$

where

- $\boldsymbol{\mu}$ is the drift (N-dimensional column vector)
- σ is the diffusion (an $N \times M$ matrix with $M \leq N$)
- $\boldsymbol{\theta}$ is a vector of model parameters
- $d\mathbf{W}$ is the increment of the vector Wiener process (M-dimensional column vector).

Estimate the parameters $\boldsymbol{\theta}$ from observations $\mathbf{X}_0, \cdots, \mathbf{X}_T$ at discrete times $t_0, \cdots, t_T$.

Maximum likelihood

Common starting point for thinking about the estimation problem is the maximum likelihood principal. Estimate $\boldsymbol{\theta}$ by minimizing the negative log-likelihood function for the observed sample, namely

$$-\log \mathcal{L}(\theta) = -\sum_{t=1}^{T} \log f_0(\mathbf{X}_t|\mathbf{X}_{t-1};\boldsymbol{\theta}),$$

with respect to the parameters $\boldsymbol{\theta}$. In this expression

- $f_0(\mathbf{X}_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$ is the true transitional density of the process;
- it is assumed that all the states are observed; and
- the sum is the contribution to the negative log-likelihood function from all the transitions.

Problems and Challenges

Problems

- Transitional density $f_0(\mathbf{X}_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$ and therefore the likelihood function seldom known in closed-form.
- Generalization to estimation in a multivariate setting is difficult.
- Multivariate models introduce the possibility of unobserved variables

Problems and Challenges

Problems

- Transitional density $f_0(\mathbf{X}_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$ and therefore the likelihood function seldom known in closed-form.
- Generalization to estimation in a multivariate setting is difficult.
- Multivariate models introduce the possibility of unobserved variables

Challenges

- To develop a simple yet comprehensive framework for consistent estimation of the parameters of multivariate SDEs.
- Handle unobserved state variables easily and effectively.

Some philosophy ...

Some philosophy ...

William Seach (c. 1285 – 1349)

Frustra fit per plura quod potest fieri per pauciora.
(It is futile to do with more things that which can be done with fewer).

Some philosophy ...

William Seach (c. 1285 – 1349)

Frustra fit per plura quod potest fieri per pauciora.
(It is futile to do with more things that which can be done with fewer).

Leonardo da Vinci (1452 – 1519)

Simplicity is the ultimate sophistication.

Some philosophy ...

William Seach (c. 1285 – 1349)

Frustra fit per plura quod potest fieri per pauciora.
(It is futile to do with more things that which can be done with fewer).

Leonardo da Vinci (1452 – 1519)

Simplicity is the ultimate sophistication.

Karl Menger (1902 – 1985) on parsimonious mathematicians

It is vain to do with fewer what requires more.

Some philosophy ...

William Seach (c. 1285 – 1349)

Frustra fit per plura quod potest fieri per pauciora.
(It is futile to do with more things that which can be done with fewer).

Leonardo da Vinci (1452 – 1519)

Simplicity is the ultimate sophistication.

Karl Menger (1902 – 1985) on parsimonious mathematicians

It is vain to do with fewer what requires more.

Sherlock Holmes, in *The Sign of Four*

When you have eliminated the impossible, Watson, whatever remains, however improbable, must be the truth.

Simplest option is to use a Gaussian ...

Consider the transition from X_t to X_{t+1} for time step Δ

1. **DML**:

$$f_1(X_{t+1}|X_t;\boldsymbol{\theta}) = \phi(X_{t+1};\boldsymbol{\theta})\,,$$

   where the mean and variance of $\phi(\cdot)$ are

$$X_t + \mu(X_t;\boldsymbol{\theta})\Delta, \qquad \sigma^2(X_t;\boldsymbol{\theta})\Delta\,.$$

2. **HPE**: model the transitional density as

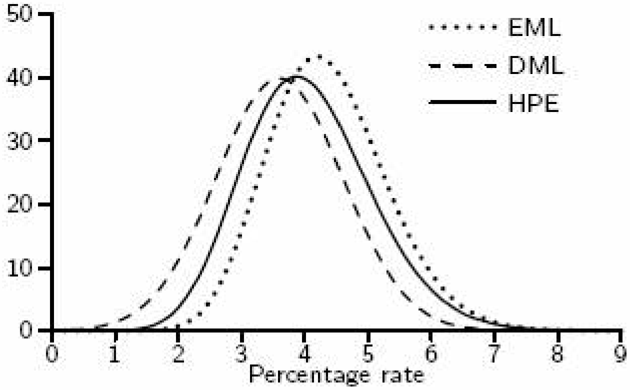
$$f(X_{t+1}|X_t;\boldsymbol{\theta}) = \frac{dZ}{dX_{t+1}}\phi(Z_{t+1};\boldsymbol{\theta})\sum_{j=0}^{\infty}\eta_j(Z_t,\Delta)\,H_j\Big(\frac{Z_{t+1}-Z_t}{\Delta}\Big)$$

3. **QML**: use $\phi(X_{t+1};\boldsymbol{\theta})$ but focus on getting the correct mean and variance of true transitional distribution.

Transitional Density of CIR Model

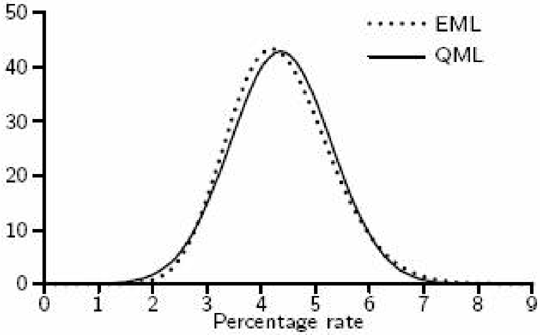
Consider the transitional density of the CIR model

$$dX = 0.2(0.06 - X)\,dt + 0.05\sqrt{X}\,dW \qquad \text{with} \quad DT = 1\,.$$



Transitional Density of CIR Model

A Gaussian with ‘true’ conditional mean and variance ...



Basic Idea

Replace the true transitional density $f_0(\mathbf{X}_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$ in the likelihood function by the multivariate Gaussian density

$$f_1(\mathbf{X}_t|\mathbf{X}_{t-1};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}}\frac{1}{|\Sigma_1|^{1/2}}\exp\Big[-\frac{1}{2}(\mathbf{X}_t-\mu_1)^{\mathrm{T}}\Sigma_1^{-1}(\mathbf{X}_t-\mu_1)\Big],$$

where μ_1 and Σ_1 are respectively the conditional mean and conditional covariance.

Crucial question

Is it possible to obtain estimates of the 'true' conditional moments, μ_0 and Σ_0, without needing to know $f_0(\mathbf{X}_t|\mathbf{X}_{t-1};\boldsymbol{\theta})$. If so we can get consistent estimates and standard errors of the parameters of the model.

Infinitesimal operator

Given a suitably differentiable function of state, ψ, the infinitesimal operator $\mathcal{A}$, is defined by

$$\mathcal{A}(\psi) = \mu(x;\boldsymbol{\theta})\frac{d\psi}{dx} + \frac{g(x;\boldsymbol{\theta})}{2}\frac{d^2\psi}{dx^2}$$

has property that it expresses the time derivative of an expected value as an expected value taken in state space

$$\frac{d\mathrm{E}\left[\psi\right]}{dt} = \int_{\mathrm{S}} \mathcal{A}(\psi)\, f(x,t)\, dx\,.$$

The infinitesimal operator may now be used to generate appropriate equations for the evolution of the first and second moments of $f_0(x,t)$ by taking $\psi = x$ and $\psi = x^2$ respectively.

Evolution of Moments

One dimensional

$$\frac{d}{dt}\int_{\mathcal{S}} x \;\, f_0(x,t)\,dx = \int_{\mathcal{S}} \mu(x) f_0(x,t)\,dx, ,$$
$$\frac{d}{dt}\int_{\mathcal{S}} x^2 f_0(x,t)\,dx = 2\int_{\mathcal{S}} x\mu(x) f_0(x,t)\,dx + \int_{\mathcal{S}} g(x) f_0(x,t)\,dx.$$

Evolution of Moments

One dimensional

$$\frac{d}{dt}\int_{\mathcal{S}} x\;\; f_0(x,t)\,dx = \int_{\mathcal{S}} \mu(x) f_0(x,t)\,dx,,$$
$$\frac{d}{dt}\int_{\mathcal{S}} x^2 f_0(x,t)\,dx = 2\int_{\mathcal{S}} x\mu(x) f_0(x,t)\,dx + \int_{\mathcal{S}} g(x) f_0(x,t)\,dx.$$

Multi-dimensional

$$\begin{aligned}
\frac{d}{dt}\int_{\mathcal{S}} \mathbf{x}\, f_0(\mathbf{x},t)\,d\mathbf{x} &= \int_{\mathcal{S}} \boldsymbol{\mu}\, f_0(\mathbf{x},t)\,d\mathbf{x}\,, \\
\frac{d}{dt}\int_{\mathcal{S}} \mathbf{x}\mathbf{x}'\;\, f_0(\mathbf{x},t)\,d\mathbf{x} &= \int_{\mathcal{S}} (\boldsymbol{\mu}\mathbf{x}' + \mathbf{x}\boldsymbol{\mu}')\, f_0(\mathbf{x},t)\,d\mathbf{x} \\
&\quad + \int_{\mathcal{S}} \mathbf{g}(x)\, f_0(\mathbf{x},t)\,d\mathbf{x}\,.
\end{aligned}$$

CIR Model

Consider the Cox, Ingersol and Ross square-root diffusion

$$dX = \lambda(\beta - X)\, ds + \sigma\sqrt{X}\, dW$$

Moment equations

$$\begin{aligned}
\frac{d}{dt}\int_0^\infty x\, f_0(x,t)\, dx &= \lambda\beta - \lambda\int_0^\infty x\, f_0(x,t)\, dx\,, \\
\frac{d}{dt}\int_0^\infty x^2\, f_0(x,t)\, dx &= (\sigma^2 + 2\lambda\beta)\int_0^\infty x\, f_0(x,t)\, dx \\
&\qquad -2\lambda\int_0^\infty x^2\, f_0(x,t)\, dx\,.
\end{aligned}$$

CIR Model

The solution of the mean equation is

$$\mu_0(t) = \beta + (X_t - \beta)e^{-\lambda s}\,,$$

which in turn allows the second equation to be integrated to obtain

$$\Sigma_0(t) = \frac{\sigma^2(1 - e^{-\lambda s})}{2\lambda}\left[\beta(1 - e^{-\lambda s}) + 2X_t e^{-\lambda s}\right].$$

Thus the conditional mean and conditional variance of the square-root process evolving from observation X_t at time t are known exactly without requiring a detailed specification of the form of the true transitional probability density function.

CKLS Model

The CKLS model is

$$dX = \alpha(\theta - X)\, dt + \sigma\, X^{\beta/2}\, dW\,.$$

Because the CKLS model has affine drift specification, the evolution of the first moment is unaffected by the introduction of the levels effect. The second moment equation becomes

$$\begin{aligned}\frac{d}{dt}\int_0^\infty x^2 f_0(x,t)\, dx &= -2\alpha\int_0^\infty x^2 f_0(x,t)\, dx + 2\alpha\theta\int_0^\infty x f_0(x,t)\, dx \\ &\quad +\sigma^2\int_0^\infty x^\beta f_0(x,t)\, dx\,.\end{aligned}$$

The difficulty now stems from the last integral in this equation. Thus in the case of non-affine models, certain moments may be determined in closed form while others must be computed numerically.

Approximating integrals

The presence of the integral

$$\sigma^2 \int_0^\infty x^\beta f_0(x, t)\, dx$$

in the variance equation means that this equation no longer has a closed-form solution and the form of the approximation $f_1(x, t)$ will matter to the determination of the moments.

There are now two sources of error:

- error due to misspecification of the transitional density;
- error due to numerical approximation of this integral.

It turns out that this integral may approximated with negligible error for all practical purposes so that the only the first source of error remains.

Optimal quadratures

When $f_0(x, t)$ is replaced by the Gaussian density function the resulting integral becomes

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_0^\infty x^\beta \exp\Big[-\frac{(x-\mu)^2}{2\sigma^2}\Big]\,dx\,.$$

Following the usual procedure of changing the variable of integration from x to z where $x = \mu + \sigma z$, then

$$\frac{1}{\sqrt{2\pi}} \int_{-\mu/\sigma}^\infty (\mu + \sigma z)^\beta\, e^{-z^2/2}\,dz\,.$$

In the context of integrals with integrands of generic form $e^{-z^2/2} f(z)$, the Gauss-Hermite quadrature can be used and this quadrature has maximum precision.

General Comments on Evolution of Moments

A few concluding comments may be made at this point regarding the computation of the moments of $f_0(x, t)$.

- If the drift and the diffusion functions are affine functions of state then the moments of the true transitional density may be computed exactly without the need to know $f_0(x, t)$.
- If the drift and diffusion functions are non-affine functions of state, then the moments cannot be computed exactly.
- In many cases of interest, the drift is affine but the diffusion is not. In these instances the problematic integrals in the solution of the second moment equation can be approximated to high-accuracy, leaving only the misspecification of the shape of $f_0(x, t)$ as the source of error.
- The leading driver of the error in the approximation arises from the first unsatisfied moment condition.

Problem of Unobserved State Variables

Three ways to deal with this problem in the contest of MLE

1. Find a proxy for the unobserved state variable.
2. Enlarge the dimension of the observations.
3. Integrate the unobservables out of the likelihood function.

Note

- For the stochastic volatility model, methods (1) and (2) above are associated with the HPE estimation method. QML is equally applicable in these situations.
- Based on QML as described here, integrating out the unobservables is straightforward and it is this process that will now be described.

Bivariate Normal

The procedure is best demonstrated with reference to the bivariate normal probability density function:

$$X = \begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \right)$$

then the transitional probability density function is given by

$$f_1(x, y) = \frac{1}{(2\pi)} \frac{1}{\sigma_x\sigma_y\sqrt{(1-\rho^2)}} \exp\left[-\frac{z}{2(1-\rho^2)} \right],$$

where

$$z = \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_x^2}$$

Now suppose y is unobserved ...

If y is an unobserved variable then the marginal probability density for x must be used in the construction of the likelihood function. This marginal density is obtained by integrating y out of the joint density function.

For the bivariate normal the marginal density for x obtained in this way is

$$f^{(m)}(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_x} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} \right],$$

that is, a Gaussian distribution with mean μ_x and variance σ_x^2.

This result generalizes to the multivariate Gaussian probability density function with K observable variables. The added complexity is that the relevant covariance matrix of the Gaussian marginal distribution is now the $K \times K$ principal minor of the original covariance matrix Σ.

Conditional density

The conditional density of y given that $x = x_0$ is

$$f^{(\mathrm{c})}(y \,|\, x = x_0) = \frac{f(\,x_0, y)}{f^{(\mathrm{m})}(x_0)}\,.$$

In the case of the bivariate Gaussian density,

$$f^{(\mathrm{c})}(y, \,|\, x = x_0) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_y\sqrt{1-\rho^2}}\exp\Big[-\frac{(y-\mu_y-\frac{\rho\sigma_y}{\sigma_x}(x_0-\mu_x))^2}{2\sigma_y^2(1-\rho^2)}\Big].$$

Consequently $f^{(\mathrm{c})}(y \,|\, x = x_0)$ is the Gaussian probability density function with respective mean and variance

$$\mu_y + \frac{\rho\sigma_y}{\sigma_x}(x_0-\mu_x) \quad \text{and} \quad \sigma_y^2(1-\rho^2).$$

CIR and CKLS Models

The performance of the QML methodology for estimating the parameters of univariate diffusions is now examined by way of a simple simulation experiment. The CIR model

$$dX = \lambda(\beta - X)\, ds + \sigma\sqrt{X}\, dW$$

and the CKLS model

$$dX = \lambda(\beta - X)\, ds + \sigma\, X^{\gamma/2}\, dW\,,$$

are used to generate 2000 samples of size 500 and 2000 respectively with true parameters $\lambda = 0.20$, $\beta = 0.08$, $\sigma = 0.10$ and, where necessary, $\gamma = 0.75$. The synthetic samples are generated with $\Delta t = 1/12$ (so that the data may be interpreted as monthly data) using Milstein's scheme with 1000 steps between observations to ensure accurate realizations of the process.

Bias and RMSE - Univariate Models

$\Delta t = 1/12$		T = 500				T = 2000		
	$\lambda(0.20)$	$\beta(0.08)$	$\sigma(0.10)$	$\gamma(0.75)$	α (0.20)	γ (0.08)	σ (0.10)	$\gamma(0.75)$
CIR	0.1016 (0.1318)	0.0014 (0.0222)	0.0001 (0.0032)		0.0229 (0.0552)	0.0005 (0.0113)	0.0001 (0.0016)	
CIR QML	0.1006 (0.1322)	0.0018 (0.0226)	0.0001 (0.0032)		0.0230 (0.0562)	0.0005 (0.0113)	0.0001 (0.0016)	
CKLS (O5)	0.0995 (0.1337)	0.0010 (0.0368)	0.0058 (0.0368)	−0.0001 (0.1316)	0.0231 (0.0569)	0.0002 (0.0125)	0.0004 (0.0146)	−0.0026 (0.0560)
CKLS (O7)	0.0995 (0.1337)	0.0009 (0.0125)	0.0058 (0.0368)	−0.0001 (0.1316)	0.0231 (0.0569)	0.0002 (0.0061)	0.0004 (0.0146)	−0.0026 (0.0560)

Bias and RMSE (in parentheses) of parameter estimates of the CIR model obtained by exact maximum likelihood and quasi-maximum likelihood. Also shown are the bias and RMSE of the CKLS model estimated by quasi-maximum likelihood and using Gauss-Hermite quadratures of order 5 and 7 respectively. Estimates are based on 2000 simulations for samples of size 500 and 2000 monthly observations.

Bivariate Feller Model

Let $\mathbf{X} = (X_1(t), X_2(t))$ be a bivariate stochastic processes with sample space S and satisfying the stochastic differential equations (SDEs)

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} dt + \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} dW_1 \\ dW_2 \end{bmatrix} \tag{1}$$

where dW_1 and dW_2 are independent increments in the Wiener processes W_1 and W_2, the vector $\boldsymbol{\mu}$ and matrix $\boldsymbol{\sigma}$ are respectively

$$\boldsymbol{\mu} = \begin{bmatrix} \alpha_1 + k_{11}X_1 + k_{12}X_2 \\ \alpha_2 + k_{21}X_1 + k_{22}X_2 \end{bmatrix}, \qquad \boldsymbol{\sigma} = \begin{bmatrix} \sqrt{X_1} & 0 \\ 0 & \sqrt{X_2} \end{bmatrix}. \tag{2}$$

The parameters k_{11}, k_{12}, k_{21}, k_{22} , α_1 and α_2 are to be estimated. Values used in the simulation are $(-0.7, 0.3, 0.4, -0.8, 0.56, 0.64)$.

Feller Model Results - Bias and MSE

Parm	True Value	500 Transitions Unconditioned	500 Transitions Conditioned	2000 Transitions Unconditioned	2000 Transitions Conditioned
k_{11}	(-0.70)	-0.1103 (0.0627)	-0.3713 (0.6796)	-0.0249 (0.0113)	-0.1326 (0.1233)
k_{12}	(0.30)	0.0168 (0.0366)	0.2874 (0.9012)	0.0050 (0.0075)	0.1739 (0.4336)
k_{21}	(0.40)	0.0235 (0.0475)	0.3047 (1.6029)	0.0062 (0.0094)	0.3942 (1.5380)
k_{22}	(-0.80)	-0.1175 (0.0699)	-0.7324 (2.7428)	-0.0275 (0.0124)	-0.6366 (2.5190)
α_1	(0.56)	0.1010 (0.0977)	0.1327 (0.1804)	0.0214 (0.0146)	0.0335 (0.0717)
α_2	(0.64)	0.1181 (0.1075)	0.3403 (1.2712)	0.0227 (0.0171)	0.2715 (1.0438)

Heston model

In the Heston model the logarithm of the asset price s and the volatility h follow the risk-neutral dynamics

$$\begin{bmatrix} ds \\ dh \end{bmatrix} = \begin{bmatrix} \alpha + \beta h \\ \kappa(\gamma - h) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)h_t} & \rho\sqrt{h_t} \\ 0 & \sigma\sqrt{h_t} \end{bmatrix} \begin{bmatrix} dW_1 \\ dW_2 \end{bmatrix}$$

where α (fixed at 0.03) is the difference between the instantaneous risk-free rate of interest and the dividend and $\beta = \lambda(1-\rho^2) - \frac{1}{2}$. This specification embodies the assumption that the market price of risk specification in the model is $\Lambda = [\lambda\sqrt{(1-\rho^2)h_t},\, 0]'$.

The parameters to be estimated are κ, γ, σ, ρ and λ.

Results - Point Estimates and RMSEs

Parm	True Value	500 Transitions Unconditioned	500 Transitions Conditioned	2000 Transitions Unconditioned	2000 Transitions Conditioned
κ	(3.00)	3.8710 (0.0474)	4.1870 (0.2080)	3.1860 (0.0190)	3.0766 (0.0531)
γ	(0.10)	0.0990 (0.0006)	0.0984 (0.0007)	0.0999 (0.0003)	0.0991 (0.0004)
σ	(0.25)	0.2502 (0.0002)	0.3609 (0.0190)	0.2501 (0.0001)	0.2892 (0.0076)
ρ	(−0.80)	−0.8010 (0.0004)	−0.7418 (0.0094)	−0.8000 (0.0002)	−0.7529 (0.0067)
λ	(4.00)	5.3270 (0.1680)	5.6880 (0.2260)	4.3105 (0.0939)	4.1878 (0.0891)

Comments on Heston Model Results

- The results for the estimation of the unconditional model are almost identical to those reported by Aït-Sahalia and Kimmel using their closed-form likelihood approximation.
- The conditioned model uses precisely half of the information of the unconditioned model and a deterioration in performance is therefore inevitable. The point estimates of the model, however, hold up remarkably well.
- The loss of information is evident in the size of the standard errors which are significantly larger than the unconditional case (apart from the parameter γ).

Concluding Comments

This is work in progress but the results are promising.

- QML yields consistent estimates of the parameters of models where the drift and diffusion are affine functions of state.
- Consistency not able to be proved for non-affine models but method works very well in practice. Good performance in the non-affine models we have considered thus far stems from the fact that the problematic integrals are well approximated by known quadratures.
- Method works well for the bivariate models considered thus far and is capable of handling unobservable state variables easily.