Financial econometrics based on stochastic differential equations and the `sde` package

S.M. Iacus (University of Milan)

Rennes, useR! 2009, July 8th - 10th

Time Series vs SDEs

Consider the AR(1) process. It is a **discrete-time** random process, defined as

$$X_t = \theta X_{t-1} + \epsilon_t, \quad X_0 = x_0, \quad \epsilon_t : \quad \text{i.i.d. random variables (noise)}$$

Consider the AR(1) process. It is a **discrete-time** random process, defined as

$$X_t = \theta X_{t-1} + \epsilon_t, \quad X_0 = x_0, \quad \epsilon_t : \quad \text{i.i.d. random variables (noise)}$$

Its **continuous-time** counter part (the Ornstein-Uhlenbeck process), written in differential form, looks like

$$\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x_0, \quad W_t : \quad \text{the Wiener process (noise)}$$

Time Series vs SDEs

Consider the AR(1) process. It is a **discrete-time** random process, defined as

$$X_t = \theta X_{t-1} + \epsilon_t, \quad X_0 = x_0, \quad \epsilon_t : \quad \text{i.i.d. random variables (noise)}$$

Its **continuous-time** counter part (the Ornstein-Uhlenbeck process), written in differential form, looks like

$$\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x_0, \quad W_t : \quad \text{the Wiener process (noise)}$$

A stochastic differential equation models a dynamical system with feedback by adding continuous time shocks

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$$

In continuous time models: time between X_t and $X_{t+\Delta t}$ matters ! The length of Δ_t is crucial as well.

In continuous time models: time between X_t and $X_{t+\Delta t}$ matters ! The length of Δ_t is crucial as well.

In time-series models: nothing happens (probabilistically) between X_t and X_{t-1}

Time Series vs SDEs

In continuous time models: time between X_t and $X_{t+\Delta t}$ matters ! The length of Δ_t is crucial as well.

In time-series models: nothing happens (probabilistically) between X_t and X_{t-1}

Why this matters? An example: according to McCrorie & Chambers (2006, J. of Econ.) and others, *“spurious Granger causality [tested with VAR models] is only a consequence of the intervals in which economic data are generated being finer than the econometrician’s sampling interval.*”

Conclusions: assume a continuous time model (SDE). Discretize that, build a VAR from the discretized SDE and the spurious Granger causality vanishes!

Time Series vs SDEs

In continuous time models: time between X_t and $X_{t+\Delta t}$ matters ! The length of Δ_t is crucial as well.

In time-series models: nothing happens (probabilistically) between X_t and X_{t-1}

Why this matters? An example: according to McCrorie & Chambers (2006, J. of Econ.) and others, *“spurious Granger causality [tested with VAR models] is only a consequence of the intervals in which economic data are generated being finer than the econometrician’s sampling interval.*”

Conclusions: assume a continuous time model (SDE). Discretize that, build a VAR from the discretized SDE and the spurious Granger causality vanishes!

Rephrasing: why using a binomial distribution if your underlying model is a Gaussian?

A few examples of SDEs

- gBm : $\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$
- CIR : $\mathrm{d}X_t = (\theta_1 + \theta_2 X_t)\mathrm{d}t + \theta_3 \sqrt{X_t}\mathrm{d}W_t$
- CKLS : $\mathrm{d}X_t = (\theta_1 + \theta_2 X_t)\mathrm{d}t + \theta_3 X_t^{\theta_4}\mathrm{d}W_t$
- nonlinear mean reversion (Aït-Sahalia)
  $\mathrm{d}X_t = (\alpha_{-1} X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)\mathrm{d}t + \beta_1 X_t^{\rho}\mathrm{d}W_t$
- double Well potential (bimodal behaviour, highly nonlinear)
  $\mathrm{d}X_t = (X_t - X_t^3)\mathrm{d}t + \mathrm{d}W_t$
- Jacobi diffusion (political polarization):
  $\mathrm{d}X_t = -\theta\left(X_t - \frac{1}{2}\right)\mathrm{d}t + \sqrt{\theta X_t(1 - X_t)}\mathrm{d}W_t$
- radial Ornstein-Uhlenbeck : $\mathrm{d}X_t = (\theta X_t^{-1} - X_t)\mathrm{d}t + \mathrm{d}W_t$
- hyperbolic diffusion : $\mathrm{d}X_t = \frac{\sigma^2}{2}\left[\beta - \gamma \frac{X_t}{\sqrt{\delta^2 + (X_t - \mu)^2}}\right]\mathrm{d}t + \sigma \mathrm{d}W_t$

Diffusion processes solutions to SDEs

From the statistical point of view, we are interested in the parametric family of diffusion process solutions of the SDE

$$\mathrm{d}X_t = b(X_t, \theta)\mathrm{d}t + \sigma(X_t, \theta)\mathrm{d}W_t, \quad X_0 = x_0, \quad t \in [0, T]$$

$\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$, where $\Theta_\alpha \subset R^p$ and $\Theta_\beta \subset R^q$.

Observations always come in discrete time form at some times $t_i = i\Delta_n$, $i = 0, 1, 2, ..., n$, where Δ_n is the length of the steps. We denote the observations by $\mathbf{X}_n := \{X_i = X_{t_i}\}_{0 \leq i \leq n}$.

From the statistical point of view, we are interested in the parametric family of diffusion process solutions of the SDE

$$\mathrm{d}X_t = b(X_t, \theta)\mathrm{d}t + \sigma(X_t, \theta)\mathrm{d}W_t, \quad X_0 = x_0, \quad t \in [0, T]$$

$\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$, where $\Theta_\alpha \subset R^p$ and $\Theta_\beta \subset R^q$.

Observations always come in discrete time form at some times $t_i = i\Delta_n$, $i = 0, 1, 2, ..., n$, where Δ_n is the length of the steps. We denote the observations by $\mathbf{X}_n := \{X_i = X_{t_i}\}_{0 \le i \le n}$.
Different sampling schemes, different statistical procedures:

1. Large sample asymptotics: Δ fixed, $T = n\Delta \to \infty$ as $n \to \infty$
2. High frequency: $T = n\Delta_n$ fixed, $\Delta_n \to 0$ as $n \to \infty$
3. Rapidly increasing design: $T = n\Delta \to \infty$, $\Delta_n \to 0$ as $n \to \infty$ under the additional condition $n\Delta_n^k \to 0$ for $k > 1$

Likelihood in discrete time

By Markov property of diffusion processes, the likelihood has this form

$$L_n(\theta) = \prod_{i=1}^{n} p_\theta\left(\Delta, X_i | X_{i-1}\right) p_\theta(X_0)$$

Problem: the transition density $p_\theta\left(\Delta, X_i | X_{i-1}\right)$ is often not available! Only for OU, CIR and gBm

By Markov property of diffusion processes, the likelihood has this form

$$L_n(\theta) = \prod_{i=1}^{n} p_\theta\left(\Delta, X_i | X_{i-1}\right) p_\theta(X_0)$$

Problem: the transition density $p_\theta\left(\Delta, X_i | X_{i-1}\right)$ is often not available! Only for OU, CIR and gBm

Solutions:

- discretization of the SDE (Euler, Milstein, Ozaki, etc)
- simulation method
- hermite polynomial expansion
- partial differential equations
- other approximations of the transition density

Local Gaussian Approximation. Euler Scheme.

By Euler discretization of the SDE : $\mathrm{d}X_t = b(X_t, \theta)\mathrm{d}t + \sigma(X_t, \theta)\mathrm{d}W_t$

$$X_{t+\Delta t} - X_t = b(X_t, \theta)\Delta t + \sigma(X_t, \theta)(W_{t+\Delta t} - W_t),$$

we get an approximate transition density which is Gaussian. This is widely seen in applied contexts. But is this approximation good or not? In general no!

For example, for gBm, the true transition density is a log-normal and the Euler schemes provides only a Gaussian approximation!

It is possible to prove that estimators are not even consistent for non negligible Δ.

Consider OU model

$$\mathrm{d}X_t = (\theta_1 - \theta_2 X_t)\mathrm{d}t + \theta_3 \mathrm{d}W_t, \qquad X_0 = x_0$$

Both true and Euler approximation are Gaussian respectively with mean and variance

$$m(\Delta, x) = xe^{-\theta_2\Delta} + \frac{\theta_1}{\theta_2}\left(1 - e^{-\theta_2\Delta}\right), \qquad v(\Delta, x) = \frac{\theta_3^2\left(1 - e^{-2\theta_2\Delta}\right)}{2\theta_2},$$

and (Euler)

$$m^{Euler}(\Delta, x) = x(1 - \theta_2\Delta) + \theta_1\Delta, \qquad v^{Euler}(\Delta, x) = \theta_3^2\Delta,$$

Only under high-frequency setting, i.e. $\Delta \to 0$, the approximation is acceptable.

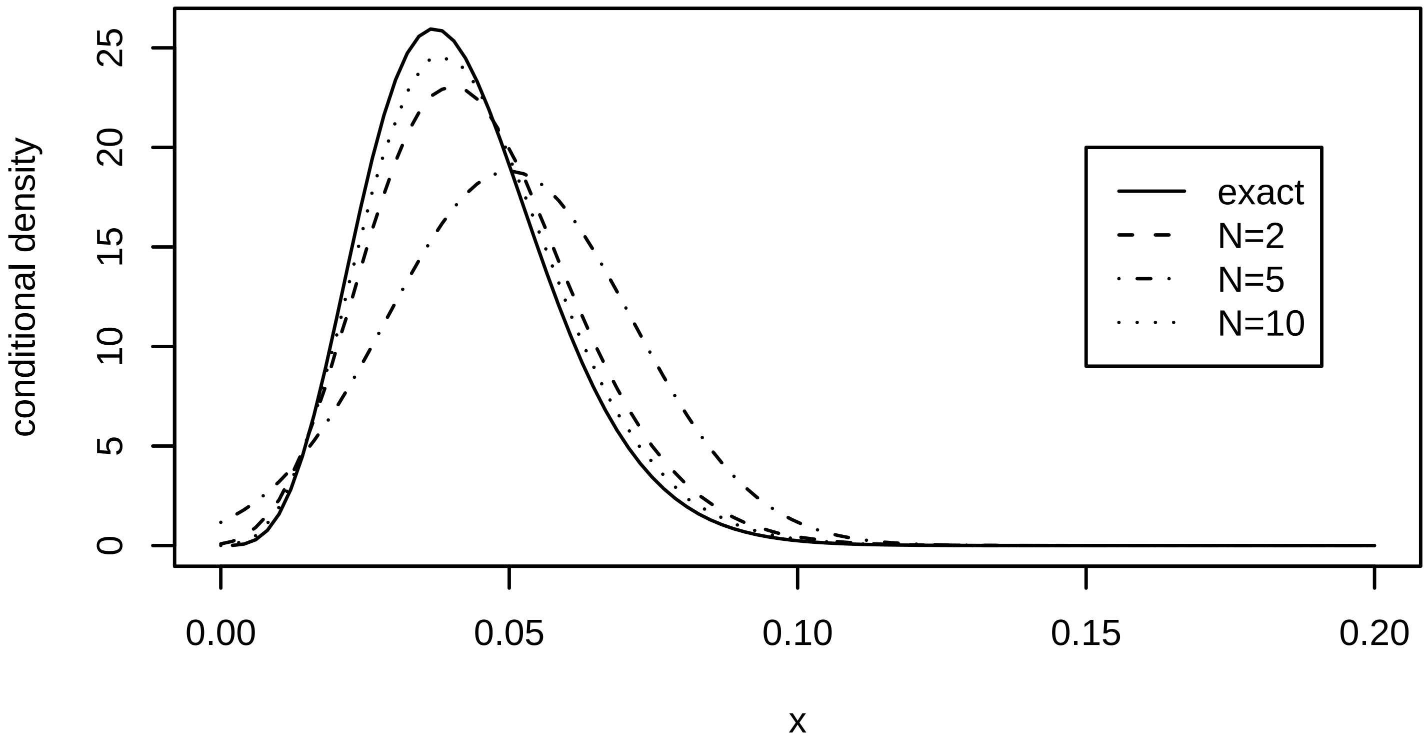
Let $p_\theta(\Delta, y|x)$ be the true transition density of $X_{t+\Delta}$ at point y given $X_t = x$. Consider a $\delta << \Delta$, for example $\delta = \Delta/N$ for N large enough, and then use the Chapman-Kolmogorov equation as follows:

$$p_\theta(\Delta, y|x) = \int p_\theta(\delta, y|z) p_\theta(\Delta - \delta, z|x) \mathrm{d}z = \mathbb{E}_z\{p_\theta(\delta, y|z)|\Delta - \delta\},$$

It means that $p_\theta(\Delta, y|x)$ is seen as the expected value over all possible transitions of the process from time $t + (\Delta - \delta)$ to $t + \Delta$, taking into account that the process was in x at time t.

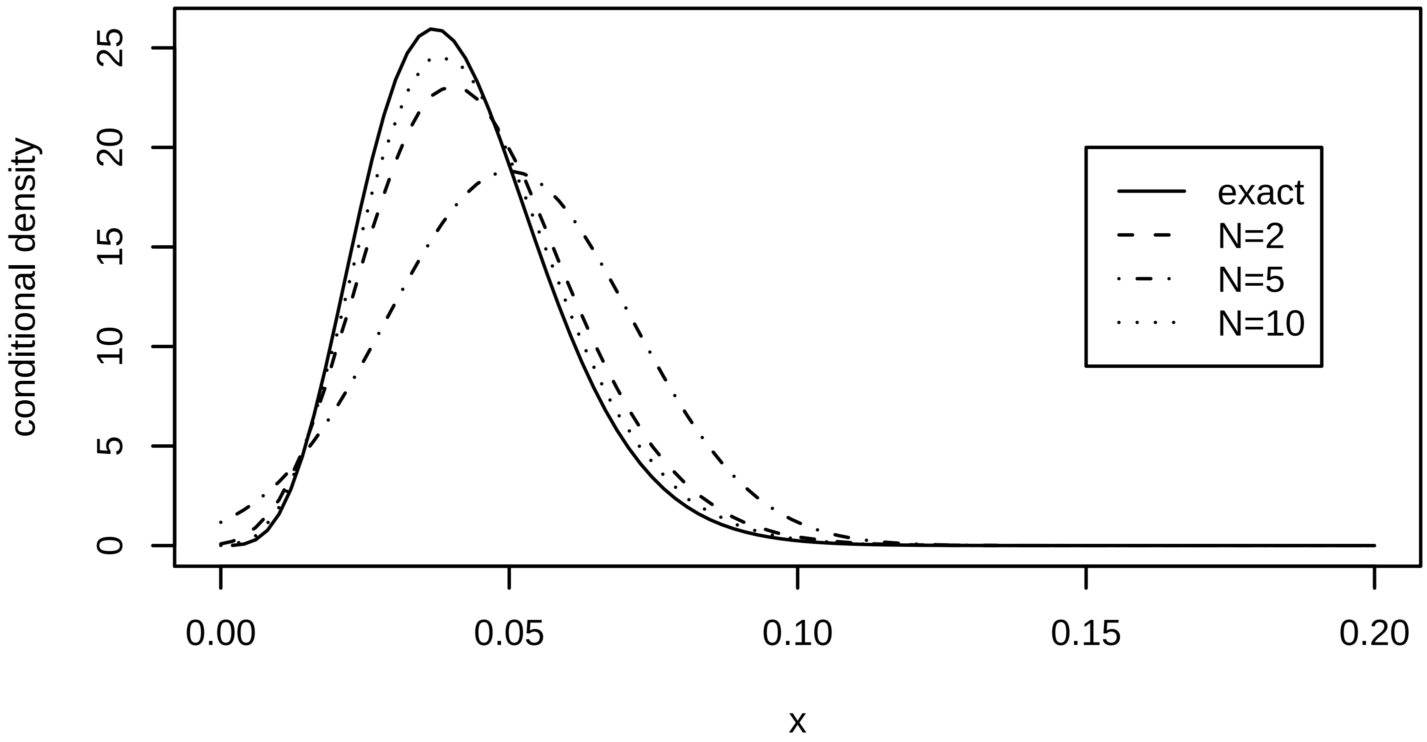
So we need simulations!

Example: approximation for the CIR model



What about N? We need many simulations

Example: approximation for the CIR model



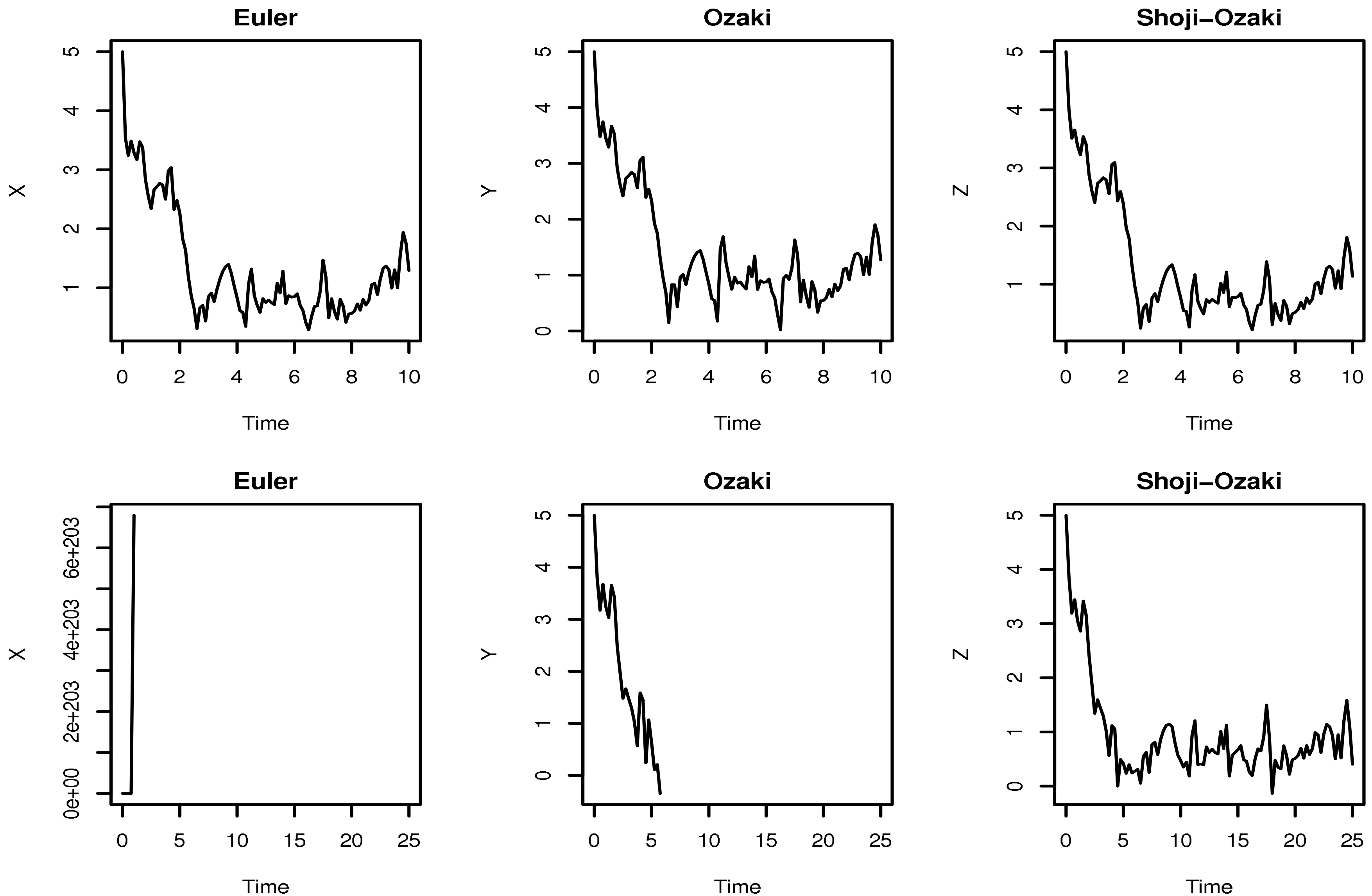
We need many simulations (N) for each time points (X_{t_i}, $X_{t_i+\Delta}$). But not all simulation schemes are stable for all models

Numerical instability. Up|Down $\Delta = 0.1 | 0.25$

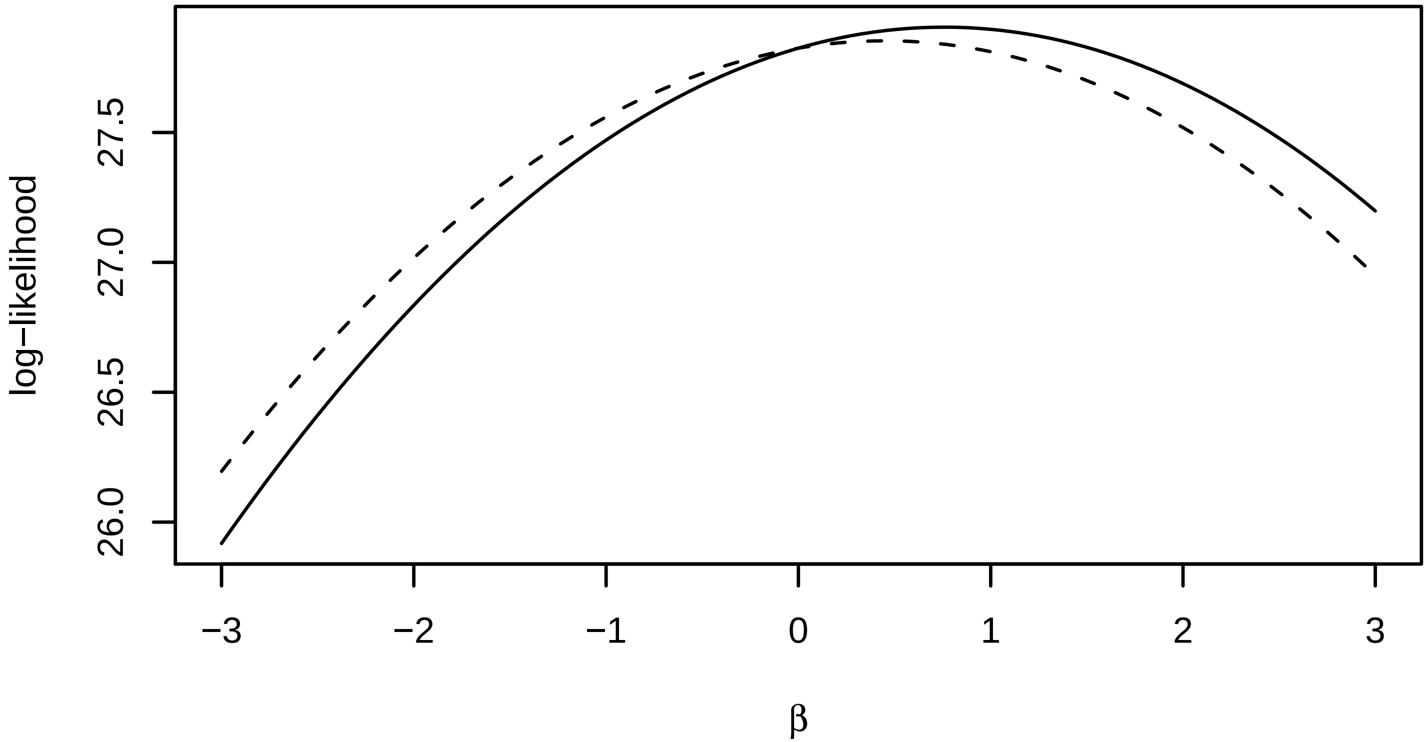
Aït-Sahalia process $\mathrm{d}X_t = (5 - 11X_t + 6X_t^2 - X_t^3)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 5$

Numerical instability. Up|Down $\Delta = 0.1|0.25$

Aït-Sahalia process $\mathrm{d}X_t = (5 - 11X_t + 6X_t^2 - X_t^3)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 5$



True likelihood (continuous line), Euler approximation (dashed line), Aït-Sahalia approximation (dotted line). Where is the dotted line? Coincides with the continuous line! Model $\mathrm{d}X_t = \beta X_t \mathrm{d}t + \mathrm{d}W_t$



no need to have Δ small, but (was) very difficult to implement!

The **sde** package implements Aït-Sahalia method. It also implements the following methods

- local Gaussian (`dcEuler`), Elerian (`dcElerian`), Ozaki (`dcOzaki`) and Shoji-Ozaki (`dcShoji`) approximations
- Simulated Likelihood Method (`dcSim`), Kessler's (`dcKessler`) and Aït-Sahalia (`HPloglik`) approximations

all of them can be passed to the `mle` function in R or used to build appropriate likelihood functions.

The **sde** package also implements many simulation schemes, including: Euler, Milstein, Milstein2, Elerian, Ozaki, Ozaki-Shoji, Exact Simulation Scheme, Simulation from conditional distribution, Predictor-Correction scheme, etc via the unique `sde.sim` function

```
sde.sim(t0 = 0, T = 1, X0 = 1, N = 100, delta, drift, sigma,
        drift.x, sigma.x, drift.xx, sigma.xx, drift.t,
        method = c("euler", "milstein", "KPS", "milstein2",
        "cdist","ozaki","shoji","EA"),
        alpha = 0.5, eta = 0.5, pred.corr = T, rcdist = NULL,
        theta = NULL, model = c("CIR", "VAS", "OU", "BS"),
        k1, k2, phi, max.psi = 1000, rh, A, M=1)
```

The `sde.sim` function

For the OU process, $\mathrm{d}X_t = -5X_t\mathrm{d}t + 3.5\mathrm{d}W_t$, it is as easy as

```
> d <- expression(-5 * x)
> s <- expression(3.5)
> sde.sim(X0=10,drift=d, sigma=s) -> X
> str(X)
 Time-Series [1:101] from 0 to 1: 10 9.32 8.79 8.89 8.48 ...
```

For the CIR model $\mathrm{d}X_t = (6 - 3X_t)\mathrm{d}t + 2\sqrt{X_t}\mathrm{d}W_t$

```
d <- expression( 6-3*x )
s <- expression( 2*sqrt(x) )
sde.sim(X0=10,drift=d, sigma=s) -> X
```

or, via model name

```
sde.sim(X0=10, theta=c(6, 3, 2), model="CIR") -> X
```

or, via exact conditional distribution `rcCIR` (also implemented in `sde`)

```
sde.sim(X0=10, theta=c(6, 3, 2), rcdist=rcCIR, method="cdist") -> X
```

Also in the `sde` package

The package also implements other estimation procedures

- estimating functions (linear, quadratic, martingale)
- GMM (but be careful, not really what you want to use with SDE!)
- approximate AIC statistics for model selection (`sdeAIC`)
- ϕ-divergence test statistics for parametric hypotheses testing (not in the book)
- change point (`cpoint`) analysis; both parametric and nonparametric
- non parametric estimation of drift (`ksdrift`) and diffusion (`ksdiff`) coefficients
- Markov Operator distance (`MOdist`) for clustering of SDE paths

The companion book: *Simulation and Inference for Stochastic Differential Equations*, with R Examples, Springer (2008).