Interval estimation of volatility function

Jan Mielniczuk, Paweł Teisseyre

Institute of Computer Science, Polish Academy of Science

October 1, 2009

1 Diffusion processes

2 Estimators of the drift and the volatility functions

3 Resampling methods

Definitions

Stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0 \tag{1}$$

W_t is a standard one dimensional Wiener process, starting from 0.

Drift function

$$\mu(X_t) = \lim_{\Delta \to 0} \Delta^{-1} E(X_{t+\Delta} - X_t | X_t)$$

Volatility function

$$\sigma^2(X_t) = \lim_{\Delta \to 0} \Delta^{-1} E((X_{t+\Delta} - X_t)^2 | X_t)$$

Applications

SDE models describe the dynamics of economic variables, e.g

1. stock prices,
2. market indexes,
3. exchange rates,
4. interest rates,
5. energy prices.

R package: SDE

Companion package to the book
S. M. Iacus, *Simulation and Inference for Stochastic Differential Equations with R examples*, Springer New York, 2008.

Example 1

Vasicek(Ornstein- Uhlenbeck) process

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t$$

R code

```
> library(sde)
> sde.sim(X0=alfa, theta=c(k*alfa, k, sigma),
  rcdist=rcOU, method="cdist")
```

Example 2

CIR process

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t$$

R code

```
> library(sde)
> sde.sim(X0=alfa, theta=c(k*alfa, k, sigma),
  rcdist=rcCIR, method="cdist")
```

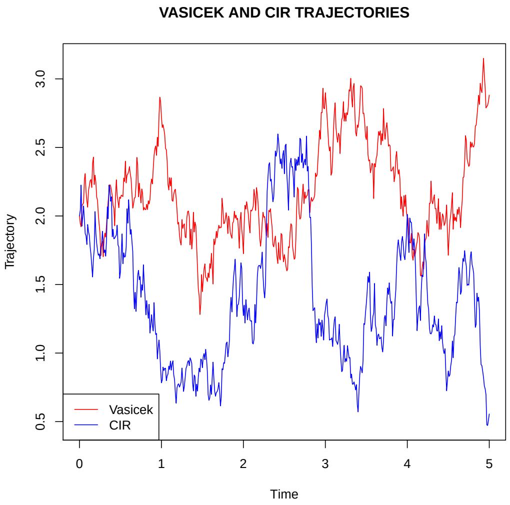
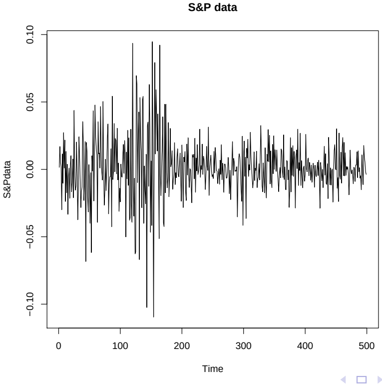


Figure: We considered the following parameters: $\kappa = 1$, $\alpha = 2$, $\sigma = 1$.

Example 3

Index S&P500

Market index published since 1957 of the prices of 500 common stocks actively traded in the United States.



Discretely sampled data

Suppose that we have observations $X_0, X_\Delta, \ldots, X_{n\Delta}$ from model (1), sampled at time points $\Delta, 2\Delta, \ldots, n\Delta$, for fixed $\Delta > 0$. For small Δ observations $X_0, X_\Delta, \ldots, X_{n\Delta}$ approximately satisfy the equation

Euler approximation

$$X_{(i+1)\Delta} - X_{i\Delta} = \mu(X_{i\Delta})\Delta + \sigma(X_{i\Delta})\sqrt{\Delta}\varepsilon_{i+1},$$

where $\{\varepsilon_i, i = 2, \ldots, n\}$ is a sequence of i.i.d. $N(0,1)$ random variables.

Estimators

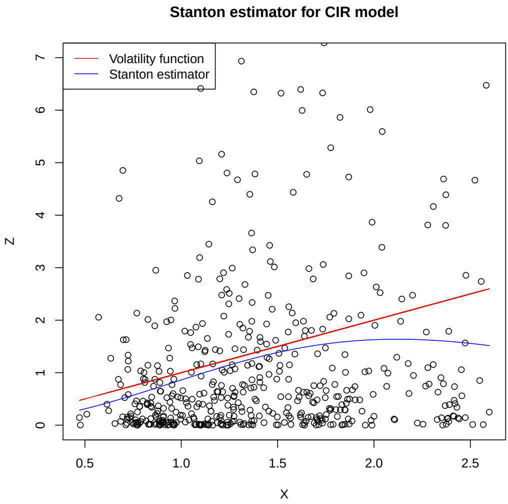
Define $Y_{i\Delta} := \Delta^{-1}(X_{(i+1)\Delta} - X_{i\Delta})$ and $Z_{i\Delta} := \Delta^{-1}(X_{(i+1)\Delta} - X_{i\Delta})^2$. Functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ can be regarded as the approximated regression functions of $(X_{i\Delta}, Y_{i\Delta})$ and $(X_{i\Delta}, Z_{i\Delta})$ respectively.

Stanton estimators

$$\hat{\mu}(x) = \frac{\sum_i Y_{i\Delta} K_h(X_{i\Delta} - x)}{\sum_i K_h(X_{i\Delta} - x)},$$

$$\hat{\sigma}^2(x) = \frac{\sum_i Z_{i\Delta} K_h(X_{i\Delta} - x)}{\sum_i K_h(X_{i\Delta} - x)},$$

where $K_h(u) := h^{-1}K(u/h)$, K is standard normal density, $h = h(n)$.


Stanton estimator for CIR model
Volatility function
Stanton estimator
Z
X
0
1
2
3
4
5
6
7
0.5
1.0
1.5
2.0
2.5

1. Kernel regression estimator (Stanton estimator)

R code

```
> library(np)
> bw = dpill(X,Z)
> npreg(X, Z, bws=bw,...)
```

2. Local linear estimator

R code

```
> library(KernSmooth)
> bw = dpill(X,Z)
> locpoly(X, Z, kernel="normal", banwidth=bw,...)
```

Assumptions

1. μ and σ are twice continously differentiable in a neighbourhood of x and satisfy Lipschitz condition on $\mathbb{R}$,
2. f_ε is bounded and satisfies Lipschitz condition on $\mathbb{R}$,
3. $\inf_{x\in\mathbb{R}} \sigma(x) > 0$ and a density of stationary distribution $f(x) > 0$,
4. K is symetric and bounded probability density having compact support,
5. $nh_n^5 \to C \geq 0$,
6. $\{X_i\}_{i\in\mathbb{Z}}$ is L^2 geometric moment contracting, i.e. $\|X_i - X_i'\| = \mathcal{O}(r^i)$ for some $0 < r < 1$, where $X_i' = J(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_i)$ and ε_0' is an independent copy of ε_0.

Assume that conditions 1-6 are satisfied. Then

Asymptotic normality of Stanton estimator

$$\sqrt{nh}(\hat{\sigma}^2(x) - \sigma^2(x) - \Delta\mu^2(x)) \xrightarrow{d} N\left(\sqrt{C}\,C_w, \frac{v(x)}{f(x)}\right),$$

where

$$v(x) = \mathbb{E}[2\mu(x)\sigma(x)\sqrt{\Delta}\varepsilon_{i+1} + (\varepsilon_{i+1}^2 - 1)\sigma^2(x)]^2 \int K^2(v)dv,$$

$$C_g = \int v^2 K(v)dv \cdot [f'(x)g'(x) + \frac{1}{2}f(x)g''(x)]/f(x),$$

$$w(x) = \Delta\mu^2(x) + \sigma^2(x).$$

Resampling methods

1. Construction interval estimates from asymptotic distribution of Stanton estimator is impossible.
2. To construct interval estimates of the volatility function we will use resampling methods.
3. The main aim of resampling is to construct several pseudo-samples with properties similar to the observed sample $X_1, \ldots, X_{n\Delta}$.

Resampling method 1

Recall that functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ can be regarded as the approximated regression functions of respectively $(X_{i\Delta}, Y_{i\Delta})$ and $(X_{i\Delta}, Z_{i\Delta})$, where $Y_{i\Delta} := \Delta^{-1}(X_{(i+1)\Delta} - X_{i\Delta})$ and $Z_{i\Delta} := \Delta^{-1}(X_{(i+1)\Delta} - X_{i\Delta})^2$.

Pair bootstrap

$$\{(X_{N_i\Delta}, Z_{N_i\Delta}), i = 1, \ldots, n-1\},$$

is generated, where $N_1, \ldots, N_n$ are i.i.d. random variables with uniform distribution on $\{1, \ldots, n-1\}$.

Resampling method 2

Autoregression bootstrap

$$X^*_{(i+1)\Delta} = \Delta\bar{\mu}(X^*_{i\Delta}) + X^*_{i\Delta} + \bar{\sigma}(X^*_{i\Delta})\varepsilon^*_{i+1}\sqrt{\Delta}, \quad i = 1, \ldots, n-1\,,$$

is generated with $X^*_\Delta = X_\Delta$, where $\bar{\mu}(\cdot)$ and $\bar{\sigma}(\cdot)$ are some estimators of $\mu(\cdot)$ and $\sigma(\cdot)$, respectively. The sequence $\{\varepsilon^*_i, i = 2, \ldots, n\}$ can be sampled randomly from $N(0, 1)$.

Resampling method 3

Subsampling

Data block $\mathcal{B}_{i,b} := (X_{i\Delta}, \ldots, X_{(i+b-1)\Delta})$ of size b,

where $i = 1, \ldots, n - b + 1$ can be interpreted as pseudosamples generated from original data.

Resampling method 3 (confidence interval)

Let $\hat{\sigma}^2_{i,b}(x)$ denote an estimator $\hat{\sigma}^2(x)$ computed from $\mathcal{B}_{i,b}$.

Theorem

Empirical distribution of $(bh_b)^{1/2}(\hat{\sigma}^2_{i,b}(x) - \hat{\sigma}^2(x))$ *approximates distribution of* $\sqrt{nh}(\hat{\sigma}^2(x) - \sigma^2(x) - \Delta\mu^2(x))$.

Approximate confidence interval for $\sigma^2(x)$

$$\left[\hat{\sigma}^2(x)(1+\eta_n) - \eta_n \sigma^{*2}_{1-\alpha/2}(x), \hat{\sigma}^2(x)(1+\eta_n) - \eta_n \sigma^{*2}_{\alpha/2}(x)\right],$$

where $\eta_n = (bh_b/nh_n)^{1/2}$ and $\sigma^{*2}_q(x)$ is a q^{th} empirical quantile of $\hat{\sigma}^2_{i,b}(x)$.

Resampling method 3 (optimal block size)

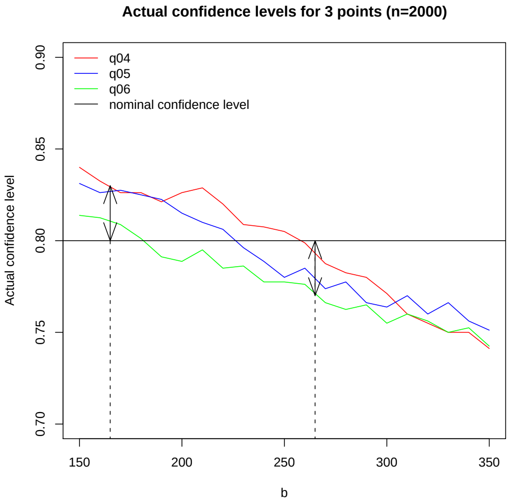
It is assumed that $b \longrightarrow \infty$ and $\frac{b}{n} \longrightarrow 0$ as $n \longrightarrow \infty$.

Problem

How to find an optimal block size $b_{opt} = b_{opt}(n)$?

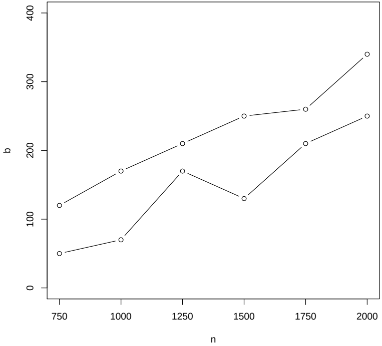
The goal is to find a relationship between b_{opt} and n.

Resampling method 3 (optimal block size)



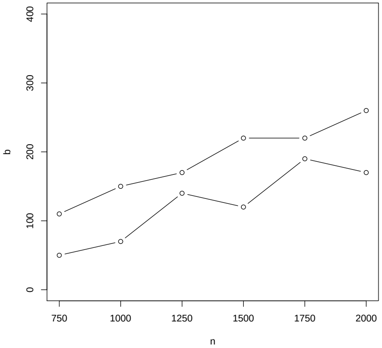
Resampling method 3 (optimal block size)

Region traced for Vasicek model.



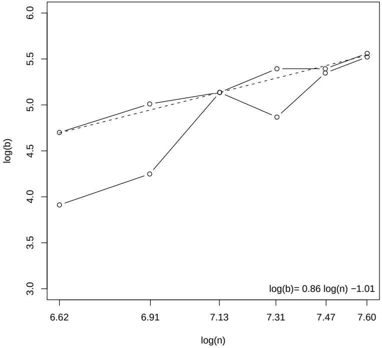
Resampling method 3 (optimal block size)

Region traced for CIR model.

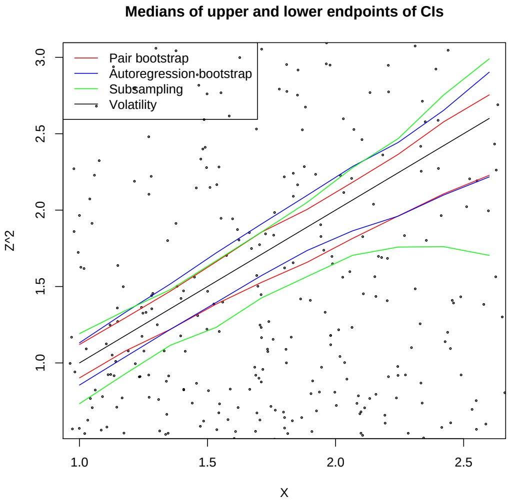


Resampling method 3 (optimal block size)

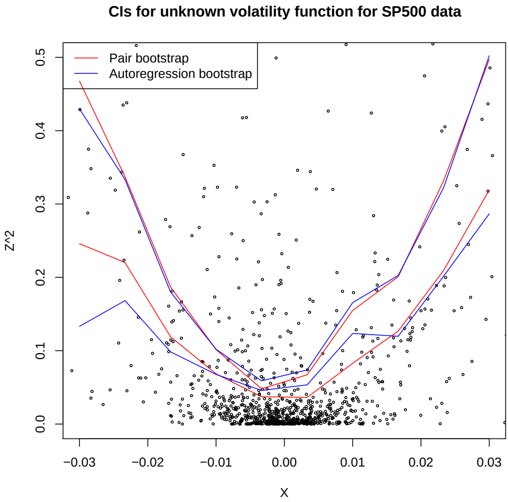
The common region traced for both models and the fitted line.



Example (Model CIR)



Example (S&P500 data)



Numerical results

Table: Coverage probabilities and width of confidence intervals for resampling methods for Vasicek model ($1 - \alpha = 0.8$)

	Autoregression bootstrap	Pair bootstrap	Subsampling
$n = 1000$			
$q_{0.4}$	0.78 (0.2)	0.78 (0.18)	0.81 (0.26)
$q_{0.5}$	0.76 (0.19)	0.81 (0.17)	0.8 (0.25)
$q_{0.6}$	0.77 (0.2)	0.79 (0.17)	0.8 (0.26)
$n = 1500$			
$q_{0.4}$	0.77 (0.15)	0.73 (0.13)	0.8 (0.19)
$q_{0.5}$	0.77 (0.15)	0.73 (0.13)	0.8 (0.19)
$q_{0.6}$	0.78 (0.15)	0.74 (0.13)	0.79 (0.19)
$n = 2000$			
$q_{0.4}$	0.79 (0.13)	0.71 (0.12)	0.81 (0.16)
$q_{0.5}$	0.77 (0.12)	0.74 (0.11)	0.82 (0.15)
$q_{0.6}$	0.78 (0.13)	0.74 (0.12)	0.81 (0.16)

Numerical results

Table: Coverage probabilities and width of confidence intervals for resampling methods for CIR model ($1 - \alpha = 0.8$)

	Autoregression bootstrap	Pair bootstrap	Subsampling
$n = 1000$			
$q_{0.4}$	0.77 (0.32)	0.76 (0.28)	0.81 (0.40)
$q_{0.5}$	0.76 (0.36)	0.76 (0.33)	0.81 (0.47)
$q_{0.6}$	0.76 (0.34)	0.76 (0.4)	0.78 (0.56)
$n = 1500$			
$q_{0.4}$	0.78 (0.24)	0.77 (0.22)	0.8 (0.30)
$q_{0.5}$	0.77 (0.28)	0.76 (0.26)	0.79 (0.34)
$q_{0.6}$	0.76 (0.33)	0.75 (0.31)	0.77 (0.41)
$n = 2000$			
$q_{0.4}$	0.79 (0.21)	0.77 (0.19)	0.81 (0.25)
$q_{0.5}$	0.76 (0.24)	0.77 (0.22)	0.78 (0.29)
$q_{0.6}$	0.77 (0.29)	0.75 (0.27)	0.78 (0.35)

Some references

1. J. Fan, *A Selective Overview of Nonparametric Methods in Financial Econometrics*, Statistical Science, Vol. 20, No. 4, pages 317–337, 2005.
2. D. N. Politis, J. P. Romano, M. Wolf, *Subsampling*, Springer New York, 1999.
3. J. Franke, J. P. Kreiss, E. Mammen, *Bootstrap of kernel smoothing in nonlinear time series*, Bernoulli, Vol. 8, No. 1, pages 1–37, 2002.
4. E. Flaichaire, *A better way to bootstrap pairs*, Economic Letters, Vo. 64, pages 257–262, 1999.