Bio-Math Summer School 2008

Middelfart, Denmark, August 2008

Estimation for diffusion-type processes

Michael Sørensen
University of Copenhagen, Denmark

http://www.math.ku.dk/∼michael

Discretely observed diffusion

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

X, b and W d-dimensional, σ $d \times d$-matrix

State space: $D \subseteq \mathbb{R}^d$

For $d = 1$, $D = (\ell, r)$, $-\infty \le \ell < r \le \infty$

Data: $X_{t_0}, \cdots, X_{t_n}$, $\Delta_i = t_i - t_{i-1}$

Often $t_i = \Delta i$

Estimating functions

$$G_n(\theta) = \sum_{i=1}^n g(\Delta_i, X_{t_i}, X_{t_{i-1}};\theta)$$

g p-dimensional

$$G_n(\theta) = \frac{1}{n}\sum_{i=1}^n g(\Delta_i, X_{t_i}, X_{t_{i-1}};\theta)$$

G_n–estimator(s): $G_n(\hat\theta_n) = 0$

To obtain consistent estimators, we assume that

$$E_\theta(g(\Delta_i, X_{t_i}, X_{t_{i-1}};\theta)) = 0$$

Martingale estimating functions

$$G_n(\theta) = \sum_{i=1}^n g(\Delta_i, X_{t_i}, X_{t_{i-1}};\theta),$$

$$g(\Delta, y, x;\theta) \quad = \quad \sum_{j=1}^N a_j(x,\Delta;\theta)[f_j(y) - \pi_\theta^\Delta f_j(x)]$$

↑ p-dimensional ↑ real valued

Transition operator: $\pi_\theta^\Delta f(x) = E_\theta(f(X_\Delta) \mid X_0 = x)$

$G_n(\theta)$ is a P_θ-martingale:

$$E_\theta(a_j(X_{t_{i-1}}, \Delta_i;\theta)[f_j(X_{t_i}) - \pi_\theta^{\Delta_i} f_j(X_{t_{i-1}})] \mid X_{t_1}, \cdots, X_{t_{i-1}}) = 0$$

Martingale estimating functions

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta),$$

$$g(\Delta, y, x; \theta) = \sum_{j=1}^{N} a_j(x, \Delta; \theta)[f_j(y) - \pi_\theta^\Delta f_j(x)]$$

- Easy asymptotics
- Simple expression for optimal estimating function
- The score function is a P_θ-martingale

Quadratic martingale estimating function

$$G_n(\theta) = \sum_{i=1}^{n} \Big\{ a_1(X_{t_{i-1}}, \Delta_i; \theta)(X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta))$$

$$+ a_2(X_{t_{i-1}}, \Delta_i; \theta) \left[(X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta))^2 - \Phi(\Delta_i, X_{t_{i-1}}; \theta) \right] \Big\}$$

$F(\Delta, x; \theta) = E_\theta(X_\Delta | X_0 = x)$ and $\Phi(\Delta, x; \theta) = \mathrm{Var}_\theta(X_\Delta | X_0 = x)$

Bibby and Sørensen (1995,1996)

A good simple choice of the weights

$a_1(x, \Delta; \theta) = \partial_\theta b(x; \theta)/\sigma^2(x; \theta)$ $\qquad$ $a_2(x, \Delta; \theta) = \partial_\theta v(x; \theta)/\sigma^4(x; \theta)$

Simulation of conditional moments

$\pi_\theta^\Delta f(x) = E_\theta(f(X_\Delta) | X_0 = x)$ is usually not explicitly known

Fix θ

Simulate numerically M independent trajectories of $\{X_t : t \in [0, \Delta]\}$ with $X_0 = x$

$$\pi_\theta^\Delta f(x) \doteq \frac{1}{M} \sum_{i=1}^{M} f(X_\Delta^{(i)})$$

Kloeden and Platen (1992)

Beskos, Papaspiliopoulos and Roberts (2006, 2007)

Variance reduction methods

Linear drift

$$X_t = X_0 - \int_0^t \beta(X_s - \mu) ds + \int_0^t \sigma(X_s; \theta) dW_s$$

$d = 1$

$$f(t) = E_\theta(X_t \,|\, X_0 = x) = x - \beta \int_0^t (f(s) - \mu) ds$$

provided $\int_0^t \sigma(X_s; \theta) dW_s$ is a proper martingale.

$$f'(t) = -\beta(f(t) - \mu), \quad f(0) = x$$

$$f(t) = x e^{-\beta t} + \mu(1 - e^{-\beta t})$$

Explicit martingale estimating functions

Kessler and Sørensen (1999)

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t$$

Generator:

$$L_\theta = \tfrac{1}{2}\sigma^2(x;\theta)\frac{d^2}{dx^2} + b(x;\theta)\frac{d}{dx},$$

φ eigenfunction for L_θ:

$$L_\theta\varphi = -\lambda_\theta\varphi$$

Under weak regularity conditions

$$\pi_\theta^\Delta\varphi(x) = E_\theta(\varphi(X_\Delta)|X_0 = x) = e^{-\lambda_\theta\Delta}\varphi(x)$$

i.e. φ is an eigenfunction for π_θ^Δ

Explicit martingale estimating functions

$Y_t = e^{\lambda_\theta t}\varphi(X_t)$

By Ito's formula

$$\begin{aligned} Y_t &= Y_0 + \int_0^t e^{\lambda_\theta s}\left[L_\theta\varphi(X_s) + \lambda_\theta\varphi(X_s)\right]ds + \int_0^t e^{\lambda_\theta s}\varphi'(X_s)\sigma(X_s;\theta)dW_s \\ &= Y_0 + \int_0^t e^{\lambda_\theta s}\varphi'(X_s)\sigma(X_s;\theta)dW_s \end{aligned}$$

Thus Y_t is a P_θ-martingale when $\int_0^t e^{\lambda_\theta s}\varphi'(X_s)\sigma(X_s;\theta)dW_s$ is a proper martingale

$$E_\theta(e^{\lambda_\theta t}\varphi(X_t)\,|\,X_0 = x) = \varphi(x)$$

Y is a martingale if for all $t > 0$

$$\int_0^t E\left(\varphi'(X_s)^2\sigma^2(X_s)\right)ds < \infty$$

Explicit martingale estimating functions

Three sets of sufficient conditions ensuring that $\pi_\theta^\Delta\varphi(x) = e^{-\lambda_\theta\Delta}\varphi(x)$:

(i) X ergodic with invariant measure μ, and $\int\varphi'(x)^2\sigma^2(x)\mu(dx) < \infty$

(ii) σ and φ' bounded

(iii) b and σ of linear growth, φ' of polynomial growth

Example

$$dX_t = -\theta\tan(X_t)dt + dW_t\ ,\theta > 0.$$

For $\theta \geq 1/2$: ergodic diffusion on the interval $(-\pi/2, \pi/2)$
The density of the invariant measure proportional to $\cos(x)^{2\theta}$

$$L_\theta\sin(x) = -(\theta + 1/2)\sin(x)$$

$$G_n(\theta) = \sum_{i=1}^n \sin(X_{t_{i-1}})\left[\sin(X_{t_i}) - e^{-(\theta+1/2)\Delta}\sin(X_{t_{i-1}})\right]$$

$$\hat\theta_n = -\Delta^{-1}\log\left(\frac{\sum_{i=1}^n\sin(X_{t_{i-1}})\sin(X_{t_i})}{\sum_{i=1}^n\sin^2(X_{t_{i-1}})}\right) - 1/2,$$

provided that $\sum_{i=1}^n\sin(X_{t_{i-1}})\sin(X_{t_i}) > 0$

Pearson diffusions

Wong (1964), Forman & Sørensen (2008)

$$dX_t = -\beta(X_t - \mu)dt + \sqrt{2\beta(aX_t^2 + bX_t + c)}dW_t, \qquad \beta > 0, \qquad d = 1$$

$$L\varphi = \beta(ax^2 + bx + c)\varphi'' + \beta(x - \mu)\varphi'$$

If φ is a polynomial of order k, then so is $L\varphi$

Thus we can find eigenfunctions that are explicit polynomials

The class of possible stationary marginal distributions is equal to Pearson's system of distributions

$Y_t = aX_t + b$ is also a Pearson diffusion

Up to location-scale transformations the following is a complete list

Pearson diffusions

- Normal distribution:
  Ornstein-Uhlenbeck process:

$$dX_t = -\beta(X_t - \mu)dt + \sqrt{2\beta c}\, dW_t, \quad c > 0$$

  $X_t \sim N(\mu, c)$
  State space: the real line
  Eigenfunctions: Hermite polynomials

- Gamma-distribution:
  Square root process:

$$dX_t = -\beta(X_t - \mu)dt + \sqrt{2\beta b X_t}\, dW_t, \quad b > 0$$

  X_t gamma-distributed with mean μ and scale parameter b
  State space: the positive real axis
  Eigenfunctions: Laguerre polynomials

Pearson diffusions

- Beta-distribution:
  Jacobi diffusions:

$$dX_t = -\beta(X_t - \mu)dt + \sqrt{2\beta a X_t(1 - X_t)}\, dW_t, \quad a > 0$$

  X_t beta-distributed with $p(x) \propto x^{\mu/a-1}(1-x)^{(1-\mu/a)-1}$
  State space: the interval $(0, 1)$
  Eigenfunctions: Jacobi polynomials

- Inverse gamma distribution:

$$dX_t = -\beta(X_t - \mu)dt + \sqrt{2\beta a}\, X_t\, dW_t, \quad a > 0$$

  Density of X_t: $\quad p(x) \propto x^{-(a^{-1}+2)} \exp(-\frac{\mu}{ax})$
  State space: the positive real axis
  Eigenfunctions: Bessel polynomials

Pearson diffusions

- F-distribution:

$$dX_t = -\beta(X_t - \mu)dt + \sqrt{2\beta a X_t(X_t + 1)}\, dW_t, \quad a > 0$$

  $(1 + a)\mu^{-1}X_t$ F-distributed with $2\mu a^{-1}$ and $2a^{-1} + 2$ degrees of freedom
  State space: the positive real axis

- t-distribution with $1 + 1/a$ degrees of freedom:

$$dX_t = -\beta(X_t - \mu)dt + \sqrt{2\beta a(X_t^2 + 1)}\, dW_t, \quad a > 0$$

  X_t is t-distribution with $1 + 1/a$ degrees of freedom and mean μ
  State space: the real line

Pearson diffusions

- Pearson's type IV distribution, a skew t-distribution

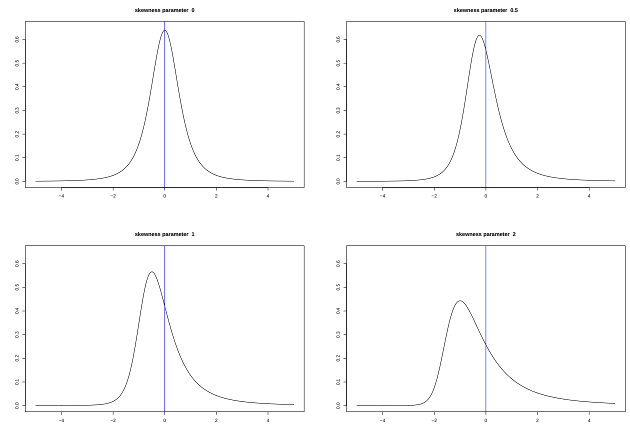
$$dZ_t = -\beta Z_t dt + \sqrt{2\beta(\nu-1)^{-1}\{Z_t^2 + 2\rho\nu^{\frac{1}{2}} Z_t + (1+\rho^2)\nu\}} dW_t, \quad \nu > 1$$

$$p(z) \propto \{(z/\sqrt{\nu}+\rho)^2+1\}^{-(\nu+1)/2} \exp\left\{\rho(\nu+1)\tan^{-1}\left(z/\sqrt{\nu}+\rho\right)\right\}$$

An expression for the normalizing constant when $\nu \in \mathbb{N}$ can be found in Nagahara (1996)

$\rho = 0$: t-distribution with ν degrees of freedom

Pearson's type IV distribution



Densities of skew t-distributions (Pearson's type IV distributions) with zero mean for $\rho = 0$, 0.5, 1, and 2 respectively

Pearson diffusions: moments

$$E(X_t^n \mid X_0 = x) = \sum_{k=0}^{n}\sum_{\ell=0}^{n} q_{n,k,\ell} \cdot e^{-\lambda_\ell t} \cdot x^k$$

$q_{n,k,n} = p_{n,k}$, $q_{n,n,\ell} = 0$ for $\ell \leq n-1$,

$$q_{n,k,\ell} = -\sum_{j=k\vee\ell}^{n-1} p_{n,j} q_{j,k,\ell} \qquad \text{for } \; k,\ell = 0,\ldots,n-1$$

λ_ℓ and $p_{n,j}$ eigenvalues and coefficients of eigenfunction polynomials

$$E(X_t^n) = a_n^{-1}\{b_n \cdot E(X_t^{n-1}) + c_n \cdot E(X_t^{n-2})\}$$

with $E(X_t^0) = 1$, and $E(X_t) = \mu$

$a_n = \lambda_n = n\{1-(n-1)a\}\beta, \quad b_n = n\{\alpha + (n-1)b\}\beta,$
$c_n = n(n-1)c\beta$ for $n = 0, 1, 2, \ldots$

Transformations of Pearson diffusions

X_t: $\varphi(x)$ eigenfunction with eigenvalue λ

T an injection

$Y_t = T(X_t):$ $\varphi(T^{-1}(y))$ eigenfunction with eigenvalue λ

- Find the stochastic differential equation for $T(X_t)$ by Ito's formula.
- $L_Y \psi(y) =$

$$\beta\left[T'(T^{-1}(y))(\mu - T^{-1}(y)) + T''(T^{-1}(y))(aT^{-1}(y)^2 + bT^{-1}(y) + c)\right]\psi'(y)$$
$$+ \beta T'(T^{-1}(y))^2[aT^{-1}(y)^2 + bT^{-1}(y) + c]\psi''(y)$$

- Check that $\psi(y) = \varphi(T^{-1}(y))$ is an eigenfunction for L

Transformations of Pearson diffusions

Jacobi diffusion state space $(-1,1)$, $\beta,\sigma>0$, $\gamma\in(-1,1)$

$$dX_t = -\beta[X_t-\gamma]dt + \sigma\sqrt{1-X_t^2}dW_t$$

Eigenfunctions: $P_n^{(\beta(1-\gamma)\sigma^{-2}-1,\,\beta(1+\gamma)\sigma^{-2}-1)}(x)$

$P_n^{(a,b)}(x)$ denotes the Jacobi polynomial of order n

$Y_t=\sin^{-1}(X_t)$ state space $(-\frac{\pi}{2},\frac{\pi}{2})$, $\rho=\beta-\frac{1}{2}\sigma^2$, $\varphi=\beta\gamma/(\beta-\frac{1}{2}\sigma^2)$

$$dY_t = -\rho\frac{\sin(Y_t)-\varphi}{\cos(Y_t)}dt + \sigma d\tilde{W}_t$$

Eigenfunctions: $P_n^{(\rho(1-\varphi)\sigma^{-2}-\frac{1}{2},\,\rho(1+\varphi)\sigma^{-2}-\frac{1}{2})}(\sin(x))$

Transformations of Pearson diffusions

Jacobi diffusion state space $(0,1)$, $\beta,\sigma>0$, $\mu\in(0,1)$

$$dX_t = -\beta(X_t-\mu)dt + \sqrt{2\beta a X_t(1-X_t)}\,dW_t$$

Eigenfunctions: Jacobi polynomials $P_n(x)$ on $(0,1)$

$Y_t=\log(X_t/(1-X_t))$ state space $(\mathbb{R}$,

$$dY_t = -\beta\left\{1-2\mu+(1-\mu)e^{Y_t}-\mu e^{-1}-8a\cosh^4(X_t/2)\right\}dt + 2\sqrt{-a\beta}\cosh(X_t/2)dW_t,$$

Eigenfunctions: $P_n(e^x/(1+e^x))$

Invariant probability distribution:

$$p(x)=\frac{e^{\kappa_1 x}}{(1+e^x)^{\kappa_1+\kappa_2}B(\kappa_1,\kappa_2)},\quad x\in\mathbb{R},$$

Generalized logistic distribution, Barndorff-Nielsen, Kent and Sørensen (1982).

Ergodic 1-dimensional diffusions

$$dX_t=b(X_t;\theta)dt+\sigma(X_t;\theta)dW_t \qquad v(x;\theta)=\sigma^2(x;\theta)$$

state space: (ℓ,r) $-\infty\le\ell<r\le\infty$

Assume that

$$\int_{x^\#}^r s(x;\theta)dx=\int_\ell^{x^\#}s(x;\theta)dx=\infty \quad\text{and}\quad \int_\ell^r\tilde{\mu}_\theta(x)dx=A(\theta)<\infty,$$

where $x^\#$ is an arbitrary point in (ℓ,r)

$$s(x;\theta)=\exp\left(-2\int_{x^\#}^x\frac{b(y;\theta)}{v(y;\theta)}dy\right) \quad\text{and}\quad \tilde{\mu}_\theta(x)=[s(x;\theta)v(x;\theta)]^{-1}$$

Scale-measure Speed-measure

Then X is ergodic with stationary distribution $\mu_\theta(x)=\tilde{\mu}_\theta(x)/A(\theta)$

Ergodic 1-dimensional diffusions

$X_t\xrightarrow{\mathcal{D}}\mu_\theta$ as $t\to\infty$

If $X_0\sim\mu_\theta$, then X is stationary and $X_t\sim\mu_\theta$ for all $t\ge0$

$$Q_\theta^\Delta(x,y)=\mu_\theta(x)p(\Delta,x,y;\theta)$$

$y\mapsto p(\Delta,x,y;\theta)$ is the transition density, i.e. the probability density function of the conditional distribution of $X_{t+\Delta}$ given that $X_t=x$ under P_θ

If $X_0\sim\mu_\theta$, then $(X_t,X_{t+\Delta})\sim Q_\theta^\Delta$ for all $t\ge0$

$f:(\ell,r)^2\mapsto\mathbb{R}$

$$Q_\theta^\Delta(f)=\int_{(\ell,r)^2}f(x,y)p(\Delta,x,y;\theta)\mu_\theta(x)dydx$$

Limit theorems

Law of large numbers: Suppose X is ergodic, and let f be a mapping $(l,r)^2 \mapsto \mathbb{R}$ satisfying that $Q_\theta^\Delta(f^2) < \infty$. Then

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}) \xrightarrow{P_\theta} Q_\theta^\Delta(f)$$

as $n \to \infty$

Martingale central limit theorem: Suppose further that

$$\int_\ell^r f(x,y)p(\Delta,x,y;\theta)dy = 0$$

for all $x \in (l,r)$. Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}) \xrightarrow{\mathcal{D}} N\left(0, Q_\theta^\Delta(f^2)\right)$$

under P_θ as $n \to \infty$. Billingsley (1961)

Uniform law of large numbers

Suppose that X is ergodic under P, so that there exists a probability measure Q^Δ on $(l,r)^2$ such that for any function $f : (l,r)^2 \mapsto \mathbb{R}$

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}) \xrightarrow{P} Q^\Delta(f)$$

as $n \to \infty$

Let the function $f : (l,r)^2 \times \Theta \mapsto \mathbb{R}$ be a continuous function of θ and locally dominated square integrable with respect to Q^Δ. Then $\theta \mapsto Q^\Delta(f(\theta))$ is continuous, and for any compact subset $M \subseteq \Theta$

$$\sup_{\theta\in M}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta};\theta) - Q^\Delta(f(\theta))\right| \xrightarrow{P} 0$$

as $n \to \infty$

Low frequency asymptotics

Data: $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$

$n \to \infty$ Δ fixed (and suppressed in $p(x,y;\theta)$ and $g(x,y;\theta)$)

θ_0: true parameter value $\theta \in \Theta \subseteq \mathbb{R}^p$

Martingale estimating functions:

$$G_n(\theta) = \sum_{i=1}^{n} g(X_{(i-1)\Delta}, X_{i\Delta};\theta)$$

g is p-dimensional

$$\int_D g(x,y;\theta)p(x,y;\theta)dy = 0 \quad \text{for all } x \in D \text{ and } \theta \in \Theta$$

Low frequency asymptotics

CONDITION LFA1:

$\theta_0 \in$ *int* Θ and there exits a neighbourhood N of θ_0 in Θ, such that:

(1) The function $g(\theta) : (x,y) \mapsto g(x,y;\theta)$ is integrable with respect to Q_{θ_0} given for all $\theta \in N$, and $g(\theta_0)$ is square integrable with respect to Q_{θ_0}

(2) The function $\theta \mapsto g(x,y;\theta)$ is continuously differentiable on N for all $(x,y) \in D^2$

(3) The functions $(x,y) \mapsto \partial_{\theta_j} g_i(x,y;\theta),\ i,j = 1,\ldots,p,$ are dominated for all $\theta \in N$ by a function which is integrable with respect to Q_{θ_0}

(4) The $p \times p$ matrix

$$W = Q_{\theta_0}\left(\partial_{\theta^T} g(\theta_0)\right)$$

is invertible

Low frequency asymptotics

CONDITION LFA2:

The functions $(x,y) \mapsto g(x,y;\theta)$ are locally dominated integrable with respect to Q_{θ_0} and $Q_{\theta_0}(g(\theta)) \neq 0$ for all $\theta \neq \theta_0$

THEOREM:

Suppose Condition LFA1 holds. Then a sequence $(\hat{\theta}_n)$ of weakly consistent G_n–estimators exists, and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N_p\left(0, W^{-1}V(W^T)^{-1}\right)$$

under P_{θ_0}, where

$$V = Q_{\theta_0}\left(g(\theta_0)g(\theta_0)^T\right).$$

If, moreover, Condition LFA2 holds, then the sequence of estimators $(\hat{\theta}_n)$ is eventually unique on any bounded subset of Θ containing θ_0

Low frequency asymptotics

THEOREM:

Under Conditions LFA1 (2) – (4)

$$W_n = \frac{1}{n}\sum_{i=1}^{n} \partial_{\theta^T} g(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_n) \xrightarrow{P_{\theta_0}} W,$$

and the probability that W_n is invertible tends to one as $n \to \infty$.

If it is, moreover, assumed that the functions $(x,y) \mapsto g_i(x,y;\theta)$, $i = 1,\ldots,p,$ are dominated for all $\theta \in N$ by a function which is square integrable with respect to Q_{θ_0}, then

$$V_n = \frac{1}{n}\sum_{i=1}^{n} g(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_n) g(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_n)^T \xrightarrow{P_{\theta_0}} V.$$

LF-Asymptotics: likelihood inference

$$U_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p(X_{(i-1)\Delta}, X_{i\Delta}; \theta),$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N_p\left(0, \mathcal{I}(\theta_0)^{-1}\right)$$

$$\mathcal{I}(\theta) = -Q_\theta\left(\partial_\theta^2 \log p(\theta)\right) \qquad \text{Fisher information}$$

Condition for eventual uniqueness:

$$Q_{\theta_0}(\partial_\theta \log p(x,y,\theta)) = -\partial_\theta \bar{K}(\theta,\theta_0) \neq 0 \quad \text{for all} \quad \theta \neq \theta_0$$

$$\bar{K}(\theta,\theta_0) = \int_D K(\theta,\theta_0;x)\,\mu_{\theta_0}(dx),$$

$$K(\theta,\theta_0;x) = \int_D \log[p(x,y;\theta_0)/p(x,y;\theta)]p(x,y;\theta_0)\,dy$$

is the Kullback-Leibler divergence between the two transition distributions

LF-Asymptotics: misspecified models

Suppose that X is ergodic under P, so that there exists a probability measure Q on D^2 such that

$$\frac{1}{n}\sum_{i=1}^{n} g(X_{(i-1)\Delta}, X_{i\Delta}; \theta) \xrightarrow{P} Q(g(\theta)) \qquad \text{as } n \to \infty$$

$$Q(g(\bar{\theta})) = 0$$

$$\hat{\theta}_n \xrightarrow{P} \bar{\theta} \qquad \sqrt{n}(\hat{\theta}_n - \bar{\theta}) \xrightarrow{\mathcal{D}} N_p\left(0, W^{-1}V(W^T)^{-1}\right)$$

$$W = Q\left(\partial_{\theta^T} g(\bar{\theta})\right) \qquad V = Q\left(g(\bar{\theta})g(\bar{\theta})^T\right) \quad \text{or more complicated}$$

Optimal estimating functions

Statistical model: $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, $\Theta \subseteq \mathbb{R}^p$

Data: $X_1, X_2, \ldots, X_n$

Estimating function (p-dimensional): $G_n(\theta) = G_n(\theta; X_1, X_2, \ldots, X_n)$

Unbiased: $E_\theta(G_n(\theta)) = 0$ for all $\theta \in \Theta$

Class of unbiased, square integrable estimating functions: $\mathcal{G}$

How do we choose a good $G_n \in \mathcal{G}$?

Godambe (1960), Durbin (1960), Godambe (1985), Godambe and Heyde (1987), Heyde (1997)

Score-function: $U_n(\theta) = \partial_\theta \log L_n(\theta)$

Likelihood function: $L_n(\theta) = f(X_1, \ldots, X_n; \theta)$

$f(x_1, \ldots, x_n; \theta) > 0$ density of $(X_1, \ldots, X_n)$ under P_θ w.r.t. a dominating measure

Godambe optimality

$p = 1$

Sensitivity: $S_{G_n}(\theta) = E_\theta(\partial_\theta G_n(\theta)) = -\text{Cov}_\theta(G_n(\theta), U_n(\theta))$

Godambe information: $K_{G_n}(\theta) = S_{G_n}(\theta)^2 / \text{Var}_\theta(G_n(\theta))$

$G^* \in \mathcal{G}$ is *Godambe-optimal* in $\mathcal{G}$ if $K_{G_n^*}(\theta) \geq K_{G_n}(\theta)$ for all $\theta \in \Theta$ and for all $G_n \in \mathcal{G}$

$p \geq 1$

$$S_{G_n}(\theta) = E_\theta(\partial_{\theta^T} G_n(\theta))$$

$$K_{G_n}(\theta) = S_{G_n}(\theta)^T E_\theta\left(G_n(\theta) G_n(\theta)^T\right)^{-1} S_{G_n}(\theta)$$

Godambe optimality

Standardized version of $G_n(\theta)$:

$$G_n^{(s)}(\theta) = -S_{G_n}(\theta)^T E_\theta\left(G_n(\theta) G_n(\theta)^T\right)^{-1} G_n(\theta)$$

Second Bartlett-identity: $E_\theta\left(G_n^{(s)}(\theta) G_n^{(s)}(\theta)^T\right) = -E_\theta(\partial_{\theta^T} G_n^{(s)}(\theta))$

$$E_\theta\left((G_n^{(s)}(\theta) - U_n(\theta))^T (G_n^{(s)}(\theta) - U_n(\theta))\right) \geq$$
$$E_\theta\left((G_n^{*(s)}(\theta) - U_n(\theta))^T (G_n^{*(s)}(\theta) - U_n(\theta))\right)$$

THEOREM: G_n^* is Godambe-optimal in $\mathcal{G}$ if

$$S_{G_n}(\theta)^{-1} E_\theta\left(G_n(\theta) G_n^*(\theta)^T\right) = S_{G_n^*}(\theta)^{-1} E_\theta\left(G_n^*(\theta) G_n^*(\theta)^T\right)$$

for all $\theta \in \Theta$ and for all $G_n \in \mathcal{G}$

Heyde optimality

$G_n(\theta)$ is a P_θ-martingale w.r.t. $\{\mathcal{F}_n\}$ for every $\theta \in \Theta$, $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$

$$E_\theta(H_i(\theta) \,|\, \mathcal{F}_{i-1}) = 0 \qquad H_i(\theta) = G_i(\theta) - G_{i-1}(\theta)$$

$$\bar{G}_n(\theta) = \sum_{i=1}^n E_\theta\left(\partial_{\theta^T} H_i(\theta) | \mathcal{F}_{i-1}\right)$$ compensator of $\partial_{\theta^T} G_n(\theta)$

$$\langle G(\theta), \tilde{G}(\theta) \rangle_n = \sum_{i=1}^n E_\theta\left(H_i(\theta) \tilde{H}_i(\theta)^T | \mathcal{F}_{i-1}\right)$$ quadratic co-characteristic

$\langle G(\theta) \rangle_n = \langle G(\theta), G(\theta) \rangle_n$

Under usual regularity conditions:

$$\langle G(\theta) \rangle_n^{-\frac{1}{2}} \bar{G}_n(\theta) (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, I_p)$$

Heyde information: $I_{G_n}(\theta) = \bar{G}_n(\theta)^T \langle G(\theta) \rangle_n^{-1} \bar{G}_n(\theta)$

Heyde optimality

$$I_{G_n}(\theta) = \bar{G}_n(\theta)^T \langle G(\theta)\rangle_n^{-1} \bar{G}_n(\theta)$$

$G^* \in \mathcal{G}$ is *Heyde-optimal* in $\mathcal{G}$ if $I_{G_n^*}(\theta) \geq I_{G_n}(\theta)$ for all $\theta \in \Theta$, all $n \in \mathbb{N}$, and all $G \in \mathcal{G}$

$$E_\theta\left(\bar{G}_n(\theta)\right) = S_{G_n}(\theta) \quad \text{and} \quad E_\theta\left(\langle G(\theta)\rangle_n\right) = E_\theta\left(G_n(\theta)G_n(\theta)^T\right)$$

THEOREM: G^* is Heyde-optimal in $\mathcal{G}$ if

$$\bar{G}_n(\theta)^{-1}\langle G(\theta), G^*(\theta)\rangle_n = \bar{G}_n^*(\theta)^{-1}\langle G^*(\theta)\rangle_n$$

for all $\theta \in \Theta$, all $n \in \mathbb{N}$, and all $G \in \mathcal{G}$

If $\bar{G}_n^*(\theta)^{-1}\langle G^*(\theta)\rangle_n$ is non-random, then G^* is also Godambe-optimal in $\mathcal{G}$

Optimal estimating functions for SDEs

$$G_n(\theta) = \sum_{i=1}^n g(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta),$$

$$g(\Delta, y, x; \theta) \quad = \quad \sum_{j=1}^N a_j(x, \Delta; \theta) h_j(\Delta, X_{t_i}, X_{t_{i-1}}; \theta)$$

$\uparrow$ p-dimensional $\qquad$ $\uparrow$ real valued

$$G_n(\theta) = \sum_{i=1}^n A(X_{t_{i-1}}, \Delta_i; \theta) h(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta)$$

$h = (h_1, \ldots, h_N)^T$

$A(x, \Delta; \theta)$ $p \times N$-matrix of weights

$\int_D h(\Delta, x, y; \theta) p(\Delta, x, y; \theta) dy = 0$ for all $x \in D$, $\Delta > 0$ and $\theta \in \Theta$

Optimal estimating functions for SDEs

$$G_n(\theta) = \sum_{i=1}^n A(X_{t_{i-1}}, \Delta_i; \theta) h(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta)$$

$$\bar{G}_n(\theta) = \sum_{i=1}^n A(X_{t_{i-1}}, \Delta_i; \theta) E_\theta(\partial_{\theta^T} h(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta) \mid X_{t_{i-1}})$$

$$\langle G(\theta), G^*(\theta)\rangle_n = \sum_{i=1}^n A(X_{t_{i-1}}, \Delta_i; \theta) V_h(X_{t_{i-1}}, \Delta_i; \theta) A^*(X_{t_{i-1}}, \Delta_i; \theta)^T$$

$$V_h(x, \Delta; \theta) = E_\theta\left(h(\Delta, X_\Delta, x; \theta) h(\Delta, X_\Delta, x; \theta)^T \mid X_0 = x\right)$$

$$B_h(x, \Delta; \theta) = E_\theta\left(\partial_{\theta^T} h(\Delta, X_\Delta, x; \theta) \mid X_0 = x\right)$$

$$B_h(x, \Delta; \theta) = V_h(x, \Delta; \theta) A^*(x, \Delta; \theta)^T$$

Optimal estimating functions for SDEs

$$A^*(x, \Delta; \theta) = B_h(x, \Delta; \theta)^T V_h(x, \Delta; \theta)^{-1}$$

Heyde-optimal and Godambe-optimal estimating function:

$$G_n^*(\theta) = \sum_{i=1}^n A^*(X_{t_{i-1}}, \Delta_i; \theta) h(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta) = \sum_{i=1}^n g^*(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta)$$

Score function:

$$U_n(\theta) = \sum_{i=1}^n \partial_\theta \log p(\Delta_i, X_{t_i}, X_{t_{i-1}}; \theta)$$

Kessler (1996): $g_i^*(\Delta, y, x; \theta)$ is the projection of $\partial_{\theta_i} \log p(\Delta, y, x; \theta)$ on the subspace spanned by the functions $y \mapsto h(\Delta, y, x; \theta)$ in $L_2(p(\Delta, y, x; \theta)dy)$, (Δ, x, θ) fixed

Optimal estimating functions for SDEs

THEOREM:

Assume that $g^*(x,y,\theta) = A^*(x;\theta)h(y,x;\theta)$ satisfies Condition LFA1.

Then a sequence $(\hat{\theta}_n)$ of G_n^*–estimators has the asymptotic distribution

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N_p\left(0, V^{-1}\right),$$

where

$$V = \mu_{\theta_0}\left(B_h(\theta_0)V_h(\theta_0)^{-1}B_h(\theta_0)^T\right)$$

μ_{θ_0} is the invariant distribution

Optimal linear estimating functions

$$K_{1,n}(\theta) = \sum_{i=1}^n A(\Delta_i, X_{t_{i-1}};\theta)[X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta)], \qquad A \text{ is a } p\times d\text{–matrix}$$

$$F(\Delta,x;\theta) = E_\theta(X_\Delta|X_0=x) \qquad \Phi(\Delta,x;\theta) = \text{Var}_\theta(X_\Delta|X_0=x)$$

$$B(x,\Delta;\theta) = \partial_\theta F(\Delta,x;\theta) \qquad V(x,\Delta;\theta) = \Phi(\Delta,x;\theta)$$

$$K^*_{1,n}(\theta) = \sum_{i=1}^n \partial_\theta F(\Delta_i, X_{t_{i-1}};\theta)\Phi(\Delta_i, X_{t_{i-1}};\theta)^{-1}[X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta)]$$

Approximately optimal linear estimating function

$$F(\Delta,x;\theta) = x + b(x;\theta)\Delta + O(\Delta^2) \qquad \Phi(\Delta,x;\theta) = \sigma(x;\theta)\sigma(x;\theta)^T\Delta + O(\Delta^2)$$

$$\tilde{K}_{1,n}(\theta) = \sum_{i=1}^n \partial_\theta b(X_{t_{i-1}};\theta)\left(\sigma(X_{t_{i-1}};\theta)\sigma(X_{t_{i-1}};\theta)^T\right)^{-1}[X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta)]$$

Optimal quadratic estimating functions

$d = 1$

$$G_n(\theta) = \sum_{i=1}^n \Big\{a_1(X_{t_{i-1}},\Delta_i;\theta)(X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta))$$

$$+ \, a_2(X_{t_{i-1}},\Delta_i;\theta)\left[(X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta))^2 - \Phi(\Delta_i, X_{t_{i-1}};\theta)\right]\Big\}$$

The optimal quadratic estimating function, $K^*_{2,n}$, is given by:

$$a_1^*(x;\theta) = \frac{\partial_\theta\Phi(x;\theta)\eta(x;\theta) - \partial_\theta F(x;\theta)\Psi(x;\theta)}{\Phi(x;\theta)\Psi(x;\theta) - \eta(x;\theta)^2}$$

$$a_2^*(x;\theta) = \frac{\partial_\theta F(x;\theta)\eta(x;\theta) - \partial_\theta\Phi(x;\theta)\Phi(x;\theta)}{\Phi(x;\theta)\Psi(x;\theta) - \eta(x;\theta)^2}$$

The Δ's have been omitted

$$\eta(x;\theta) = E_\theta([X_\Delta - F(x;\theta)]^3|X_0=x)$$

$$\Psi(x;\theta) = E_\theta([X_\Delta - F(x;\theta)]^4|X_0=x) - \Phi(x;\theta)^2$$

Optimal quadratic estimating functions

Gaussian approximation: $\eta(x;\theta) \doteq 0 \quad \Psi(x;\theta) \doteq 2\Phi(x;\theta)^2$

Result: Approximate Gaussian score function

A further approximation:

$$\partial_\theta F(\Delta,x;\theta) \doteq \partial_\theta b(x;\theta)\Delta \qquad \Phi(\Delta,x;\theta) \doteq \sigma^2(x;\theta)\Delta$$

Approximately optimal quadratic estimating function

$$\tilde{K}_{2,n}(\theta) = \sum_{i=1}^n \left\{ \frac{\partial_\theta b(X_{t_{i-1}};\theta)}{v(X_{t_{i-1}};\theta)}[X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta)] \right.$$

$$\left. + \frac{\partial_\theta v(X_{t_{i-1}};\theta)}{2v^2(X_{t_{i-1}};\theta)\Delta_i}[(X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta))^2 - \Phi(\Delta_i, X_{t_{i-1}};\theta)] \right\}$$

$v(x;\theta) = \sigma^2(x;\theta)$

The square root process

$$dX_t = (\alpha + \beta X_t)dt + \sigma\sqrt{X_t}\,dW_t, \quad X_0 = x_0 > 0.$$

$K^*_{2,n}$ explicitly, but $K^*_{2,n}(\alpha,\beta,\sigma) = 0$ must be solved numerically

$\tilde{K}_{2,n}$ gives the explicit estimators (provided first expression positive):

$$e^{\tilde{\beta}_n\Delta} = \frac{n\sum_{i=1}^n X_{t_i}/X_{t_{i-1}} - (\sum_{i=1}^n X_{t_i})(\sum_{i=1}^n X_{t_{i-1}}^{-1})}{n^2 - (\sum_{i=1}^n X_{t_{i-1}})(\sum_{i=1}^n X_{t_{i-1}}^{-1})}$$

$$\tilde{\alpha}_n = \frac{\tilde{\beta}_n(ne^{\tilde{\beta}_n\Delta} - \sum_{i=1}^n X_{t_i}/X_{t_{i-1}})}{(1-e^{\tilde{\beta}_n\Delta})\sum_{i=1}^n X_{t_{i-1}}^{-1}}$$

$$\tilde{\sigma}_n^2 = \frac{\sum_{i=1}^n X_{t_{i-1}}^{-1}[X_{t_i} - [(\tilde{\alpha}_n + \tilde{\beta}_n)x)e^{\tilde{\beta}_n)\Delta} - \tilde{\alpha}_n]/\tilde{\beta}_n)]^2}{\sum_{i=1}^n X_{t_{i-1}}^{-1}\phi^{\#}(\Delta, X_{t_{i-1}}; \tilde{\alpha}_n, \tilde{\beta}_n)},$$

$$\phi^{\#}(\Delta, x; \alpha, \beta) = \tfrac{1}{2}[(\alpha + 2\beta x)e^{2\beta\Delta} - 2(\alpha+\beta x)e^{\beta\Delta} + \alpha]\beta^{-2}$$

Bibby and Sørensen (1995,1996)

The square root process

Simulation study in Bibby and Sørensen (1995): Empirical mean of 500 estimates of β in the square root process ($\alpha = 10$, $\beta = -1$, and $\sigma = 1$)

Δ	# obs.	mean K^*	mean $\tilde{K}$	S.E. K^*	S.E. $\tilde{K}$
0.5	200	-1.048	-1.049	0.207	0.207
	500	-1.030	-1.030	0.120	0.120
	1000	-1.010	-1.011	0.083	0.083
1.0	200	-1.059	-1.059	0.213	0.213
	500	-1.031	-1.031	0.119	0.120
	1000	-1.005	-1.005	0.082	0.082
2.0	200	-1.030	-1.035	0.240	0.250
	500	-1.045	-1.046	0.200	0.204
	1000	-1.022	-1.022	0.134	0.139

Approximately optimal estimating functions

$$E_\theta(f(X_{t+\Delta})\,|\,X_t) = \sum_{i=0}^k \frac{\Delta^i}{i!}L_\theta^i f(X_t) + O(\Delta^{k+1})$$

under weak conditions on f and the diffusion model,
see e.g. Kessler (1997)

$$L_\theta = \tfrac{1}{2}\sigma^2(x;\theta)\frac{d^2}{dx^2} + b(x;\theta)\frac{d}{dx},$$

$$E_\theta(X_\Delta|X_0 = x) = x + \Delta b(x;\theta) + \tfrac{1}{2}\Delta^2\left\{b(x;\theta)\partial_x b(x;\theta) + \tfrac{1}{2}v(x;\theta)\partial_x^2 b(x;\theta)\right\} + O(\Delta^3)$$

Var$_\theta(X_\Delta|X_0 = x) =$

$$\Delta v(x;\theta) + \Delta^2\left[\tfrac{1}{2}b(x;\theta)\partial_x v(x;\theta) + v(x;\theta)\left\{\partial_x b(x;\theta) + \tfrac{1}{4}\partial_x^2 v(x;\theta)\right\}\right] + O(\Delta^3)$$

$v(x;\theta) = \sigma^2(x;\theta)$

Example

$$dX_t = -\theta\tan(X_t)dt + dW_t\ , \theta > 0$$

Eigenfunctions:

$$\varphi_i(x;\theta) = C_i^\theta(\sin(x)),$$

where C_i^θ is a Gegenbauer polynomial of order i

Eigenvalues:

$$i(\theta + i/2),\ i = 0, 1, \ldots$$

$$\varphi_i(x;\theta) = \sum_{m=0}^i \binom{\theta - 1 + m}{m}\binom{\theta - 1 + i - m}{i - m}\cos[(2m - i)(\pi/2 - x)]$$

$A^*(x,\Delta;\theta)$ can be found explicitly

Example

$$dX_t = -\theta \tan(X_t)dt + dW_t \ , \theta > 0$$

Optimal estimating function based on $\sin(x)$:

$$G^*(\theta) = \sum_{i=1}^{n} \frac{\sin(X_{t_{i-1}})[\sin(X_{t_i}) - e^{-(\theta+\frac{1}{2})\Delta}\sin(X_{t_{i-1}})]}{\frac{1}{2}(e^{2(\theta+1)\Delta} - 1)/(\theta+1) - (e^{\Delta} - 1)\sin^2(X_{t_{i-1}})}$$

When Δ is small we can use the approximately optimal estimating function

$$\tilde{G}(\theta) = \sum_{i=1}^{n} \frac{\sin(X_{t_{i-1}})[\sin(X_{t_i}) - e^{-(\theta+\frac{1}{2})\Delta}\sin(X_{t_{i-1}})]}{\cos^2(X_{t_{i-1}})},$$

$$\tilde{\theta}_n = -\Delta^{-1}\log\left(\frac{\sum_{i=1}^n \tan(X_{t_{i-1}})\sin(X_{t_i})/\cos(X_{t_{i-1}})}{\sum_{i=1}^n \tan^2(X_{t_{i-1}})}\right) - \frac{1}{2},$$

provided $\sum_{i=1}^n \tan(X_{t_{i-1}})\sin(X_{t_i})/\cos(X_{t_{i-1}}) > 0$

Explicit optimal estimating functions

Consider

$$h_i(\Delta, x, y; \theta) = \varphi_i(y;\theta) - e^{-\lambda_i(\theta)\Delta}\varphi_i(x;\theta), \qquad i = 1, \ldots, N,$$

where $\varphi_i(y;\theta)$ is an eigenfunction of the generator with eigenvalue $\lambda_i(\theta)$, for which $\pi_\theta^\Delta \varphi_i(x;\theta) = e^{-\lambda_i(\theta)\Delta}\varphi_i(x;\theta)$

Suppose

$$\varphi_i(x;\theta) = \Pi_i(\kappa(x);\theta),$$

where κ is a real function independent of θ, and Π_i is a polynomial of degree i:

$$\Pi_i(y;\theta) = \sum_{j=0}^{i} a_{i,j}(\theta)\, y^j$$

Explicit optimal estimating functions

Optimal weight matrix:

$$B_h(x,\Delta;\theta)_{ij} = \sum_{k=0}^{j} \partial_{\theta_i} a_{j,k}(\theta) \int_\ell^r \kappa(y)^k p(\Delta, x, y;\theta)dy - \partial_{\theta_i}[e^{-\lambda_j(\theta)\Delta}\varphi_j(x;\theta)]$$

$i = 1, \ldots, p, \ \ j = 1, \ldots, N$

$$V_h(\Delta, x;\theta)_{ij} = \sum_{r=0}^{i}\sum_{s=0}^{j} a_{i,r}(\theta)a_{j,s}(\theta)\int_\ell^r \kappa(y)^{r+s}p(\Delta,x,y;\theta)dy \\ - e^{-[\lambda_i(\theta)+\lambda_j(\theta)]\Delta}\varphi_i(x;\theta)\varphi_j(x;\theta)$$

$i, j = 1, \ldots, N$

Explicit optimal estimating functions

Thus to find the optimal estimating function based on the first N eigenfunctions, we need to find the moments

$$\int_\ell^r \kappa(y)^i p(\Delta, x, y;\theta)dy \quad \text{for} \quad 1 \le i \le 2N$$

If we integrate both sides of

$$\varphi_i(y;\theta) = \sum_{j=0}^{i} a_{i,j}(\theta)\,\kappa(y)^j$$

with respect to $p(\Delta, x, y;\theta)$ for $i = 1, \ldots, 2N$, we obtain a system of linear equations

$$e^{-\lambda_i(\theta)\Delta}\varphi_i(x;\theta) = \sum_{j=0}^{i} a_{i,j}(\theta)\int_\ell^r \kappa(y)^j p(\Delta,x,y;\theta)dy, \qquad i = 1, \ldots, 2N$$

Optimal estimating functions

$$G_n^\bullet(\theta) = \sum_{i=1}^{n} A^*(X_{(i-1)\Delta}; \tilde{\theta}_n) h(X_{(i-1)\Delta}, X_{i\Delta}; \theta)$$

where $\tilde{\theta}_n$ is a weakly $\sqrt{n}$-consistent estimator of θ_0

$\tilde{\theta}_n$ e.g. obtained from

$$G_n(\theta) = \sum_{i=1}^{n} A^\circ(X_{(i-1)\Delta}; \theta) h(X_{(i-1)\Delta}, X_{i\Delta}; \theta)$$

for some simple choice of $A^\circ(x;\theta)$

Under regularity conditions:

$$\sqrt{n}(\hat{\theta}_n^\bullet - \theta_0) \xrightarrow{\mathcal{D}} N_p\left(0, V^{-1}\right),$$

where $V = \mu_{\theta_0}\left(B_h(\theta_0) V_h(\theta_0)^{-1} B_h(\theta_0)^T\right)$

Jump diffusions

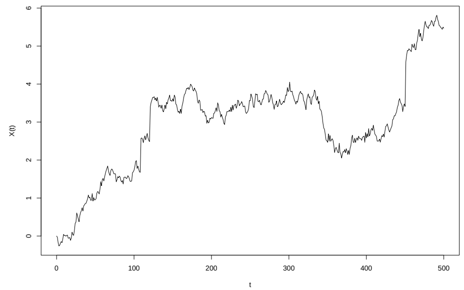
$$dX_t = \alpha dt + \sigma dW_t + dZ_t$$

$$Z_t = \sum_{j=0}^{N_t} Y_j$$

N is a Poisson process with intensity λ

Y_j, $j = 1, 2, \ldots$, are i.i.d. normal with mean μ and variance τ^2

W, N and the Y_js are independent



$\alpha = 0.0001, \sigma = 0.1, \lambda = 0.01, \mu = 1,$ and $\tau = 0.1.$

Jump diffusions

$$\Delta = 1$$

$$G_n(\theta) = \sum_{i=1}^{n} A(X_{i-1}, \theta) h(X_i, X_{i-1}; \theta)$$

$$h(x, y; \theta) = \begin{pmatrix} y - F(x;\theta) \\ (y - F(x;\theta))^2 - \Phi(x;\theta) \\ e^y - \kappa(x;\theta) \end{pmatrix}$$

$$\begin{aligned} F(x;\theta) &= \mathsf{E}_\theta(X_i | X_{i-1} = x) = x + \alpha + \lambda\mu \\ \Phi(x;\theta) &= \mathsf{Var}_\theta(X_i | X_{i-1} = x) = \sigma^2 + \lambda(\mu^2 + \tau^2) \\ \kappa(x;\theta) &= \mathsf{E}_\theta(e^{X_i} | X_{i-1} = x) = \exp\left(x + \alpha + \tfrac{1}{2}\sigma^2 + \lambda(e^{\mu + \frac{1}{2}\tau^2} - 1)\right) \end{aligned}$$

Explicit expression for the optimal weight matrix $A(x, \theta)$

Jump diffusions

Parameter	True value	Mean	Standard error
α	0.0001	-0.0009	0.0070
σ	0.1	0.0945	0.0180
λ	0.01	0.0155	0.0209
μ	1	0.9604	0.5126
τ	0.1	0.2217	0.3156

500 observations (500 simulated estimates)

A dangerous approximation

$$g(\Delta, y, x; \theta) = \sum_{j=1}^{N} a_j(x, \Delta; \theta)[f_j(y) - \pi_\theta^\Delta f_j(x)]$$

Simplify by $\quad \pi_\theta^\Delta f(x) \simeq f(x) + \left[b(x;\theta)f'(x) + \frac{1}{2}v(x;\theta)f''(x)\right]\Delta$?

Consider

$$dX_t = -\beta\,(X_t - \alpha)\,dt + \psi(X_t;\rho)dW_t,$$

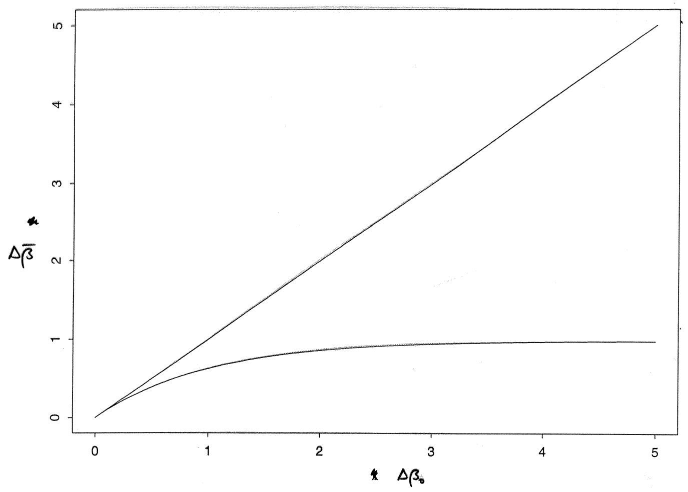
and estimate β and α by solving

$$\sum_{i=1}^{n} a(X_{t_{i-1}};\theta)[X_{t_i} - (X_{t_{i-1}} + b(X_{t_{i-1}};\theta))\Delta] = 0.$$

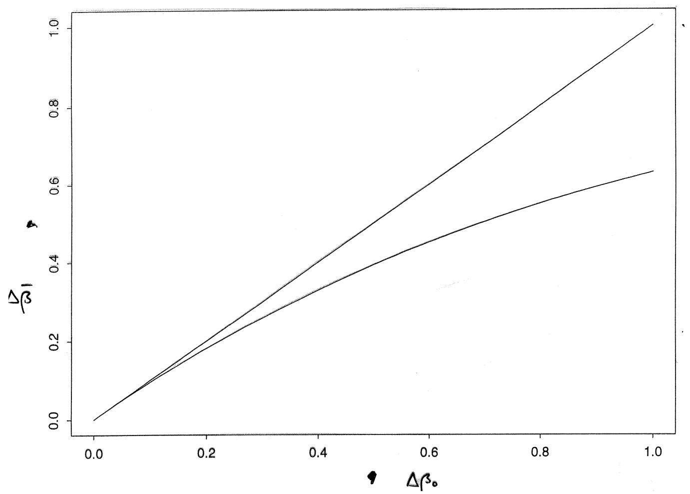
Then $\hat{\alpha}_n \to \alpha_0$ and $\hat{\beta}_n \to \bar{\beta}$, where

$$\bar{\beta} = \frac{1 - e^{-\beta_0\Delta}}{\Delta} \leq \frac{1}{\Delta}$$

A dangerous approximation



A dangerous approximation



A dangerous approximation

Δ	# obs.	mean $\hat{\beta}_n$	SE $\hat{\beta}_n$		Δ	# obs.	mean $\hat{\beta}_n$	SE $\hat{\beta}_n$
0.5	200	0.81	0.12		1.5	200	0.52	0.05
0.5	500	0.80	0.07		1.5	500	0.52	0.03
0.5	1000	0.79	0.05		1.5	1000	0.52	0.02
1.0	200	0.65	0.07		2.0	200	0.43	0.03
1.0	500	0.64	0.04		2.0	500	0.43	0.02
1.0	1000	0.63	0.03		2.0	1000	0.43	0.02

Square root process with $\beta = 1, \mu = 10$ and $\sigma = 1$ 500 simulations

Simulation study in Bibby & Sørensen (1995)

CKLS model - diffusion parameters

$$dX_t = -\beta\,(X_t - \alpha)\,dt + \sigma X_t^{\gamma} dW_t$$

Assume further that $\gamma_0 = 0.5$ (i.e. data are from the square root process)

This implies that

$$X_t \sim \Gamma(2\alpha_0\beta_0\sigma_0^{-2}, 2\beta_0\sigma_0^{-2})$$

provided that

$$\lambda_0 = 2\alpha_0\beta_0\sigma_0^{-2} > 1$$

$$\begin{aligned}\bar{\sigma}^2(\gamma) &= \left(\frac{\sigma_0^2}{2\beta_0}\right)^{1-2\gamma} \frac{\Gamma(\lambda_0 - 2\gamma)}{\Delta\Gamma(\lambda_0)} \left[\sigma_0^2(e^{-\beta_0\Delta} - e^{-2\beta_0\Delta})(\lambda_0 - 2\gamma)/\beta_0\right. \\ &\qquad \left. + \alpha_0(e^{-2\beta_0\Delta} - 2e^{-\beta_0\Delta} + 1)\right]\end{aligned}$$

CKLS model - diffusion parameters

$$\bar{\sigma}^2(\gamma) = \sigma_0^2\left(\frac{\sigma_0^2}{2\beta_0}\right)^{1-2\gamma} \frac{\Gamma(\lambda_0 + (1-2\gamma))}{\Gamma(\lambda_0)}\left[1 + \frac{3\gamma\sigma_0^2/2 - \alpha_0\beta_0}{\alpha_0\beta_0 - \gamma\sigma_0^2}\beta_0\Delta + O\left((\beta_0\Delta)^2\right)\right]$$

If we restrict the model by assuming that $\gamma = 0.5$, we find that

$$\begin{aligned}\bar{\sigma}^2 &= \frac{\sigma_0^2}{\Delta\beta_0}\left[e^{-\beta_0\Delta} - e^{-2\beta_0\Delta} + \frac{e^{-2\beta_0\Delta} - 2e^{-\beta_0\Delta} + 1}{2 - \sigma_0^2/(\alpha_0\beta_0)}\right] \\ &= \sigma_0^2\left(1 + \frac{3/2 - \lambda_0}{\lambda_0 - 1}\beta_0\Delta + O\left(\beta_0\Delta)^2\right)\right)\end{aligned}$$

CKLS model - diffusion parameters

$\bar{\gamma}$, two solutions:

$$\bar{\gamma} = \frac{1}{4}\left[\lambda_0(1 + q_0) \pm \sqrt{\lambda_0^2(1 + q_0)^2 - 4(\lambda_0 - 1)}\right]$$

where $\lambda_0 = 2\alpha_0\beta_0\sigma_0^{-2} > 1$ and

$$q_0 = \frac{e^{-2\beta_0\Delta} - 2e^{-\beta_0\Delta} + 1}{2(e^{-\beta_0\Delta} - e^{-2\beta_0\Delta})} > 0.$$

As $\Delta \to 0$

$$q_0 \to 0$$

and

$$\bar{\gamma} \to \begin{cases} 0.5 \\ 0.5(\lambda_0 - 1) \end{cases}$$

Jacobi diffusion

$$dX_t = -\beta[X_t - (m + \gamma z)]dt + \sigma\sqrt{z^2 - (X_t - m)^2}dW_t$$

Asymptotic information at $(\beta, \gamma, \sigma^2) = (0.02, 0, 0.01)$:

Eigenfunction no.	1	2	1 & 2
Inf. for $\hat{\beta}$	47.4	44.8	49.2
Inf. for $\hat{\sigma}^2$	0	759	5016

For optimal estimating functions based on more than two eigenfunctions, the information is not increased by more than 1 - 3 per cent

Larsen & Sørensen (2007)

High frequency asymptotics

Scalar diffusion process

$$dX_t = b(X_t;\alpha)dt + \sigma(X_t;\beta)dW_t$$

$\theta = (\alpha,\beta) \in \Theta \subseteq \mathbb{R}^2$ θ_0 is the true parameter value

State space: (ℓ, r)

Ergodic with invariant measure μ_θ

$v(x;\beta) = \sigma^2(x;\beta)$

Data: $X_{t_0^n}, \ldots, X_{t_n^n}$ $t_i^n = i\Delta_n$, $i = 0,\ldots,n$

High frequency asymptotic scenario:

$$n \to \infty \qquad \Delta_n \to 0 \qquad n\Delta_n \to \infty$$

Condition HFA1: the process

- $\int_{x^\#}^r s(x;\theta)dx = \int_\ell^{x^\#} s(x;\theta)dx = \infty$ and $\int_\ell^r x^k \tilde{\mu}_\theta(x)dx < \infty$

for all $k \in \mathbb{N}$, where $x^\#$ is an arbitrary point in (ℓ, r),

$s(x;\theta) = \exp\left(-2\int_{x^\#}^x \frac{b(y;\alpha)}{v(y;\beta)}dy\right)$ and $\tilde{\mu}_\theta(x) = [s(x;\theta)v(x;\beta)]^{-1}$

- $\sup_t E_\theta(|X_t|^k) < \infty$ for all $k \in \mathbb{N}$
- $b, \sigma \in C_{p,4,1}((\ell,r)\times\Theta)$

Technical definitions

$C_{p,k_1,k_2,k_3}(\mathbb{R}_+ \times (\ell,r)^2 \times \Theta)$ is the class of real functions $f(t,y,x;\theta)$ satisfying that

- $f(t,y,x;\theta)$ is k_1 times continuously differentiable with respect t, k_2 times continuously differentiable with respect y, and k_3 times continuously differentiable with respect α and with respect to β
- f and all partial derivatives $\partial_t^{i_1}\,\partial_y^{i_2}\,\partial_\alpha^{i_3}\,\partial_\beta^{i_4} f$, $i_j = 1,\ldots k_j$, $j = 1,2$, $i_3 + i_4 \le k_3$, are of polynomial growth in x and y uniformly for θ in a compact set (for fixed t)

$C_{p,k_1,k_2}((\ell,r)\times\Theta)$ for $f(y;\theta)$ and $C_{p,k_1,k_2}((\ell,r)^2\times\Theta)$ for $f(y,x;\theta)$

are defined similarly

$|R(\Delta,y,x;\theta)| \le F(y,x;\theta)$

F is of polynomial growth in y and x uniformly for θ in a compact set

Condition HFA2: the estimating function

$$G_n(\theta) = \sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \qquad 2-\text{dimensional}$$

- For some $\kappa \ge 2$

$E_\theta(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}) = \Delta_n^\kappa R(\Delta_n, X_{t_{i-1}^n};\theta)$ for all $\theta \in \Theta$

- The function $g(\Delta,y,x;\theta)$ has an expansion in powers of Δ

$$g(\Delta,y,x;\theta) = g(0,y,x;\theta) + \Delta g^{(1)}(y,x;\theta) + \tfrac{1}{2}\Delta^2 g^{(2)}(y,x;\theta) + O(\Delta^3)$$

- The function $R(\Delta,y,x;\theta)$ in the expansion of g is differentiable with respect to θ, and

$$\begin{aligned} g(\Delta,y,x;\theta) &\in C_{p,6,2}((\ell,r)^2\times\Theta) \text{ for fixed } \Delta,\\ g^{(1)}(y,x;\theta) &\in C_{p,4,2}((\ell,r)^2\times\Theta),\\ g^{(2)}(y,x;\theta) &\in C_{p,2,2}((\ell,r)^2\times\Theta) \end{aligned}$$

HF-asymptotics: Theorem 1

Suppose

- Conditions HFA1 and HFA2
- The identifiability condition that

$$\gamma(\theta,\theta_0) = \int_\ell^r [b(x,\alpha_0) - b(x,\alpha)]\partial_y g(0,x,x;\theta)\mu_{\theta_0}(x)dx + \tfrac{1}{2}\int_\ell^r [v(x,\beta_0) - v(x,\beta)]\partial_y^2 g(0,x,x;\theta)\mu_{\theta_0}(x)dx \neq 0$$

for all $\theta \neq \theta_0$

- The matrix $S = \int_\ell^r A_{\theta_0}(x)\mu_{\theta_0}(x)dx$ is invertible, where

$$A_\theta(x) = \begin{pmatrix} \partial_\alpha b(x;\alpha)\partial_y g_1(0,x,x;\theta) & \frac{1}{2}\partial_\beta v(x;\beta)\partial_y^2 g_1(0,x,x;\theta) \\ \partial_\alpha b(x;\alpha)\partial_y g_2(0,x,x;\theta) & \frac{1}{2}\partial_\beta v(x;\beta)\partial_y^2 g_2(0,x,x;\theta) \end{pmatrix}$$

HF-asymptotics: Theorem 1

Then a consistent estimator $\hat\theta_n = (\hat\alpha_n, \hat\beta_n)$ that solves the estimating equation $G_n(\theta) = 0$ exists and is unique in any compact subset of Θ containing θ_0 with a probability that goes to one as $n \to \infty$.

For a martingale estimating function or more generally if $n\Delta^{2\kappa-1} \to 0$,

$$\sqrt{n\Delta_n}(\hat\theta_n - \theta_0) \xrightarrow{\mathcal{D}} N_2\left(0, S^{-1}V_0(S^T)^{-1}\right)$$

under P_{θ_0}, where

$$V_0 = \int_\ell^r v(x,\beta_0)\partial_y g(0,x,x;\theta_0)\partial_y g(0,x,x;\theta_0)^T \mu_{\theta_0}(x)dx.$$

HF-asymptotics: Optimal rate

Gobet (2002):

A discretely sampled diffusion is LAN in the high frequency asymptotics considered here, and the optimal rate of convergence is

For parameters in the drift coefficient: $\sqrt{n\Delta_n}$

For parameters in the diffusion coefficient: $\sqrt{n}$

Jacobsen's condition:

$$\partial_y g_2(0,x,x;\theta) = 0$$

for all $x \in (\ell, r)$ and $\theta \in \Theta$

Jacobsen (2001): small Δ-optimality

HF-asymptotics: Theorem 2

Suppose

- Conditions HFA1 and HFA2
- The identifiability condition that

$$\int_\ell^r [b(x,\alpha_0) - b(x,\alpha)]\partial_y g_1(0,x,x;\theta)\mu_{\theta_0}(x)dx \neq 0 \quad \text{when } \alpha \neq \alpha_0$$

$$\int_\ell^r [v(x,\beta_0) - v(x,\beta)]\partial_y^2 g_2(0,x,x;\theta)\mu_{\theta_0}(x)dx \neq 0 \quad \text{when } \beta \neq \beta_0$$

- $S_{11} \neq 0$ and $S_{22} \neq 0$
- $\partial_y g_2(0,x,x;\theta) = 0$

HF-asymptotics: Theorem 2

Then a consistent estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ that solves the estimating equation $G_n(\theta) = 0$ exists and is unique in any compact subset of Θ containing θ_0 with a probability that goes to one as $n \to \infty$

If, moreover,

$$\partial_\alpha \partial_y^2 g_2(0, x, x; \theta) = 0,$$

then for a martingale estimating function or more generally if $n\Delta^{2(\kappa-1)} \to 0$,

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1/S_{11}^2 & 0 \\ 0 & W_2/S_{22}^2 \end{pmatrix} \right)$$

where

$$W_1 = \int_\ell^r v(x;\beta_0)[\partial_y g_1(0,x,x;\theta_0)]^2 \mu_{\theta_0}(x)dx$$

and

$$W_2 = \frac{1}{2}\int_\ell^r v(x;\beta_0)^2[\partial_y^2 g_2(0,x,x;\theta_0)]^2 \mu_{\theta_0}(x)dx$$

HF-asymptotics: Efficiency

Gobet (2002):

A discretely sampled diffusion is LAN in the high frequency asymptotics considered here, and the Fisher information is

$$\begin{pmatrix} \int_\ell^r \frac{(\partial_\alpha b(x;\alpha_0))^2}{v(x;\beta_0)} \mu_{\theta_0}(x)dx & 0 \\ 0 & \frac{1}{2}\int_\ell^r \left[\frac{\partial_\beta v(x;\beta_0)}{v(x;\beta_0)}\right]^2 \mu_{\theta_0}(x)dx \end{pmatrix}$$

Condition for efficiency:

$\partial_y g_1(0,x,x;\theta) = \partial_\alpha b(x;\alpha)/v(x;\beta)$ $\qquad$ $\partial_y^2 g_2(0,x,x;\theta) = \partial_\beta v(x;\beta)/v(x;\beta)^2$

for all $x \in (\ell, r)$ and $\theta \in \Theta$

Jacobsen (2001): small Δ-optimality

HF-asymptotics: Quadratic estimating functions

$$\sum_{i=1}^n \begin{pmatrix} a_1(X_{t_{i-1}^n}, \Delta; \theta)(X_{t_i^n} - F(\Delta, X_{t_{i-1}^n}; \theta)) \\ a_2(X_{t_{i-1}^n}, \Delta; \theta)\left[(X_{t_i^n} - F(\Delta, X_{t_{i-1}^n}; \theta))^2 - \phi(\Delta, X_{t_{i-1}^n}; \theta)\right] \end{pmatrix}$$

$$\begin{aligned} F(\Delta, x; \theta) &= E_\theta(X_\Delta | X_0 = x) = x + O(\Delta) \\ \phi(\Delta, x; \theta) &= \text{Var}_\theta(X_\Delta | X_0 = x) = O(\Delta) \end{aligned}$$

$$g(0,y,x;\theta) = \begin{pmatrix} a_1(x,0;\theta)(y-x) \\ a_2(x,0;\theta)(y-x)^2 \end{pmatrix}$$

$$\begin{aligned} \partial_y g_2(0,y,x;\theta) &= 2a_2(x,\Delta;\theta)(y-x) && \text{Jacobsen's condition satisfied} \\ \partial_y g_1(0,x,x;\theta) &= a_1(x,0;\theta) = \partial_\alpha b(x;\alpha)/v(x;\beta) && \text{Approximately} \\ \partial_y^2 g_2(0,x,x;\theta) &= 2a_2(x,0;\theta) = \partial_\beta v(x;\beta)/v(x;\beta)^2 && \text{optimal} \end{aligned}$$

HF-asymptotics: Estimating functions

$$G_n(\theta) = \sum_{i=1}^n A(X_{t_{i-1}^n}, \Delta_n; \theta)[f(X_{t_i^n}) - \pi_\theta^{\Delta_n} f(X_{t_{i-1}^n})]$$

$f(y) = (f_1(y), \ldots, f_N(y))^T$

$A(x, \Delta; \theta)$ a $2 \times N$-matrix of weights

$\pi_\theta^\Delta f(x) = E_\theta(f(X_\Delta) \,|\, X_0 = x)$ is the transition operator

HF-asymptotics: Theorem 3

Suppose Condition HFA1 is satisfied and that the functions f_j are six times continuously differentiable

Jacobsen (2002):

A sufficient condition that it is possible to find a specification of the weight matrix $A(x, \Delta; \theta)$ such that the estimating function $G_n(\theta)$ gives estimators that are rate optimal and efficient is that

- $N \geq 2$
- and that the matrix

$$D(x) = \begin{pmatrix} f_1'(x) & f_1''(x) \\ f_2'(x) & f_2''(x) \end{pmatrix}$$

is invertible for μ_θ-almost all x

For a d-dimensional diffusion: $N \geq d(d+3)/2$

HF-asymptotics: Theorem 3

For $N = 2$, the estimators are rate optimal and efficient when

$$A(x, 0; \theta) = \begin{pmatrix} \partial_\alpha b(x;\alpha)/v(x;\beta) & c(x;\theta) \\ 0 & \partial_\beta v(x;\beta)/v(x;\beta)^2 \end{pmatrix} D(x)^{-1}$$

for any function $c(x;\theta)$

Godambe-Heyde optimality

$$G_n(\theta) = \sum_{i=1}^n A^*(X_{t_{i-1}^n}, \Delta_n; \theta)[f(X_{t_i^n}) - \pi_\theta^{\Delta_n} f(X_{t_{i-1}^n})]$$

is Godambe-Heyde optimal if

$$A^*(x, \Delta; \theta) E_\theta \left([f(X_\Delta) - \pi_\theta^\Delta f(x)][f(X_\Delta) - \pi_\theta^\Delta f(x)]^T \mid X_0 = x \right) = \partial_\theta \pi_\theta^\Delta f^T(x)$$

for μ_θ-almost all x

HF-asymptotics: Theorem 4

Suppose Condition HFA1 is satisfied, that the functions f_j are six times continuously differentiable, that $N \geq 2$ and that $D(x)$ is invertible for μ_θ-almost all x

Then

$$g^*(\Delta, y, x; \theta) = \begin{pmatrix} 1 & 0 \\ 0 & 2\Delta \end{pmatrix} A^*(x, \Delta; \theta)[f(y) - \pi_\theta^\Delta f(x)]$$

satisfies that

$$\partial_y g_2^*(0, x, x; \theta) = 0$$

and

$$\partial_y g_1^*(0, x, x; \theta) = \partial_\alpha b(x;\alpha)/v(x;\beta) \qquad \partial_y^2 g_2^*(0, x, x; \theta) = \partial_\beta v(x;\beta)/v(x;\beta)^2$$

for all $x \in (\ell, r)$ and $\theta \in \Theta$

HF-asymptotics: Kessler

$$H_n^{(k)}(\theta) = \sum_{i=1}^{n} \frac{\partial_\theta r_k(\Delta_i, X_{t_{i-1}};\theta)}{\Gamma_{k+1}(\Delta_i, X_{t_{i-1}};\theta)}[X_{t_i} - r_k(\Delta_i, X_{t_{i-1}};\theta)]$$
$$+ \sum_{i=1}^{n} \frac{\partial_\theta \Gamma_{k+1}(\Delta_i, X_{t_{i-1}};\theta)}{2\Gamma_{k+1}(\Delta_i, X_{t_{i-1}};\theta)^2}[(X_{t_i} - r_k(\Delta_i, X_{t_{i-1}};\theta))^2 - \Gamma_{k+1}(\Delta_i, X_{t_{i-1}};\theta)]$$

$$r_k(\Delta, x;\theta) = \sum_{i=0}^{k} \frac{\Delta^i}{i!} L_\theta^i f(x), \qquad \text{where } f(x) = x$$

For fixed x, y and θ the function $(y - r_k(\Delta, x;\theta))^2$ is a polynomial of order $2k$ in Δ. Define $g_{x,\theta}^j(y)$, $j = 0, 1, \cdots, k$ by

$$(y - r_k(\Delta, x;\theta))^2 = \sum_{j=0}^{k} \Delta^j g_{x,\theta}^j(y) + O(\Delta^{k+1})$$

$$\Gamma_k(\Delta, x;\theta) = \sum_{j=0}^{k} \Delta^j \sum_{r=0}^{k-j} \frac{\Delta^r}{r!} L_\theta^r g_{x,\theta}^j(x)$$

HF-asymptotics: Kessler

THEOREM:

Under regularity conditions

$$\hat\theta_{k,n} \overset{P_{\theta_0}}{\to} \theta_0 \quad \text{as } \Delta_n \to 0 \text{ and } n\Delta_n \to \infty$$

If, moreover, $n\Delta^{2k+1} \to 0$, then

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat\alpha_n - \alpha_0) \\ \sqrt{n}(\hat\beta_n - \beta_0) \end{pmatrix} \overset{\mathcal{D}}{\longrightarrow} N\left(0, I(\theta)^{-1}\right),$$

where $I(\theta)$ is the Fisher information matrix

$$\begin{pmatrix} \int_\ell^r \frac{(\partial_\alpha b(x;\alpha_0))^2}{v(x;\beta_0)} \mu_{\theta_0}(x)dx & 0 \\ 0 & \frac{1}{2}\int_\ell^r \left[\frac{\partial_\beta v(x;\beta_0)}{v(x;\beta_0)}\right]^2 \mu_{\theta_0}(x)dx \end{pmatrix}$$

HF-asymptotics, fixed interval

$$dX_t = b(X_t;\alpha)dt + \sigma(X_t;\beta)dW_t \qquad X \ d\text{-dimensional}$$

DATA: $X_{i/n}$, $i = 1, \ldots, n$

α cannot be consistently estimated when $n \to \infty$

Genon-Catalot and Jacod (1993) proposed an approximate Gaussian log-likelihood function for β

Under regularity conditions, the estimator $\hat\beta_n$ is consistent and

$$\sqrt{n}(\hat\beta_n - \beta_0) \overset{\mathcal{D}}{\longrightarrow} Z,$$

where the distribution of Z is a normal variance mixture

Small diffusion - high frequency asymptotics

Sørensen and Uchida (2003), Gloter and Sørensen (2005)

$$dX_t = b(X_t, \alpha)dt + \varepsilon\sigma(X_t, \beta)dW_t, \quad X_0 = x_0 \qquad X \ d\text{-dimensional}$$

$\varepsilon > 0$ $\quad \theta = (\alpha, \beta) \in \overline{\Theta}_1 \times \overline{\Theta}_2$ $\quad \theta_0$ is the true parameter value

$\Theta_i \subseteq \mathbb{R}$ open, convex, bounded $\quad \theta_0 \in \Theta_1 \times \Theta_2$

Data: $X_0, X_{1/n}, X_{2/n} \ldots, X_1$

Asymptotic scenario:

$$n \to \infty \qquad \varepsilon_n \to 0 \qquad \liminf_{n\to\infty} \varepsilon_n n^\rho > 0 \qquad \text{for some } \rho > 0$$

Small diffusion - high frequency asymptotics

Define $\xi_t(x,\alpha)$ by

$$\partial_t \xi_t(x,\alpha) = b(\xi_t(x,\alpha),\alpha), \quad \xi_0(x,\alpha) = x,$$

and

$$\tilde{\delta}_n(x,\alpha) = \xi_{1/n}(x,\alpha)$$

Approximate Gaussian log-likelihood:

$$\tilde{U}_{\varepsilon,n}(\theta) = -\sum_{k=1}^{n} \left\{ \log\det V_{k-1}(\beta) + \varepsilon^{-2} n P_k(\alpha)^T V_{k-1}(\beta)^{-1} P_k(\alpha) \right\}$$

$$P_k(\alpha) = X_{k/n} - \tilde{\delta}_n(X_{(k-1)/n},\alpha) \qquad V_k(\beta) = \sigma(X_{k/n},\beta)\sigma(X_{k/n},\beta)^T$$

Small diffusion - high frequency asymptotics

Approximations:

$\delta_n(x,\alpha)$ an approximations to $\tilde{\delta}_n(x,\alpha)$ such that for any compact subset K of $\mathbb{R}^d$

$$\sup_{x\in K,\alpha\in\overline{\Theta}_1} |\delta_n(x,\alpha) - \tilde{\delta}_n(x,\alpha)| \le c(K)\varepsilon n^{-3/2}$$

Similar bounds hold for the first two derivatives of δ_n and $\tilde{\delta}_n$ with respect to α plus further regularity conditions

Approximate Gaussian log-likelihood:

$$U_{\varepsilon,n}(\theta) = -\sum_{k=1}^{n} \left\{ \log\det V_{k-1}(\beta) + \varepsilon^{-2} n P_k(\alpha)^T V_{k-1}(\beta)^{-1} P_k(\alpha) \right\},$$

$$P_k(\alpha) = X_{k/n} - \delta_n(X_{(k-1)/n},\alpha) \qquad V_k(\beta) = \sigma(X_{k/n},\beta)\sigma(X_{k/n},\beta)^T$$

Small diffusion - high frequency asymptotics

$$\mathcal{L}_\alpha(f)(x) = \sum_{i=1}^{d} b_i(x,\alpha)\partial_{x_i} f(x)$$

$$\delta_n^k(x,\alpha) = \sum_{j=1}^{k} (\mathcal{L}_\alpha)^{j-1}(b(\cdot,\alpha))(x)\frac{n^{-j}}{j!}$$

δ_n^k satisfies the regularity conditions when $k - 1/2 \ge \rho$

$\delta_n^1(x,\alpha) = n^{-1}b(x,\alpha)$ Sørensen and Uchida (2003)

$$\delta_n^2(x,\alpha) = n^{-1}b(x,\alpha) + \frac{1}{2}n^{-2}\sum_{i=1}^{d} b_i(x,\alpha)\partial_{x_i} b(x,\alpha)$$

Small diffusion - high frequency asymptotics

Suppose $\hat{\theta}_{\varepsilon,n} = (\hat{\alpha}_{\varepsilon,n},\hat{\beta}_{\varepsilon,n})$ maximize the approximate Gaussian log-likelihood $U_{\varepsilon,n}(\theta)$

Under regularity conditions

$$\begin{pmatrix} \varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \end{pmatrix} \to N\left(0, \begin{pmatrix} I_b(\theta_0) & 0 \\ 0 & I_\sigma(\theta_0) \end{pmatrix}^{-1}\right)$$

where

$$I_b^{i,j}(\theta_0) = \int_0^1 \left(\partial_{\alpha_i} b(X_s^0,\alpha_0)\right)^T [\sigma\sigma^T]^{-1}(X_s^0,\beta_0)\left(\partial_{\alpha_j} b(X_s^0,\alpha_0)\right) ds$$

$$I_\sigma^{i,j}(\theta_0) = \frac{1}{2}\int_0^1 \mathrm{tr}\left[\left(\partial_{\beta_i}[\sigma\sigma^T]\right)[\sigma\sigma^T]^{-1}\left(\partial_{\beta_j}[\sigma\sigma^T]\right)[\sigma\sigma^T]^{-1}(X_s^0,\beta_0)\right] ds$$

Asymptotic scenarios - 1

DATA: $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$

Possibly time-series from several individuals

LARGE SAMPLE ASYMPTOTICS:

$n \to \infty$ or

number of individuals $\to \infty$ (Pedersen, 2000)

HIGH FREQUENCY ASYMPTOTICS:

$\Delta \to 0$ and $n \to \infty$

$n\Delta \to \infty$: Prakasa-Rao (1983), Yoshida (1992), Kessler (1997), Sørensen (2005)

$n\Delta$ constant: Dohnal (1987), Jacod and Genon-Catalot (1993)

Asymptotic scenarios - 2

SMALL DIFFUSION ASYMPTOTICS:

$\sigma(X;\theta) = \epsilon g(x,\theta)$ $\quad \epsilon \to 0$ $\quad$ Sørensen (2000)

SMALL DIFFUSION/HIGH FREQUENCY ASYMPTOTICS:

$\sigma(X;\theta) = \epsilon g(x,\theta)$ $\quad \epsilon \to 0, \Delta \to 0$ and $n \to \infty$

Genon-Catalot (1990), Sørensen and Uchida (2003), Gloter and Sørensen (2006)

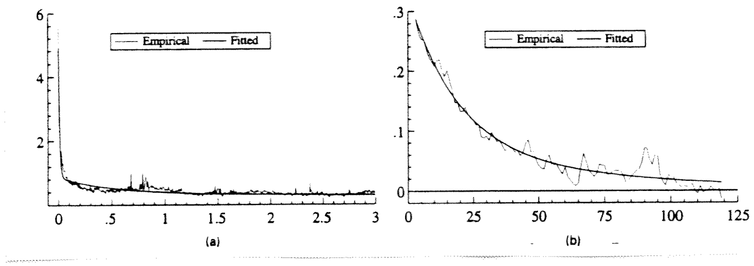
SMALL VOLATILITY OF VOLATILITY ASYMPTOTICS

Sørensen and Yoshida (2000)

German mark – US dollar: 5-minute returns

Empirical and fitted autocovariance functions (x-axis days)

Barndorff-Nielsen & Shephard (2001)



$$\rho(t) = \phi_1 \exp(-\beta_1 t) + \ldots + \phi_4 \exp(-\beta_4 t)$$

Sums of diffusions

$$\begin{aligned} Y_t &= X_{1,t} + \ldots + X_{M,t} \\ dX_{i,t} &= -\beta_i(X_{i,t} - \mu_i) + \sigma_i(X_{i,t})dW_{i,t}, \quad i = 1, \ldots, M, \end{aligned}$$

Autocorrelation function of Y

$$\rho(t) = \phi_1 \exp(-\beta_1 t) + \ldots + \phi_M \exp(-\beta_M t)$$

$$\phi_i = \frac{\text{Var}(X_{i,t})}{\text{Var}(X_{1,t}) + \cdots + \text{Var}(X_{M,t})}$$

Sums of diffusions with linear drift and prescribed marginal distribution were studied in Bibby, Skovgaard and Sørensen (2005)

Sums of Pearson diffusions were considered in Forman and Sørensen (2008)

Integrated diffusions

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t, \quad X_0 \sim \mu_\theta \quad d = 1$$

Data:

$$Y_i = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} X_s\,ds, \qquad i = 1,\ldots,n$$

For instance ice-core data, Ditlevsen, Ditlevsen and Andersen (2002)

Bollersev and Wooldridge (1992)

Gloter (2000, 2006)

Ditlevsen and Sørensen (2004)

Comte, Genon-Catalot and Rozenholc (2008)

Partially observed diffusions

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t$$

X, b D-dimensional, W m-dimensional, σ $D \times m$-matrix

Data:

$$Y_i = k(X_{t_i}) + Z_i \qquad i = 1,\ldots,n$$

$k : \mathbb{R}^D \mapsto \mathbb{R}^d$ $(d < D)$

$\{Z_i\}$ i.i.d. sequence with mean zero independent of X

Hypoelliptic diffusions

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \sigma(x;\theta) = \begin{pmatrix} \gamma(x,\theta) \\ 0 \end{pmatrix}$$

Integrated diffusions:

$$\begin{aligned} dX_{1,t} &= b(X_t;\theta)dt + \gamma(X_t;\theta)dW_t \\ dX_{2,t} &= X_{1,t}dt \end{aligned}$$

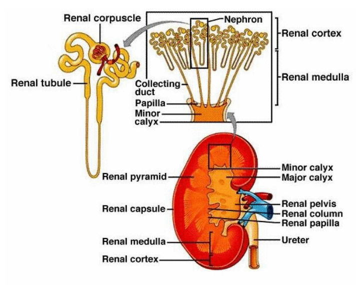
$Y_i = X_{2,t_i} - X_{2,t_{i-1}}$

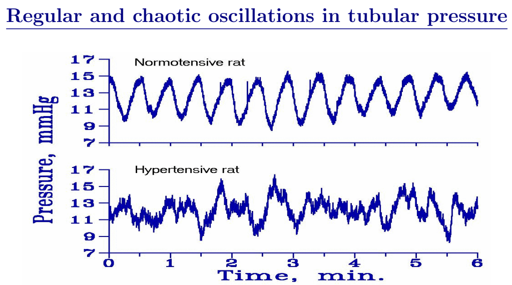
Harmonic oscillator, molecular dynamics

$$\begin{aligned} dX_{1,t} &= -(\beta_1 X_{1,t} + \beta_2 X_{2,t})dt + \gamma dW_t \\ dX_{2,t} &= X_{1,t}dt \end{aligned}$$

Feedback in a rat nephron

Ditlevsen, Yip, Marsh and Holstein-Rathlou (2007)





$$dZ(t) = -\beta(Z(t) - \alpha_0)dt + \sigma dW(t)$$

Feedback in a rat nephron

$$\begin{aligned} dP_t &= \left[F_{\text{filt}}(P_t, r) - F_{\text{reab}} - F_{\text{hen}}(P_t)\right] C^{-1} dt \\ dr &= v dt \\ dv &= \left[\omega^{-1}\left(P_{\text{av}}(P_t, r) - P_{\text{eq}}(r, X_{3,t}, Z_t)\right) - \gamma v\right] dt \\ dX_{1,t} &= 3\left[F_{\text{hen}}(P_t) - X_{1,t}\right] T^{-1} dt \\ dX_{2,t} &= 3(X_{1,t} - X_{2,t}) T^{-1} dt \\ dX_{3,t} &= 3(X_{2,t} - X_{3,t}) T^{-1} dt \end{aligned}$$

$$P_{\text{eq}}(r, X_{3,t}, Z_t) = \\ 2.4 \cdot e^{10(r-1.4)} + 1.6(r-1) + \Psi(X_{3,t}, Z_t)\left(\frac{4.7}{1 + e^{-13(r-0.4)}} + 7.2(r + 0.9)\right)$$

$$\Psi(X_{3,t}, Z_t) = \Psi_{\text{max}} - \frac{\Psi_{\text{max}} - \Psi_{\text{min}}}{1 + \exp\left\{Z_t(X_{3,t}/F_{\text{hen}_0} - S)\right\}}$$

Estimation?

$Y_1, \ldots, Y_n$ is a non-Markovian process

Martingale estimating functions?

$$\sum_{i=1}^{n} a(Y_{i-1}; \theta)\left[Y_i^2 - E_\theta\left(Y_i^2 | Y_{i-1}\right)\right] \ ?$$

$$\sum_{i=1}^{n} a(Y_{i-1}, \ldots, Y_0; \theta)\left[Y_i^2 - E_\theta\left(Y_i^2 | Y_{i-1}, \ldots, Y_0\right)\right] \ ?$$

Another look at martingale estimating functions

Assume that $E_\theta(f(Y_i)^2) < \infty$.

$$G_n(\theta) = \sum_{i=1}^{n} \underbrace{a(Y_{i-1}, Y_{i-2}, \ldots; \theta)}_{\in \mathcal{P}_{i-1}}[f(Y_i) - \underbrace{E_\theta\left(f(Y_i) \,|\, Y_{i-1}, \ldots, Y_1\right)}_{\in \mathcal{P}_{i-1}}]$$

$\mathcal{P}_{i-1}$ = set of all functions $g(Y_{i-1}, \ldots, Y_1)$ such that $E_\theta(g(Y_{i-1}, \ldots, Y_1)^2) < \infty$

This is a set of predictors of $f(Y_i)$ given $Y_1, \ldots, Y_{i-1}$ in which

$$E_\theta\left(f(Y_i) \,|\, Y_{i-1}, \ldots, Y_1\right)$$

is the minimum mean square error predictor of $f(Y_i)$

= the L^2-projection of $f(Y_i)$ onto $\mathcal{P}_{i-1}$

Prediction-based estimating functions

Sørensen (2000)

We still assume that $E_\theta(f(Y_i)^2) < \infty$

$$G_n(\theta) = \sum_{i=1}^{n} \underbrace{\pi_{i-1}(Y_{i-1}, \ldots, Y_1; \theta))}_{\in \mathcal{P}_{i-1}}[f(Y_i) - \underbrace{\bar{\pi}_{i-1}(Y_{i-1}, \ldots, Y_1; \theta)}_{\in \mathcal{P}_{i-1}}],$$

$\mathcal{P}_{i-1}$ = linear space of square integrable predictors of $f(Y_i)$ given $Y_1, \ldots, Y_{i-1}$

$$\begin{aligned} \bar{\pi}_{i-1}(\theta) &= \text{minimum mean square error predictor of } f(Y_i) \text{ in } \mathcal{P}_{i-1} \\ &= L^2\text{-projection of } f(Y_i) \text{ onto } \mathcal{P}_{i-1} \end{aligned}$$

Normal equations for $\bar{\pi}_{i-1}$:

$$E_\theta\left(\pi_{i-1}[f(Y_i) - \bar{\pi}_{i-1}(\theta)]\right) = 0 \quad \text{for all } \pi_{i-1} \in \mathcal{P}_{i-1}$$

Thus $E_\theta(G_n(\theta)) = 0$

An example

$$f(x) = x^2 \qquad \theta \in \mathbb{R} \qquad E_\theta(Y_i^4) < \infty.$$

$$\pi_{i-1} = a_0 + a_1 Y_{i-1}^2 + \cdots + a_q Y_{i-q}^2$$

The minimum mean square error predictor of $f(Y_i)$ is given by

$$\bar{\pi}_{i-1}(\theta) = \bar{a}_0(\theta) + \bar{a}_1(\theta) Y_{i-1}^2 + \cdots + \bar{a}_q(\theta) Y_{i-q}^2$$

where

$$\begin{pmatrix} c_1(\theta) \\ c_2(\theta) \\ \vdots \\ c_q(\theta) \end{pmatrix} = \begin{pmatrix} c_0(\theta) & c_1(\theta) & \cdots & c_{q-1}(\theta) \\ c_1(\theta) & c_0(\theta) & \cdots & c_{q-2}(\theta) \\ \vdots & \vdots & & \vdots \\ c_{q-1}(\theta) & c_{q-2}(\theta) & \cdots & c_0(\theta) \end{pmatrix} \begin{pmatrix} \bar{a}_1(\theta) \\ \bar{a}_2(\theta) \\ \vdots \\ \bar{a}_q(\theta) \end{pmatrix}$$

and

$$\bar{a}_0(\theta) = \mathsf{E}_\theta\left(Y_1^2\right)\left[1 - \bar{a}_1(\theta) - \cdots - \bar{a}_q(\theta)\right]$$

with $c_i(\theta) = \mathsf{Cov}_\theta\left(Y_1^2, Y_{1+i}^2\right), \quad i = 0, 1, \ldots, q$

Optimal estimating function

$$\sum_{i=q+1}^{n} \left(a_0 + a_1 Y_{i-1}^2 + \cdots + a_q Y_{i-q}^2\right)\left[Y_i^2 - \bar{a}_0(\theta) - \bar{a}_1(\theta) Y_{i-1}^2 - \cdots - \bar{a}_q(\theta) Y_{i-q}^2\right]$$

Optimal choice of $a_0, a_1, \ldots, a_q$?

Suppose that $E_\theta(Y_i^8) < \infty$.

We want weights $a_0^*(\theta), a_1^*(\theta), \ldots, a_q^*(\theta)$ such that the estimating function

$$\begin{aligned} G_n^*(\theta) &= \sum_{i=q+1}^{n} \left(a_0^*(\theta) + a_1^*(\theta) Y_{i-1}^2 + \cdots + a_q^*(\theta) Y_{i-q}^2\right) \\ &\qquad \times \left[Y_i^2 - \bar{a}_0(\theta) - \bar{a}_1(\theta) Y_{i-1}^2 - \cdots - \bar{a}_q(\theta) Y_{i-q}^2\right] \end{aligned}$$

satisfies

$$-E_\theta(\partial_\theta G_n(\theta)) = E_\theta(G_n(\theta) G_n^*(\theta))$$

for all G_n in our class of estimating functions

Optimal estimating function

$$\begin{aligned} H_0^{(i)}(\theta) &= Y_i^2 - \bar{a}_0(\theta) - \bar{a}_1(\theta) Y_{i-1}^2 - \cdots - \bar{a}_q(\theta) Y_{i-q}^2 \\ H_k^{(i)}(\theta) &= Y_{i-k}^2 \left[Y_i^2 - \bar{a}_0(\theta) - \bar{a}_1(\theta) Y_{i-1}^2 - \cdots - \bar{a}_q(\theta) Y_{i-q}^2\right], \end{aligned}$$

$k = 1, \ldots, q,$

$$H^{(i)} = \left(H_0^{(i)}(\theta), \ldots, H_q^{(i)}(\theta)\right)^T$$

Then

$$G_n(\theta) = a(\theta)^T \sum_{i=q+1}^{n} H^{(i)}(\theta),$$

so that

$$E_\theta(G_n(\theta) G_n^*(\theta)) = a(\theta)^T M_n(\theta) a^*(\theta)$$

with $a(\theta)^T = (a_0(\theta), a_1(\theta), \ldots, a_q(\theta))$ and $a^*(\theta) = (a_0^*(\theta), a_1^*(\theta), \ldots, a_q^*(\theta))^T$ and

$$M_n(\theta) = \text{ the covariance matrix of } \sum_{i=q+1}^{n} H^{(i)}(\theta)$$

Optimal estimating function

$$\begin{aligned} M_n(\theta) = (n-q) E_\theta\left(H^{(q)}(\theta) H^{(q)}(\theta)^T\right) + \\ \sum_{k=1}^{n-q-1} (n-q-k)\left[E_\theta\left(H^{(q)}(\theta) H^{(q+k)}(\theta)^T\right) + E_\theta\left(H^{(q+k)}(\theta) H^{(q)}(\theta)^T\right)\right] \end{aligned}$$

Usually advisable to use the optimal weights calculated at a preliminary consistent estimator $\bar{\theta}$

$$(a_0^*(\bar{\theta}), a_1^*(\bar{\theta}), \ldots, a_q^*(\bar{\theta}))$$

$\bar{\theta}$ e.g. the estimator given by

$$\sum_{i=q+1}^{n} \left[Y_i^2 - \bar{a}_0(\bar{\theta}) - \bar{a}_1(\bar{\theta}) Y_{i-1}^2 - \cdots - \bar{a}_q(\bar{\theta}) Y_{i-q}^2\right] = 0$$

Optimal estimating function

$$\begin{aligned}\partial_\theta G_n(\theta) &= \sum_{i=q+1}^{n} \left(\partial_\theta a_0(\theta) + \partial_\theta a_1(\theta) Y_{i-1}^2 + \cdots + \partial_\theta a_q(\theta) Y_{i-q}^2\right) \\ &\qquad \times \left[Y_i^2 - \bar{a}_0(\theta) - \bar{a}_1(\theta) Y_{i-1}^2 - \cdots - \bar{a}_q(\theta) Y_{i-q}^2\right] \\ &- \sum_{i=q+1}^{n} \left(a_0(\theta) + a_1(\theta) Y_{i-1}^2 + \cdots + a_q(\theta) Y_{i-q}^2\right) \\ &\qquad \times \left[\partial_\theta \bar{a}_0(\theta) + \partial_\theta \bar{a}_1(\theta) Y_{i-1}^2 + \cdots + \partial_\theta \bar{a}_q(\theta) Y_{i-q}^2\right]\end{aligned}$$

$$-E_\theta(\partial_\theta G_n(\theta)) = (n-q)\, a(\theta)^T \bar{C}(\theta) \partial_\theta \bar{a}(\theta)$$

where

$$\bar{C}(\theta) = E_\theta\left(ZZ^T\right) \qquad Z^T = (1, Y_q^2, \ldots, Y_1^2)$$

$$\bar{a}(\theta)^T = \left(\bar{a}_1(\theta)^T, \ldots, \bar{a}_N(\theta)^T\right)$$

Optimal estimating function

$$\begin{aligned}-E_\theta(\partial_\theta G_n(\theta)) &= (n-q)\, a(\theta)^T \bar{C}(\theta) \partial_\theta \bar{a}(\theta) \\ &= a(\theta)^T M_n(\theta) a^*(\theta) = E_\theta(G_n(\theta) G_n^*(\theta))\end{aligned}$$

for all $a(\theta)$ when

$$M_n(\theta) a^*(\theta) = (n-q) \bar{C}(\theta) \partial_\theta \bar{a}(\theta)$$

or

$$a^*(\theta) = \bar{M}_n(\theta)^{-1} \bar{C}(\theta) \partial_\theta \bar{a}(\theta)$$

with

$$\bar{M}_n(\theta) = \frac{1}{n-q} M_n(\theta)$$

More general prediction-based estimating functions

$$G_n(\theta) = \sum_{i=s+1}^{n} \sum_{j=1}^{N} \Pi_j^{(i-1)}(\theta) \left[f_j(Y_i) - \breve{\pi}_j^{(i-1)}(\theta)\right]$$

$$\Pi_j^{(i-1)}(\theta) = \begin{pmatrix} \pi_{1,j}^{(i-1)}(\theta) \\ \vdots \\ \pi_{p,j}^{(i-1)}(\theta) \end{pmatrix}$$

$$\theta \in \Theta \subseteq \mathbb{R}^p$$

$$G_n(\theta) = A(\theta) \sum_{i=s+1}^{n} H^{(i)}(\theta)$$

A useful estimating function

Often a reasonable estimating function is given by

$$\sum_{i=1}^{n} \left\{ \pi_1^{(i-1)}(\theta) \left[Y_i - \breve{\pi}_1^{(i-1)}(\theta)\right] + \pi_2^{(i-1)}(\theta) \left[Y_i^2 - \breve{\pi}_2^{(i-1)}(\theta)\right] \right\}$$

$$\begin{aligned}\pi_1^{(i-1)} &= \alpha_{1,0} + \alpha_{1,1} Y_{i-1} + \cdots + \alpha_{1,s} Y_{i-s} \\ \pi_2^{(i-1)} &= \alpha_{2,0} + \alpha_{2,1} Y_{i-1} + \cdots + \alpha_{2,s} Y_{i-s} + \alpha_{2,s+1} Y_{i-1}^2 + \cdots + \alpha_{2,2s} Y_{i-s}^2\end{aligned}$$

To calculate the minimum mean square error predictors, we need

$$E_\theta(Y_1^\kappa Y_k^j), \quad 0 \le \kappa \le j \le 2, \quad k = 1, \ldots, s.$$

To calculate the optimal estimating function, we need

$$E_\theta[Y_{t_1}^{k_1} Y_{t_2}^{k_2} Y_{t_3}^{k_3} Y_{t_4}^{k_4}]$$

$t_1 \le t_2 \le t_3 \le t_4$ and $k_1 + k_2 + k_3 + k_4 \le 8$, where $k_i \in \mathbb{N}_0$

Measurement errors

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t$$

Data: $Y_i = X_{t_i} + Z_i, \quad i = 1, \ldots, n,$

Z_i are independent and identically distributed and independent of X

$$\begin{aligned} E_\theta(Y_1^{k_1} Y_2^{k_2}) &= E_\theta\left((X_{t_1}+Z_1)^{k_1}(X_{t_2}+Z_2)^{k_2}\right) \\ &= \sum_{i_1=0}^{k_1}\sum_{i_2=0}^{k_2}\binom{k_1}{i_1}\binom{k_2}{i_2} E_\theta(X_{t_1}^{i_1}X_{t_2}^{i_2}) E_\theta(Z_1^{k_1-i_1}) E_\theta(Z_2^{k_2-i_2}). \end{aligned}$$

$$E_\theta(X_{t_1}^{i_1}X_{t_2}^{i_2}) = E_\theta(X_{t_1}^{i_1} E_\theta(X_{t_2}^{i_2} \,|X_{t_1}\,))$$

For Pearson diffusions

$$E_\theta(X_{t_1}^{i_1} E_\theta(X_{t_2}^{i_2} \,|X_{t_1}\,)) = \sum_{k=0}^{i_2}\sum_{\ell=0}^{i_2} q_{i_2,k,\ell} e^{-\lambda_\ell(t_2-t_1)} E_\theta(X_{t_1}^{i_1+k})$$

Sums of Pearson diffusions

Explicit optimal prediction-based estimating functions
Forman & Sørensen (2006)

Sum of 2 skew t-diffusions

$$dX_{i,t} = -\beta_i X_{i,t}dt + \sqrt{2\beta_i(\nu_i-1)^{-1}\{X_{i,t}^2 + 2\rho\sqrt{\nu_i}X_{i,t} + (1+\rho^2)\nu_i\}}dW_{i,t}$$

$$Y_t = X_{1,t} + X_{2,t}$$

Assume β_1, β_2 and $\phi_1 = \nu_1(\nu_1-2)^{-1}/\{\nu_1(\nu_1-2)^{-1} + \nu_2(\nu_2-2)^{-1}\}$ known

$$\Delta = 1$$

$$f(x) = x^2 \qquad \pi_{i-1} = a_0 + a_1 Y_{i-1}$$

Sum of skew t-diffusions

$$\sum_{i=2}^{n}\left[\begin{array}{c} Y_i^2 - \sigma^2 - \kappa Y_{i-1} \\ Y_{i-1}(Y_i^2 - \sigma^2 - \kappa Y_{i-1}) \end{array}\right] = 0$$

$$\sigma^2 = (1+\rho^2)\left\{\frac{\nu_1}{\nu_1-2} + \frac{\nu_2}{\nu_2-2}\right\}, \quad \kappa = 4\rho\left\{\frac{\sqrt{\nu_1}}{\nu_1-3}\phi_1 e^{-\theta_1\Delta} + \frac{\sqrt{\nu_2}}{\nu_2-3}\phi_2 e^{-\theta_2\Delta}\right\}$$

$$\begin{aligned} \hat\kappa &= \frac{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}Y_i^2 - (\frac{1}{n-1}\sum_{i=2}^n Y_{i-1})(\frac{1}{n-1}\sum_{i=2}^n Y_i^2)}{\frac{1}{n-1}\sum_{i=2}^n Y_{i-1}^2 - (\frac{1}{n-1}\sum_{i=2}^n Y_{i-1})^2} \\ \hat\sigma^2 &= \frac{1}{n-1}\sum_{i=2}^n Y_i^2 + \hat\kappa\,\frac{1}{n-1}\sum_{i=2}^n Y_{i-1} \end{aligned}$$

Integrated diffusions

Ditlevsen & Sørensen (2004)

A reasonable prediction based estimating function:

$$\sum_{i=1}^{n}\left\{\pi_1^{(i-1)}(\theta)\left[Y_i - \bar\pi_1^{(i-1)}(\theta)\right] \;+\; \pi_2^{(i-1)}(\theta)\left[Y_i^2 - \bar\pi_2^{(i-1)}(\theta)\right]\right\}$$

$$\begin{aligned} \pi_1^{(i-1)} &= \alpha_{1,0} + \alpha_{1,1}Y_{i-1} + \cdots + \alpha_{1,s_1}Y_{i-s_1} \\ \pi_2^{(i-1)} &= \alpha_{2,0} + \alpha_{2,1}Y_{i-1} + \cdots + \alpha_{2,s_2}Y_{i-s_2} + \alpha_{2,s_2+1}Y_{i-1}^2 + \cdots + \alpha_{2,2s_2}Y_{i-s_2}^2 \end{aligned}$$

Integrated square root process

$$dX_t = -\beta(X_t - \mu)dt + \sigma\sqrt{X_t}\, dW_t.$$

$$\pi_1^{(i-1)} = \alpha_{1,0} + \alpha_{1,1}Y_{i-1} \qquad \pi_2^{(i-1)} = \alpha_{2,0}$$

$$\begin{aligned}\bar{\pi}_1^{(i-1)}(Y_{i-1};\theta) &= \mu\,(1 - \bar{a}_{11}(\beta)) + \bar{a}_{11}(\beta)Y_{i-1} \\ \bar{\pi}_2^{(i-1)}(\theta) &= \mu^2 + \mu\sigma^2\beta^{-3}\Delta^{-2}(e^{-\beta\Delta} - 1 + \beta\Delta)\end{aligned}$$

$$\bar{a}_{11}(\beta) = \frac{(1 - e^{-\beta\Delta})^2}{2(\beta\Delta - 1 + e^{-\beta\Delta})}$$

Integrated square root process

$$\sum_{i=1}^{n} \begin{pmatrix} 1 \\ Y_{i-1} \\ 0 \end{pmatrix} [Y_i - \bar{\pi}_1^{(i-1)}(Y_{i-1};\theta)] + \sum_{i=1}^{n} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} [Y_i^2 - \bar{\pi}_2^{(i-1)}(\theta)] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{aligned}\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} Y_i + \frac{a_{11}(\hat{\beta})Y_n - Y_1}{(n-1)(1 - a_{11}(\hat{\beta}))} \\ \sum_{i=2}^{n} Y_{i-1}Y_i &= \hat{\mu}(1 - a_{11}(\hat{\beta}))\sum_{i=2}^{n} Y_{i-1} + a_{11}(\hat{\beta})\sum_{i=2}^{n} Y_{i-1}^2 \\ \hat{\sigma}^2 &= \frac{\hat{\beta}^3\Delta^2\sum_{i=2}^{n}\left(Y_i^2 - \hat{\mu}^2\right)}{(n-1)\hat{\mu}(e^{-\hat{\beta}\Delta} - 1 + \hat{\beta}\Delta)}\end{aligned}$$

A Pearson Stochastic Volatility Model

$$dX_t = (\alpha + \beta V_t)dt + \sqrt{V_t}dW_t$$

Volatility process:

$$V_t = V_{1,t} + \cdots + V_{m,t}$$

$$dV_{i,t} = -\beta_i(V_{i,t} - \mu_i)dt + \sqrt{2\beta(a_iV_{i,t}^2 + b_iV_{i,t} + c_i)}dB_{i,t}$$

$W, B_1, \ldots, B_m$ are independent standard Wiener processes

DATA: $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$

$Y_i = X_{i\Delta} - X_{(i-1)\Delta} = \alpha\Delta + \beta S_i + \sqrt{S_i}A_i$

$$S_i = \int_{(i-1)\Delta}^{i\Delta} V_t dt \qquad A_i \sim N(0,1), \text{ i.i.d.}$$

Explicit optimal prediction based estimating functions:
Forman & Sørensen (2006)

Example: Generalized Heston Model

$$dV_t^{(i)} = -\theta_i\left(V_t^{(i)} - \varphi_i\lambda\sigma\right)dt + \sqrt{2\sigma\theta_iV_t^{(i)}}dB_t^{(i)}$$

$\lambda > 0$, $\sigma > 0$, $\varphi_i > 0$, and $\varphi_1 + \cdots + \varphi_m = 1$

V_t gamma distributed with shape parameter λ and scale parameter σ

Autocorrelation function of V:

$$\rho(u) = \varphi_1\exp(-\theta_1u) + \cdots + \varphi_m\exp(-\theta_mu)$$

Bibby and Sørensen (2004)

Generalization of Heston (1993) ($m = 1$).

Diffusion compartment models

$$dX_t = (B(\theta)X_t - b(\theta))dt + \sigma(X_t;\theta)dW_t$$

X, $b(\theta)$ and W D-dimensional
$B(\theta)$ and $\sigma(x;\theta)$ $D \times D$-matrices

$B(\theta)$ flow between compartments
$b(\theta)$ input/output

DATA: e.g. $Y_i = X_{1,t_i} + Z_i$

A Gaussian diffusion compartment models

$$\begin{pmatrix} dX_{1,t} \\ dX_{2,t} \end{pmatrix} = \begin{pmatrix} -\beta_1 & \beta_2 \\ \beta_1 & -(\beta_2+\beta_3) \end{pmatrix} \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \tau_1 & 0 \\ 0 & \tau_2 \end{pmatrix} \begin{pmatrix} dW_{1,t} \\ dW_{2,t} \end{pmatrix}$$

$\beta_1, \beta_2, \beta_3, \tau_1, \tau_2 > 0$
Data: $Y_i = X_{1,i\Delta}$, $i = 1, \ldots, n$

Likelihood inference: Bibby (1995)

Explicit optimal prediction-based estimating function: Düring (2002)

$$\sum_{i=1}^{n} \left\{ \pi_1^{(i-1)}(\theta) \left[Y_i - \bar{\pi}_1^{(i-1)}(\theta) \right] + \pi_2^{(i-1)}(\theta) \left[Y_i^2 - \bar{\pi}_2^{(i-1)}(\theta) \right] \right\}$$

$$\begin{aligned} \pi_1^{(i-1)} &= \alpha_{1,0} + \alpha_{1,1}Y_{i-1} + \cdots + \alpha_{1,s}Y_{i-s} \\ \pi_2^{(i-1)} &= \alpha_{2,0} + \alpha_{2,1}Y_{i-1} + \cdots + \alpha_{2,s}Y_{i-s} + \alpha_{2,s+1}Y_{i-1}^2 + \cdots + \alpha_{2,2s}Y_{i-s}^2 \end{aligned}$$

A non-Gaussian diffusion compartment model

Joint work with Maria Düring

$$\begin{pmatrix} dX_{1,t} \\ dX_{2,t} \end{pmatrix} = \left[\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} -\beta_1 & \beta_2 \\ \beta_1 & -(\beta_2+\beta_3) \end{pmatrix} \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \end{pmatrix} \right] dt + \begin{pmatrix} \tau_1\sqrt{X_{1,t}} & 0 \\ 0 & \tau_2\sqrt{X_{2,t}} \end{pmatrix} \begin{pmatrix} dW_{1,t} \\ dW_{2,t} \end{pmatrix}$$

$\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2 > 0$

Condition for existence of a unique, strictly positive solution:
$\alpha_i \geq \frac{1}{2}\sigma_i^2$, $i = 1, 2$

Data: $Y_i = X_{1,i\Delta}$, $i = 1, \ldots, n$

A non-Gaussian diffusion compartment model

A reasonable prediction based estimating function:

$$\sum_{i=1}^{n} \left\{ \pi_1^{(i-1)}(\theta) \left[Y_i - \bar{\pi}_1^{(i-1)}(\theta) \right] + \pi_2^{(i-1)}(\theta) \left[Y_i^2 - \bar{\pi}_2^{(i-1)}(\theta) \right] \right\}$$

$$\begin{aligned} \pi_1^{(i-1)} &= \alpha_{1,0} + \alpha_{1,1}Y_{i-1} + \cdots + \alpha_{1,s}Y_{i-s} \\ \pi_2^{(i-1)} &= \alpha_{2,0} + \alpha_{2,1}Y_{i-1} + \cdots + \alpha_{2,s}Y_{i-s} + \alpha_{2,s+1}Y_{i-1}^2 + \cdots + \alpha_{2,2s}Y_{i-s}^2 \end{aligned}$$

The predictors can by found explicitly, e.g. by the Durbin-Levinson algorithm, see e.g. Brockwell and Davis (1991)

The moments needed for the optimal estimating function must be found numerically

By methods in Down, Meyn and Tweedie (1995) the process can be shown to be geometrically α-mixing

Stochastic delay differential equations

$$dX_t = \int_{-r}^{0} X_{t+s} a_\theta(ds) dt + \sigma dW_t$$

a_θ is a measure on $[-r, 0]$

$$dX_t = \sum_{k=1}^{N} \theta_k X_{t-r_k} dt + \sigma dW_t$$

$$dX_t = -b \int_{-r}^{0} X_{t+s} e^{as} ds\, dt + \sigma dW_t$$

Joint work with Uwe Küchler (Humboldt-University of Berlin):

Pseudo-likelihood, prediction-based estimating functions

Non-martingale central limit theorem

Suppose $\sum_{i=r}^{n} g(X_{i\Delta}, X_{(i-1)\Delta}, \ldots, X_{(i-r)\Delta})$

- X is stationary and geometrically α-mixing
- $E\left(g(X_{r\Delta}, \ldots, X_0)\right) = 0$
- $E\left((g_j(X_{r\Delta}, \ldots, X_0))^{2+\delta}\right) < \infty$ for a $\delta > 0$, $j = 1, \ldots, p$

Then (see e.g. Doukhan, 1994)

$$M = E\left(g(X_{r\Delta}, \ldots, X_0) g(X_{r\Delta}, \ldots, X_0)^T\right) + \sum_{k=1}^{\infty} \left[E\left(g(X_{r\Delta}, \ldots, X_0) g(X_{(r+k)\Delta}, \ldots, X_{k\Delta})^T\right) + E_\theta\left(g(X_{(r+k)\Delta}, \ldots, X_{k\Delta}) g(X_{r\Delta}, \ldots, X_0)^T\right)\right]$$

converges, and if the limit M is strictly positive definite, then

$$\frac{1}{\sqrt{n}} \sum_{i=r}^{n} g(X_{i\Delta}, \ldots, X_{(i-r)\Delta}) \xrightarrow{\mathcal{D}} N_p(0, M) \quad \text{as} \quad n \to \infty$$

Geometric α-mixing

$\mathcal{F}_t = \sigma\{X_s \,|\, s \le t\}$ $\qquad$ $\mathcal{F}^t = \sigma\{X_s \,|\, s \ge t\}$

A stochastic process X is called α-mixing, if

$$\sup_{A \in \mathcal{F}_t, B \in \mathcal{F}^{t+u}} |P_{\theta_0}(A) P_{\theta_0}(B) - P_{\theta_0}(A \cap B)| \le \alpha(u)$$

for all $t > 0$ and $u > 0$, where

$$\alpha(u) \to 0 \qquad \text{as } u \to \infty.$$

X is called geometrically α-mixing is there exist $c_1, c_2 > 0$ such that

$$\alpha(u) \le c_1 e^{-c_2 u}, \qquad \text{for all } u > 0.$$

Geometrically α-mixing one-dimensional diffusions

$$dX_t = b(X_t; \theta) dt + \sigma(X_t; \theta) dW_t$$

Conditions for geometric α-mixing

(i) The function b is continuously differentiable with respect to x and σ is twice continuously differentiable respect to x, $\sigma(x; \theta) > 0$ for all $x \in (\ell, r)$, and there exists a constant $K_\theta > 0$ such that $|b(x; \theta)| \le K_\theta(1 + |x|)$ and $\sigma^2(x; \theta) \le K_\theta(1 + x^2)$ for all $x \in (\ell, r)$

(ii) $\sigma(x; \theta)\mu_\theta(x) \to 0$ as $x \downarrow \ell$ and $x \uparrow r$

(iii) $1/\gamma(x; \theta)$ has a finite limit as $x \downarrow \ell$ and $x \uparrow r$, where $\gamma(x; \theta) = \partial_x \sigma(x; \theta) - 2b(x; \theta)/\sigma(x; \theta)$

Genon-Catalot, Jeantheau & Larédo (2000)

$\mu_\theta(x)$ density of invariant measure