Prediction-based estimating functions for integrated diffusions, stochastic delay differential equations and diffusion compartment models

Michael Sørensen

Department of Applied Mathematics and Statistics
University of Copenhagen
Denmark

www.math.ku.dk/∼michael

Martingale estimating functions

DATA: $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$

MODEL: Markov process model parametrized by $\theta \in \Theta \subseteq \mathbb{R}^p$

ESTIMATING FUNCTION:

$$G_n(\theta) = \sum_{i=1}^{n} a(X_{t_{i-1}};\theta)\left[f(X_{t_i}) - E_\theta\left(f(X_{t_i})|\, X_{t_{i-1}}\right)\right]$$

θ-martingale

ESTIMATOR: $G_n(\hat{\theta}) = 0$

Simple Stochastic Volatility Model

$$\begin{aligned} dX_t &= \sqrt{v_t} dW_t \\ dv_t &= b(v_t;\theta)dt + c(v_t;\theta)dB_t \end{aligned}$$

W and B are independent standard Wiener processes

We assume that v_t is stationary

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

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

Simple Stochastic Volatility Model

$$Y_i = X_{i\Delta} - X_{(i-1)\Delta} = \int_{(i-1)\Delta}^{i\Delta} \sqrt{v_t} dW_t$$

$\{Y_i\}$ is a stationary process

$$Y_i = \sqrt{S_i} Z_i \qquad \text{where} \qquad S_i = \int_{(i-1)\Delta}^{i\Delta} v_t dt,$$

and where $\{Z_i\}$ are i.i.d., independent of $\{S_i\}$, and $Z_i \sim N(0,1)$.

Martingale estimating functions:

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

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

Sums of diffusion processes

$$Y_t = \sum_{i=1}^{m} X_t^{(i)}$$

$$dX_t^{(i)} = -\theta_i(X_t^{(i)} - \mu_i)dt + \phi_i(X_t^{(i)})dW_t^{(i)},$$

$\theta_i > 0$, $i = 1, \ldots, m$

$W^{(i)}$, $i = 1, \ldots, m$, are independent Wiener processes

Bibby, Skovgaard, and Sørensen (2003)

Jacobsen and Sørensen (2004)

DATA: $Y_0, Y_\Delta, Y_{2\Delta}, \ldots, Y_{n\Delta}$

Diffusion compartment models

$$dX_t = A(\theta)X_t dt + \sigma(X_t;\theta)dW_t,\ X_0 = x_0$$

$$X = \left(X_t^{(1)}, \ldots, X_t^{(d)}\right)$$

$X_t^{(j)}$ is the amount of material in the jth compartment at time t

$Y = \left(X_t^{(1)}, \ldots, X_t^{(m)}\right)$ ($m < d$)

DATA: $Y_0, Y_\Delta, Y_{2\Delta}, \ldots, Y_{n\Delta}$

Bibby (1995) proposed an algorithm similar to the EM-algorithm based on martingale estimating functions for estimation

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 $f(Y_i)$ on $\mathcal{P}_{i-1}$

Prediction-based estimating functions

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

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

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

where $\pi_{i-1}(\theta) \in \mathcal{P}_{i-1}$, and

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

Normal equations:

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

Hence $G_n(\theta)$ is an unbiased estimating function

Stochastic volatility model

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

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

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

Stochastic volatility model: Moments

Barndorff-Nielsen and Shephard (1998):

$$\xi(\theta) = E_\theta(v_t) \qquad \omega(\theta) = \mathrm{Var}_\theta(v_t) \qquad r(u;\theta) = \mathrm{Cov}_\theta(v_t, v_{t+u})/\omega(\theta)$$

$$\begin{aligned} \mathrm{E}_\theta(Y_1^2) &= \Delta\xi(\theta) \\ \mathrm{Var}_\theta\left(Y_1^2\right) &= 6\omega(\theta)R^*(\Delta;\theta) + 2\Delta^2\xi(\theta)^2 \end{aligned}$$

where $\qquad R^*(t;\theta) = \int_0^t \int_0^s r(u;\theta)duds$

$$\begin{aligned} \mathrm{Cov}_\theta(Y_1^2, Y_{1+i}^2) &= \omega(\theta)\left[R^*(\Delta(i+1);\theta) - 2R^*(\Delta i;\theta) + R^*(\Delta(i-1);\theta)\right] \\ &= \omega(\theta)\int_{(i-1)\Delta}^{i\Delta}\int_s^{s+\Delta} r(u;\theta)duds \end{aligned}$$

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

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

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

$$M_n(\theta) = (n-q)E_\theta\left(H^{(q)}(\theta)H^{(q)}(\theta)^T\right) +$$

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

Usually advisible 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

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

$$\bar{a}(\theta)^T = (\bar{a}_0(\theta), \ldots, \bar{a}_N(\theta))$$

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

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

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

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

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

More general prediction-based estimating functions

$$\pi_{\ell,j}^{(i-1)}(\theta) = \sum_{k=0}^{q_j} a_{\ell jk}(\theta) Z_{jk}^{(i-1)}$$

$$Z_{jk}^{(i-1)} = h_{jk}(Y_{i-1}, \ldots, Y_{i-s}), \quad k = 1, \ldots, q_j, \quad j = 1, \ldots, N$$

$$\bar{\pi}_{\ell,j}^{(i-1)}(\theta) = \sum_{k=0}^{q_j} \bar{a}_{\ell jk}(\theta) Z_{jk}^{(i-1)}$$

$$\bar{a}(\theta) = \left(\bar{a}_{111}(\theta), \ldots, \bar{a}_{pq_N N}(\theta)\right)^T ,$$

Consistency and asymptotic normality

CONDITIONS:

- Y is stationary and α-mixing with mixing coefficients $\alpha_k(\theta)$, $k = 1, 2, \ldots$, and that there exists a $\delta > 0$ such that

$$\sum_{k=1}^{\infty} \alpha_k(\theta)^{\delta/(2+\delta)} < \infty$$

and

$$E_{\theta_0}\left(\left|H^{(r)}(\theta)_j\right|^{2+\delta}\right) < \infty \quad \text{for all} \quad j$$

- The matrix

$$M(\theta) = E_\theta\left(H^{(r)}(\theta)H^{(r)}(\theta)^T\right) + \\ \sum_{k=1}^{\infty}\left[E_\theta\left(H^{(r)}(\theta)H^{(r+k)}(\theta)^T\right) + E_\theta\left(H^{(r+k)}(\theta)H^{(r)}(\theta)^T\right)\right]$$

is strictly positive definite

Consistency and asymptotic normality

CONDITIONS:

- The vector $\bar{a}(\theta)$ and the matrix $A(\theta)$ are twice continuously differentiable with respect to θ

- The matrices $\partial_{\theta^T}\bar{a}(\theta_0)$ and $A(\theta_0)$ have rank p

Consistency and asymptotic normality

Under these conditions:

For every $n \geq s+1$, an estimator $\hat{\theta}_n$ exists that solves the estimating equation $G_n(\hat{\theta}_n) = 0$ with a probability tending to one as $n \to \infty$. Moreover,

$$\hat{\theta}_n \to \theta_0$$

in probability and

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

as $n \to \infty$

Stochastic Volatility Model

$$dX_t = \sqrt{v_t} dW_t, \quad dv_t = b(v_t)dt + c(v_t)dB_t$$

Suppose the volatility process v is α-mixing with mixing coefficients $\alpha_t(\theta_0)$, $t > 0$. Then Y is α-mixing with the mixing coefficients $\tilde{\alpha}_k(\theta_0) < \alpha_k(\theta_0)$, $k = 1, 2 \dots$

$$\begin{aligned}
&|P\left((Y_1, \dots, Y_t) \in M_1\right) P\left((Y_{t+k}, \dots, Y_{t+k+\ell}) \in M_2\right) - \\
- \quad & P\left(\{(Y_1, \dots, Y_t) \in M_1\} \cap \{(Y_{t+k}, \dots, Y_{t+k+\ell}) \in M_2\}\right)| \\
= \quad & \Bigg| E\left[P\left(\left(\sqrt{S_1} Z_1, \dots, \sqrt{S_t} Z_t\right) \in M_1 \Big| \mathcal{F}^v\right)\right] \\
& \times E\left[P\left(\left(\sqrt{S_{t+k}} Z_{t+k}, \dots, \sqrt{S_{t+k+\ell}} Z_{t+k+\ell}\right) \in M_2 \Big| \mathcal{F}^v\right)\right] - \\
& - E\left[P\left(\left\{\left(\sqrt{S_1} Z_1, \dots, \sqrt{S_t} Z_t\right) \in M_1\right\}\right.\right. \\
& \cap \quad \left.\left.\left\{\left(\sqrt{S_{t+k}} Z_{t+k}, \dots, \sqrt{S_{t+k+\ell}} Z_{t+k+\ell}\right) \in M_2\right\} \Big| \mathcal{F}^v\right)\right] \Bigg| \\
= \quad & |E\left[f_1\left(S_1, \dots, S_t\right)\right] E\left[f_2\left(S_{t+k}, \dots, S_{t+k+\ell}\right)\right] \\
& \qquad - E\left[f_1\left(S_1, \dots, S_t\right) f_2\left(S_{t+k}, \dots, S_{t+k+\ell}\right)\right]| \\
= \quad & |\text{Cov}\left(f_1\left(S_1, \dots, S_t\right), f_2\left(S_{t+k}, \dots, S_{t+k+\ell}\right)\right)| \leq \alpha_k(\theta_0)
\end{aligned}$$

Integrated diffusions

Joint work with Susanne Ditlevsen

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

Data:

$$Y_i = \int_0^\Delta X_{(i-1)\Delta+s}\, d\nu(s), \;\; i = 1, \ldots, n$$

ν is a probability measure on $[0, \Delta]$

For instance:

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

Integrated diffusions: Estimation

A reasonable prediction based estimating function:

$$
\begin{aligned}
f_1(y) = y \qquad & \pi_1^{(i-1)} &=& \ \alpha_{1,0} + \alpha_{1,1} Y_{i-1} + \cdots + \alpha_{1,s_1} Y_{i-s_1} \\
f_2(y) = y^2 \qquad & \pi_2^{(i-1)} &=& \ \alpha_{2,0} + \alpha_{2,1} Y_{i-1} + \cdots + \alpha_{2,s_2} Y_{i-s_2} \\
& & & + \alpha_{2,s_2+1} Y_{i-1}^2 + \cdots + \alpha_{2,2s_2} Y_{i-s_2}^2
\end{aligned}
$$

The minimum mean square error predictors can by found by the Durbin-Levinson algorithm from moments of the type

$$E_\theta(Y_1^m) = \frac{m!}{\Delta^m} \int_0^\Delta \int_0^{s_1} \cdots \int_0^{s_{m-1}} E_\theta(X_{s_1} \cdots X_{s_m}) ds_m \cdots ds_2 ds_1$$

and

$$E_\theta\left(Y_1^{m_1} Y_k^{m_2}\right) = \Delta^{-(m_1+m_2)} \cdot$$

$$\int_0^\Delta \cdots \int_0^\Delta E_\theta \left(X_{s_1} \cdots X_{s_{m_1}} X_{((k-1)\Delta + s_{(m_1+1)})} \cdots X_{((k-1)\Delta + s_{(m_1+m_2)})} \right) ds_1 \cdots ds_{(m_1+m_2)}$$

Example: CIR-diffusion

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

$$f_1(y) = y \qquad \pi_1^{(i-1)} = \alpha_{1,0} + \alpha_{1,1} Y_{i-1}$$

$$f_2(y) = y^2 \qquad \pi_2^{(i-1)} = \alpha_{2,0}$$

$$\begin{aligned} E_\theta(Y_1^2) &= \mu^2 + \mu\sigma^2\beta^{-3}\Delta^{-2}(e^{-\beta\Delta} - 1 + \beta\Delta) \\ E_\theta(Y_1 Y_k) &= \mu^2 + \tfrac{1}{2}\mu\sigma^2\beta^{-3}\Delta^{-2}(e^{\beta\Delta} - 1)^2 e^{-k\beta\Delta} \qquad \text{for } k > 1 \end{aligned}$$

$$\begin{aligned} \bar{\pi}_1^{(i-1)}(Y_{i-1};\theta) &= \mu\,(1 - \bar{a}_{11}(\beta)) + \bar{a}_{11}(\beta) Y_{i-1} \\ \bar{\pi}_2^{(i-1)}(\theta) &= \bar{a}_{20}(\theta) \end{aligned}$$

$$\bar{a}_{11}(\beta) = \frac{(1 - e^{-\beta\Delta})^2}{2(\beta\Delta - 1 + e^{-\beta\Delta})} \qquad \bar{a}_{20}(\theta) = E_\theta(Y_1^2)$$

Example: CIR-diffusion

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

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

Example: Ornstein-Uhlenbeck process

$$dX_t = -\beta X_t dt + \sigma dW_t$$

$$f_1(y) = y \qquad \pi_1^{(i-1)} = \alpha_{1,0} + \alpha_{1,1} Y_{i-1}$$

$$f_2(y) = y^2 \qquad \pi_2^{(i-1)} = \alpha_{2,0}$$

$$\begin{aligned} E_\theta(Y_1^2) &= \sigma^2 \beta^{-3} \Delta^{-2} (\beta\Delta - 1 + e^{-\beta\Delta}) \\ E_\theta(Y_1 Y_k) &= \tfrac{1}{2} \sigma^2 \beta^{-3} \Delta^{-2} (1 - e^{\beta\Delta})^2 e^{-k\beta\Delta} \qquad \text{for } k > 1 \end{aligned}$$

$$\bar{\pi}_1^{(i-1)}(Y_{i-1};\theta) = \frac{(1 - e^{-\beta\Delta})^2}{2(\beta\Delta - 1 + e^{-\beta\Delta})} Y_{i-1}$$

$$\bar{\pi}_2^{(i-1)}(\theta) = \frac{\sigma^2(\beta\Delta - 1 + e^{-\beta\Delta})}{\beta^3 \Delta^2}$$

Ornstein-Uhlenbeck process: Optimal estimation

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

$$\frac{(1-e^{-\hat{\beta}\Delta})^2}{2(\hat{\beta}\Delta - 1 + e^{-\hat{\beta}\Delta})} = \frac{\sum_{i=2}^{n} Y_{i-1}Y_i}{\sum_{i=2}^{n} Y_{i-1}^2}$$

$$\hat{\sigma}^2 = \frac{\hat{\beta}^3\Delta^2}{\hat{\beta}\Delta - 1 + e^{-\hat{\beta}\Delta}} \frac{1}{n-1} \sum_{i=2}^{n} Y_i^2$$

No solution if $\frac{1}{n}\sum_{i=2}^{n} Y_{i-1}Y_i \leq 0$

Asymptotics

Suppose

$$f_j(y) = y^{\kappa_{j0}} \qquad \pi_j^{(i-1)} = \alpha_{j,0} + \alpha_{j,1} Y_{i-l_{j1}}^{\kappa_{j1}} + \cdots + \alpha_{j,q_j} Y_{i-l_{jq_j}}^{\kappa_{jq_j}}$$

with $\kappa_{jk}, l_{jk} \in \mathbb{N}$, $j = 1, \ldots, N$

Sufficient conditions on X to ensure consistency and asymptotic normality of estimators:

• X is stationary and α-mixing with mixing coefficients $\alpha_t(\theta)$, $t > 0$, satisfying that for some $\delta > 0$

$$\sum_{k=1}^{\infty} \alpha_{k\Delta}(\theta)^{\delta/(2+\delta)} < \infty$$

• $E_\theta\left(|X_0|^{4\kappa+\epsilon}\right) < \infty$ for some $\epsilon > 0$, where $\kappa = \max\{\kappa_{10}, \ldots, \kappa_{Nq_N}\}$

Sufficient condition for geometric α-mixing

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \qquad \text{state space } (\ell, r)$$

$$\int_\ell^{x_0} s(x)dx = \int_{x_0}^r s(x)dx = \infty \;\text{ and }\; \int_\ell^r m(x)dx < \infty,$$

$$s(x) = \exp\left(-2\int_{x_0}^x \frac{b(u)}{\sigma^2(u)}du\right) \qquad m(x) = \frac{1}{\sigma^2(x)s(x)} \qquad x_0 \in (\ell, r)$$

(i) The function b is continuously differentiable and σ is twice continuously differentiable on (ℓ, r), $\sigma(x) > 0$ for all $x \in (\ell, r)$, and there exists a constant $K > 0$ such that $|b(x)| \leq K(1 + |x|)$ and $\sigma^2(x) \leq K(1 + x^2)$ for all $x \in (\ell, r)$

(ii) $\sigma(x)m(x) \to 0$ as $x \downarrow \ell$ and $x \uparrow r$

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

Genon-Catalot, Jeantheau and Larédo (2000) (Bernoulli)

Stochastic delay differential equations

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

$$dX_t = \int_{-r}^{0} X_{t+s} a_\theta(ds) dt + \sigma dW_t$$ a_θ is a measure on $[-r, 0]$

Initial condition: The distribution of $\{X_s \mid s \in [-r, 0]\}$ is the stationary distribution.

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

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

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

Pseudo likelihood function

$$
\begin{aligned}
\tilde{L}_n(\theta) &= \prod_{i=k}^{n-1} p(X_{(i+1)\Delta} | X_{i:i+1-k}; \theta) \\
&= \prod_{i=k}^{n-1} \frac{\exp\left(-\frac{1}{2v_k(\theta)}\left(X_{(i+1)\Delta} - \phi_k(\theta)^T X_{i:i+1-k}\right)^2\right)}{\sqrt{2\pi v_k(\theta)}}
\end{aligned}
$$

$$X_{i:j} = (X_{i\Delta}, \ldots, X_{j\Delta})^T \; i > j \geq 1$$

The pseudo score function

$$
\begin{aligned}
\partial_\theta \log(\tilde{L}_n(\theta)) &= \sum_{i=k}^{n-1} \frac{\partial_\theta \phi_k(\theta)^T X_{i:i+1-k}}{v_k(\theta)} \left(X_{(i+1)\Delta} - \phi_k(\theta)^T X_{i:i+1-k}\right) \\
&\quad + \frac{\partial_\theta v_k(\theta)}{2v_k(\theta)^2} \sum_{i=k}^{n-1} \left[\left(X_{(i+1)\Delta} - \phi_k(\theta)^T X_{i:i+1-k}\right)^2 - v_k(\theta)\right]
\end{aligned}
$$

is a prediction-based estimating function

How to calculate the pseudo score function

The quantities $\phi_k(\theta)$, $v_k(\theta)$, $\partial_\theta \phi_k(\theta)$, and $\partial_\theta v_k(\theta)$ can be found from $K_\theta(t) = E_\theta(X_0 X_t)$ by the Durbin-Levinson algorithm and its derivative v.r.t. θ

The covariance function

$$K_\theta(t) = E_\theta(X_0 X_t), \quad t \geq 0$$

satisfies the continuous time analogue of the Yule-Walker equations

$$\partial_t K_\theta(t) = \int_{-r}^{0} K_\theta(t+s) a_\theta(ds)$$

Example 1

$$dX_t = [aX_t + bX_{t-r}]\,dt + \sigma dW_t \qquad\qquad r > 0, \sigma > 0$$

When X is stationary:

$$K_\theta(0) = \begin{cases} \dfrac{\sigma^2(b\sinh(\lambda(a,b)r) - \lambda(a,b))}{2\lambda(a,b)[a + b\cosh(\lambda(a,b)r)]}, & |b| < -a \\ \sigma^2(br-1)/(4b), & b = a \\ \dfrac{\sigma^2(b\sin(\lambda(a,b)r) - \lambda(a,b))}{2\lambda(a,b)[a + b\cos(\lambda(a,b)r)]}, & b < -|a|, \end{cases} \qquad \lambda(a,b) = \sqrt{|a^2 - b^2|}$$

$$K_\theta(t) = \begin{cases} K_\theta(0)\cosh(\lambda(a,b)t) - \sigma^2(2\lambda(a,b))^{-1}\sinh(\lambda(a,b)t), & |b| < -a \\ K_\theta(0) - t/2, & b = a \\ K_\theta(0)\cos(\lambda(a,b)t) - \sigma^2(2\lambda(a,b))^{-1}\sin(\lambda(a,b)t), & b < -|a|, \end{cases}$$

$t \in [0, r]$

$$K_\theta(t) = b\int_r^t e^{a(t-s)}K_\theta(s-r)ds + e^{a(t-r)}K_\theta(r), \qquad\qquad t \in [r, 2r]$$

Example 2

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

When X is stationary:

$$K_\theta(t) = \frac{\sin\left(\sqrt{2b}(\frac{1}{2} - t)\right)}{2\sqrt{2b}\cos\left(\sqrt{b/2}\right)} + \frac{1}{2b}$$

Optimal prediction-based estimating function

$$G_n^*(\theta) = \partial_\theta \log(\tilde{L}_n(\theta)) + G_n^{\#}(\theta)$$

Both the optimal prediction-based estimator and the pseudo-likelihood estimator are asymptotically normal, and there is an explicit expression for the information loss suffered by using $\partial_\theta \log(\tilde{L}_n(\theta))$ in stead of $G_n^*(\theta)$

A diffusion compartment model

Joint work with Maria Düring: A two-compartment model

$$\begin{pmatrix} dX_t^{(1)} \\ dX_t^{(2)} \end{pmatrix} = \left(\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} -\kappa_1 & \kappa_2 \\ \kappa_1 & -(\kappa_2 + \kappa_3) \end{pmatrix} \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \end{pmatrix} \right) dt$$

$$+ \begin{pmatrix} \sigma_1 \sqrt{X_t^{(1)}} & 0 \\ 0 & \sigma_2 \sqrt{X_t^{(2)}} \end{pmatrix} \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix}$$

$\alpha_1, \alpha_2, \kappa_1, \kappa_2, \kappa_3, \sigma_1, \sigma_2 > 0$

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

Data: $Y_i = X_{i\Delta}^{(1)}$, $i = 1, \ldots, n$.

Diffusion compartment model: Estimation

A reasonable prediction based estimating function:

$$
\begin{aligned}
&f_1(y) = y \qquad && \pi_1^{(i-1)} &&= \alpha_{1,0} + \alpha_{1,1} Y_{i-1} + \cdots + \alpha_{1,s} Y_{i-s} \\
&f_2(y) = y^2 \qquad && \pi_2^{(i-1)} &&= \alpha_{2,0} + \alpha_{2,1} Y_{i-1} + \cdots + \alpha_{2,s} Y_{i-s} \\
&&&&& \quad + \alpha_{2,s+1} Y_{i-1}^2 + \cdots + \alpha_{2,2s} Y_{i-s}^2
\end{aligned}
$$

The predictors can by found explicitly, e.g. by the Durbin-Levinson algorithm

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

Diffusion compartment model: Asymptotics

If $\alpha_i > \frac{1}{2}\sigma_i^2$, $i = 1, 2$, results by Down, Meyn and Tweedie (1995) can be used to show that the two-dimensional process is geometrically α-mixing

The necessary moment and differentiability conditions hold too, so the prediction-based estimator exists for n large enough, and is consistent and asymptotically normal