On Parameter Estimation for Threshold Models

Yu.A. Kutoyants

Laboratoire de Statistique & Processus,

Université du Maine,

FRANCE

Le Mans, SAPS-VII, March 19, 2009

Threshold Autoregressive Time Series

Consider time series

$$X_{j+1} = \rho_1 \, X_j \, \mathbb{1}_{\{|X_j|<\vartheta\}} + \rho_2 \, X_j \, \mathbb{1}_{\{|X_j|\geq\vartheta\}} + \varepsilon_{j+1}, \qquad j = 0, \ldots, n-1,$$

where ε_j are i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$, $\rho_1 \neq \rho_2$ and $|\rho_2| < 1$. We suppose that σ^2, ρ_1, ρ_2 are known and $\vartheta \in \Theta = (\alpha, \beta)$ is unknown parameter. Our goal is to estimate ϑ by observations $X^n = (X_0, X_1, \ldots, X_n)$ and to describe the asymptotic behavior of estimators as $n \to \infty$. Tong (1983), Chan (1993), Hansen (2000), Fan and Yao (2003).

Threshold Ornstein-Uhlenbeck Process

Let the observed process be

$$\mathrm{d}X_t = -\rho_1 X_t \, \mathbb{1}_{\{|X_t|<\vartheta\}}\mathrm{d}t - \rho_2 X_t \, \mathbb{1}_{\{|X_t|\geq\vartheta\}}\mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq T$$

where W_t is Wiener process, $\rho_1 \neq \rho_2$ and $\rho_2 > 0$. We suppose that σ, ρ_1, ρ_2 are known and $\vartheta \in \Theta = (\alpha, \beta)$ is unknown parameter. Our goal is to estimate ϑ by observations $X^T = (X_t, 0 \leq t \leq T)$ and to describe the asymptotic behavior of estimators as $T \to \infty$.

The process $(X_t)_{t\geq 0}$ is an ergodic diffusion with the invariant density

$$f(\vartheta, x) = \frac{1}{G(\vartheta)} \left[e^{-\frac{\rho_1 \, x^2}{\sigma^2}} \mathbb{1}_{\{|x|\leq\vartheta\}} + e^{-\frac{\rho_2 \, x^2}{\sigma^2} + \frac{(\rho_2-\rho_1)\vartheta^2}{\sigma^2}} \mathbb{1}_{\{|x|>\vartheta\}} \right]$$

Let us denote $p_2(x,\vartheta) = G(\vartheta)^{-1} e^{\frac{(\rho_2-\rho_1)\vartheta^2}{\sigma^2}} 1\!\!1_{\{|x|>\vartheta\}}$,
$p_1(x,\vartheta) = G(\vartheta)^{-1} 1\!\!1_{\{|x|\leq\vartheta\}}$. Then this density can be written as a mixture of two Gaussian densities

$$f(\vartheta,x) = p_1(x,\vartheta)\; e^{-\frac{\rho_1\, x^2}{\sigma^2}} + p_2(x,\vartheta)\, e^{-\frac{\rho_2\, x^2}{\sigma^2}}$$

The likelihood ratio function is

$$\begin{aligned}\ln L\left(\vartheta, X^T\right) = &-\frac{\rho_1}{\sigma^2}\int_0^T X_t 1\!\!1_{\{|X_t|\leq\vartheta\}}\,\mathrm{d}X_t - \frac{\rho_2}{\sigma^2}\int_0^T X_t 1\!\!1_{\{|X_t|\geq\vartheta\}}\,\mathrm{d}X_t \\ &-\frac{\rho_1^2}{2\sigma^2}\int_0^T X_t^2 1\!\!1_{\{|X_t|<\vartheta\}}\,\mathrm{d}t - \frac{\rho_2^2}{2\sigma^2}\int_0^T X_t^2 1\!\!1_{\{|X_t|\geq\vartheta\}}\,\mathrm{d}t\end{aligned}$$

and the MLE $\hat{\vartheta}_T$ and BE $\tilde{\vartheta}_T$ are defined as usual by the relations

$$L\left(\hat{\vartheta}_T, X^T\right) = \sup_{\theta\in\Theta} L\left(\theta, X^T\right) \quad \text{and} \quad \tilde{\vartheta}_T = \frac{\int_\alpha^\beta \theta\, p(\theta)\, L\left(\theta, X^T\right)\mathrm{d}\theta}{\int_\alpha^\beta p(\theta)\, L\left(\theta, X^T\right)\mathrm{d}\theta}.$$

The normalized likelihood ratio

$$Z_T(u) = \frac{L\left(\vartheta + \frac{u}{T}, X^T\right)}{L\left(\vartheta, X^T\right)}, \qquad u \in \mathbb{U}_T = \left(T\left(\alpha - \vartheta\right), T\left(\beta - \vartheta\right)\right)$$

converges to the random process

$$Z(u) = \exp\left\{\Gamma_\vartheta \, W(u) - \frac{|u|}{2} \, \Gamma_\vartheta^2\right\}, \qquad u \in R$$

where $W(\cdot)$ is a two-sided Wiener process and

$$\Gamma_\vartheta^2 = \frac{2 \, \left(\rho_2 - \rho_1\right)^2 \, \vartheta^2}{G(\vartheta) \, \sigma^2} e^{-\frac{\rho_1^2 \vartheta^2}{\sigma^2}}.$$

Let us introduce

$$Z_0(u) = \exp\left\{W(u) - \frac{|u|}{2}\right\}, \qquad u \in R$$

.

Introduce two r.v.'s $\hat{u}$ and $\tilde{u}$ are defined by

$$Z_0\left(\hat{u}\right) = \sup_u Z_0\left(u\right), \qquad \tilde{u} = \frac{\int_R u\, Z_0\left(u\right)\ {\rm d}u}{\int_R Z_0\left(u\right)\ {\rm d}u}.$$

The proof is based on the two remarkable Theorems 1.10.1 and 1.10.2 by Ibragimov and Khasminskii (1981). The MLE and BE are

- Uniformly consistent
- Have different limit distributions

$$T\left(\hat{\vartheta}_T - \vartheta\right) \Longrightarrow \frac{\hat{u}}{\Gamma_\vartheta}, \qquad T\left(\tilde{\vartheta}_T - \vartheta\right) \Longrightarrow \frac{\tilde{u}}{\Gamma_\vartheta}$$

- The moments converge: for any $p > 0$

$$\mathbf{E}_\vartheta \left|T\left(\hat{\vartheta}_T - \vartheta\right)\right|^p \longrightarrow \mathbf{E}\left|\frac{\hat{u}}{\Gamma_\vartheta}\right|^p, \quad \mathbf{E}_\vartheta \left|T\left(\tilde{\vartheta}_T - \vartheta\right)\right|^p \longrightarrow \mathbf{E}\left|\frac{\tilde{u}}{\Gamma_\vartheta}\right|^p$$

For the MLE we have (ϑ_0 is the true value) :

$$\mathbf{P}_{\vartheta_0}\left\{T\left(\hat\vartheta_T-\vartheta_0\right)<x\right\}=$$

$$=\mathbf{P}\left\{\sup_{T(\theta-\vartheta_0)<x} L\left(\vartheta,X^T\right)>\sup_{T(\theta-\vartheta_0)\geq x} L\left(\vartheta,X^T\right)\right\}$$

$$=\mathbf{P}\left\{\sup_{T(\theta-\vartheta_0)<x} \frac{L\left(\vartheta,X^T\right)}{L\left(\vartheta_0,X^T\right)}>\sup_{T(\theta-\vartheta_0)\geq x} \frac{L\left(\vartheta,X^T\right)}{L\left(\vartheta_0,X^T\right)}\right\}$$

$$=\mathbf{P}\left\{\sup_{u<x} Z_T\left(u\right)>\sup_{u\geq x} Z_T\left(u\right)\right\}\rightarrow\mathbf{P}\left\{\sup_{u<x} Z\left(u\right)>\sup_{u\geq x} Z\left(u\right)\right\}$$

$$=\mathbf{P}\left(\frac{\hat u}{\Gamma_{\vartheta_0}}<x\right),\quad \text{i.e.}\quad T\left(\hat\vartheta_T-\vartheta_0\right)\Longrightarrow\frac{\hat u}{\Gamma_{\vartheta_0}}.$$

where we put $\vartheta=\vartheta_0+T^{-1}u$.

For the BE we change the variable $\theta = \vartheta_0 + u/T \equiv \vartheta_u$

$$\tilde{\vartheta}_T = \frac{\int_\alpha^\beta \theta p\left(\theta\right) L\left(\theta, X^T\right) \mathrm{d}\theta}{\int_\alpha^\beta p\left(\theta\right) L\left(\theta, X^T\right) \mathrm{d}\theta} = \vartheta_0 + \frac{1}{T} \frac{\int_{\mathbb{U}_T} u p\left(\vartheta_u\right) L\left(\vartheta_u, X^T\right) \mathrm{d}u}{\int_{\mathbb{U}_T} p\left(\vartheta_u\right) L\left(\vartheta_u, X^T\right) \mathrm{d}u},$$

Then using the convergence $p\left(\vartheta_u\right) \longrightarrow p\left(\vartheta_0\right)$ we can write

$$\mathbf{P}_{\vartheta_0}\left\{T\left(\tilde{\vartheta}_T - \vartheta_0\right) < x\right\} = \mathbf{P}\left\{\frac{\int_{\mathbb{U}_T} u\, p\left(\vartheta_u\right) Z_T\left(u\right)\, \mathrm{d}u}{\int_{\mathbb{U}_T} p\left(\vartheta_u\right) Z_T\left(u\right)\, \mathrm{d}u} < x\right\}$$

$$\longrightarrow \mathbf{P}\left\{\frac{\int_R u\, Z\left(u\right)\, \mathrm{d}u}{\int_R Z\left(u\right)\, \mathrm{d}u} < x\right\} = \mathbf{P}\left(\frac{\tilde{u}}{\Gamma_{\vartheta_0}} < x\right)$$

Moreover, the Bayesian estimators are asymptotically efficient in the sense of the following lower bound: for all estimators $\bar{\vartheta}_T$

$$\underline{\lim}_{\delta\to 0}\ \underline{\lim}_{T\to\infty}\ \sup_{|\vartheta-\vartheta_0|<\delta} T^2\, \mathbf{E}_\vartheta \left(\bar{\vartheta}_T-\vartheta\right)^2 \geq \frac{\mathbf{E}\tilde{u}^2}{\Gamma_{\vartheta_0}^2}$$

and for bayesian estimators we have

$$\lim_{\delta\to 0}\ \lim_{T\to\infty}\ \sup_{|\vartheta-\vartheta_0|<\delta} T^2\, \mathbf{E}_\vartheta \left(\tilde{\vartheta}_T-\vartheta\right)^2 = \frac{\mathbf{E}\tilde{u}^2}{\Gamma_{\vartheta_0}^2}.$$

The proofs can be found in K. (2004), Section 3.4.

Remind that the quantities $\mathbf{E}\hat{u}^2$ and $\mathbf{E}\tilde{u}^2$ were calculated by Terent'ev (1968) and Rubin and Song (1995) respectively

$$\mathbf{E}\hat{u}^2 = 26 \quad > \quad \mathbf{E}\tilde{u}^2 = 16\zeta\left(3\right) \sim 19,3$$

where $\zeta\left(\cdot\right)$ is Riemann zeta function.

Misspecification

Suppose that the observed process is

$$\mathrm{d}X_t = -\rho_1 X_t \, 1\!\!\mathrm{I}_{\{|X_t|<\vartheta\}}\mathrm{d}t - \rho_2 X_t \, 1\!\!\mathrm{I}_{\{|X_t|\geq\vartheta\}}\mathrm{d}t + h\left(X_t\right)\mathrm{d}t + \sigma \mathrm{d}W_t,$$

where $h\left(\cdot\right)$ is unknown function (contamination) and the statistician to construct the estimators uses this model without $h\left(\cdot\right)$ (wrong model). Remind that in regular case the MLE and BE are usually not consistent. In our (singular) case the consistent estimation is possible. For example, suppose that $\rho_1 > \rho_2$ and that the invariant density is symmetric function. If the function $h\left(\cdot\right)$ satisfies the condition

$$\vartheta\left(\rho_2 - \rho_1\right) < h\left(\vartheta\right) - h\left(-\vartheta\right) < \vartheta\left(\rho_1 - \rho_2\right),$$

then the MLE is consistent.

It is possible to describe the properties of estimators in the case when all three parameters $(\rho_1, \rho_2, \vartheta) = (\vartheta_1, \vartheta_2, \vartheta_3) = \vartheta$ are unknown and we observe

$$\mathrm{d}X_t = -\vartheta_1 X_t \,\mathbb{1}_{\{|X_t|<\vartheta_3\}}\mathrm{d}t - \vartheta_2 X_t \,\mathbb{1}_{\{|X_t|\geq\vartheta_3\}}\mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq T.$$

Under condition of identifiability,$(\vartheta_1 \neq \vartheta_2)$ the both estimators (MLE $\hat{\vartheta}_T$, BE $\tilde{\vartheta}_T$) are consistent and $\sqrt{T}\left(\hat{\vartheta}_{1,T} - \vartheta_1\right)$, $\sqrt{T}\left(\hat{\vartheta}_{2,T} - \vartheta_2\right)$ and $T\left(\hat{\vartheta}_{3,T} - \vartheta_3\right)$ are asymptotically independent

$$\sqrt{T}\left(\hat{\vartheta}_{1,T} - \vartheta_1\right) \Longrightarrow \mathcal{N}\left(0, \frac{\sigma^2}{\mathbf{E}_\vartheta \xi^2 \mathbb{1}_{\{|\xi|<\vartheta_3\}}}\right),$$

$$\sqrt{T}\left(\hat{\vartheta}_{2,T} - \vartheta_2\right) \Longrightarrow \mathcal{N}\left(0, \frac{\sigma^2}{\mathbf{E}_\vartheta \xi^2 \mathbb{1}_{\{|\xi|\geq\vartheta_3\}}}\right),$$

$$T\left(\hat{\vartheta}_{3,T} - \vartheta_3\right) \Longrightarrow \frac{\hat{u}}{\Gamma_\vartheta}, \qquad T\left(\tilde{\vartheta}_{3,T} - \vartheta_3\right) \Longrightarrow \frac{\tilde{u}}{\Gamma_\vartheta}$$

Other Threshold Models.

Simple Threshold model. Suppose that the observed process is

$$\mathrm{d}X_t = \rho_1 \, 1\!\!\mathrm{I}_{\{X_t<\vartheta\}}\mathrm{d}t - \rho_2 \, 1\!\!\mathrm{I}_{\{X_t\geq\vartheta\}}\mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq T$$

where $\rho_i > 0$ and $\vartheta \in (\alpha, \beta)$. Then this process is ergodic with exponential type invariant density. The normalized LR

$$Z_T(u) = \frac{L\left(\vartheta + \frac{u}{T}, X^T\right)}{L\left(\vartheta, X^T\right)} \Longrightarrow \exp\left\{\Gamma_\vartheta \, W(u) - \frac{|u|}{2}\,\Gamma_\vartheta^2\right\}$$

where $W(\cdot)$ is a two-sided Wiener process

$$\Gamma_\vartheta^2 = \frac{2\,\rho_2\rho_1\left(\rho_2+\rho_1\right)}{\sigma^4}$$

and we have the corresponding asymptotic properties of the MLE and BE.

Simple Switching. Suppose that in this model $\rho_1 = \rho_2 = \rho > 0$. Then the observed process is

$$\mathrm{d}X_t = -\rho \,\mathrm{sgn}\,(X_t - \vartheta)\,\mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq T$$

where $\rho_i > 0$ and $\vartheta \in (\alpha, \beta)$. The invariant density is $f(\vartheta, x) = e^{-2|x-\vartheta|}$ and the normalized LR

$$Z_T(u) = \frac{L\left(\vartheta + \frac{u}{T}, X^T\right)}{L\left(\vartheta, X^T\right)} \Longrightarrow \exp\left\{\Gamma_\vartheta\, W(u) - \frac{|u|}{2}\,\Gamma_\vartheta^2\right\}, \qquad \Gamma_\vartheta^2 = \frac{4\,\rho^3}{\sigma^4}$$

and we have the corresponding asymptotic properties of the MLE and BE: $T\left(\hat\vartheta_T - \vartheta\right) \Rightarrow \hat u/\Gamma_\vartheta$.

Note that the same normalization and the same limit process we have for the model (Küchler, K.)

$$\mathrm{d}X_t = -\rho\, X_{t-\vartheta}\,\mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq T$$

The observation window $(-\infty,\infty)$ can be essentially reduced. Let us put

$$\vartheta^\star_{\sqrt{T}}=\frac{1}{\sqrt{T}}\int_0^{\sqrt{T}} X_t\,\mathrm{d}t\rightarrow\vartheta,\qquad T^{1/4}\left(\vartheta^\star_{\sqrt{T}}-\theta\right)\Longrightarrow\mathcal{N}\left(0,d^2\left(\vartheta\right)\right)$$

and introduce the *window*

$$\mathbb{B}_T=\left[\vartheta^\star_{\sqrt{T}}-T^{1/8},\vartheta^\star_{\sqrt{T}}+T^{1/8}\right].$$

The MLE and BE we define with the help of the LR $L\left(\vartheta,X^T_{\sqrt{T}}\right)$

$$=\exp\left\{-\frac{\rho}{\sigma^2}\int_{\sqrt{T}}^{T}\mathrm{sgn}\left(X_t-\vartheta\right)1\!\!\mathrm{I}_{\{X_t\in\mathbb{B}_T\}}\mathrm{d}X_t-\frac{\rho^2}{2\sigma^2}\int_{\sqrt{T}}^{T}1\!\!\mathrm{I}_{\{X_t\in\mathbb{B}_T\}}\mathrm{d}t\right\}$$

Then these estimators have the same as. properties as if $\mathbb{B}_T=(-\infty,\infty)$.

Multy Threshold O-U Process. Suppose that the observed process is

$$\mathrm{d}X_t = -\sum_{l=1}^{k+1} \rho_l \, X_t \; 1\!\!1_{\{\vartheta_{l-1} < X_t \leq \vartheta_l\}} \mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq T$$

where $\rho_1 > 0$, $\rho_{k+1} > 0$, $\rho_l \neq \rho_m > 0$, $\vartheta_0 = -\infty$, $\vartheta_{k+1} = \infty$ and $\vartheta = (\vartheta_1, \ldots, \vartheta_k)$. Then this process is ergodic. The normalized LR

$$Z_T(\mathbf{u}) = \frac{L\left(\vartheta + \frac{\mathbf{u}}{T}, X^T\right)}{L\left(\vartheta, X^T\right)} \Longrightarrow \prod_{l=1}^{k} \exp\left\{\Gamma_l \; W_l\left(u_l\right) - \frac{|u_l|}{2} \, \Gamma_l^2\right\}$$

where $W_l(\cdot)$ are independent two-sided Wiener processes. The estimators $\hat{\vartheta}_T$ and $\tilde{\vartheta}_T$ have asymptotically independent components, say,

$$T\left(\hat{\vartheta}_{l,T} - \vartheta_l\right) \Longrightarrow \frac{\hat{u}_l}{\Gamma_l}.$$

Nonlinear Threshold Model. Suppose that

$$\mathrm{d}X_t = S_1\left(X_t\right)\, 1\!\!1_{\left\{|X_t|<\vartheta\right\}}\mathrm{d}t + S_2\left(X_t\right)\, 1\!\!1_{\left\{|X_t|\geq\vartheta\right\}}\mathrm{d}t + \sigma\left(X_t\right)\mathrm{d}W_t,$$

where $S_2(x)$ and $\sigma\left(x\right)$ are such that the process is ergodic. Then, if

$$|S_1(\vartheta) - S_2(\vartheta)| + |S_1(-\vartheta) - S_2(-\vartheta)| > 0, \qquad \vartheta \in \left(\alpha, \beta\right),$$

then the LR

$$Z_T\left(u\right) = \frac{L\left(\vartheta + \frac{u}{T}, X^T\right)}{L\left(\vartheta, X^T\right)} \Longrightarrow \exp\left\{\Gamma_\vartheta\, W\left(u\right) - \frac{|u|}{2}\,\Gamma_\vartheta^2\right\}$$

where $W\left(\cdot\right)$ is a two-sided Wiener process

$$\Gamma_\vartheta^2 = \frac{\left(S_2\left(\vartheta\right) - S_1\left(\vartheta\right)\right)^2}{\sigma\left(\vartheta\right)^2}\, f\left(\vartheta, \vartheta\right) + \frac{\left(S_2\left(-\vartheta\right) - S_1\left(-\vartheta\right)\right)^2}{\sigma\left(-\vartheta\right)^2}\, f\left(\vartheta, -\vartheta\right)$$

and we have the corresponding asymptotic properties of the MLE and BE.

Threshold Autoregressive Time Series

(joint work with N.H.Chan)

Consider the discrete time model

$$X_{j+1} = \rho_1\, X_j\, 1\!\!1_{\{|X_j|<\vartheta\}} + \rho_2\, X_j\, 1\!\!1_{\{|X_j|\geq\vartheta\}} + \varepsilon_{j+1}, \qquad j = 0,\dots,n-1,$$

where ε_j are i.i.d. $\mathcal{N}\left(0,\sigma^2\right)$, $\rho_1 \neq \rho_2$ and $|\rho_2| < 1$. The unknown parameter $\vartheta \in \Theta = (\alpha,\beta)$, $\alpha > 0$. The likelihood function is written as

$$L\left(\vartheta, X^n\right) = \frac{f_0\left(X_0\right)}{\left(2\pi\sigma^2\right)^{n/2}} \exp\Big\{-\frac{1}{2\sigma^2}\sum_{j=0}^{n-1}\Big(X_{j+1} - \rho_1\, X_j\, 1\!\!1_{\{|X_j|\leq\vartheta\}} - \rho_2\, X_j\, 1\!\!1_{\{|X_j|>\vartheta\}}\Big)^2\Big\}.$$

Note that $(X_j)_{j\geq 1}$ is geometrically mixing and denote its stationary density function by $f(\vartheta, x)$. It is solution of the integral equation

$$
\begin{aligned}
f(\vartheta, y) = \int_{-\infty}^{\infty} \Bigg[\frac{1}{\rho_1} f\left(\vartheta, \frac{x}{\rho_1}\right) \; 1\!\!1_{\left\{ \frac{|x|}{\rho_1} < \vartheta \right\}} \\
+ \frac{1}{\rho_2} f\left(\vartheta, \frac{x}{\rho_2}\right) \; 1\!\!1_{\left\{ \frac{|x|}{\rho_2} \geq \vartheta \right\}} \Bigg] \varphi(y - x) \, \mathrm{d}x \\
= \int_{-\infty}^{\infty} f(\vartheta, x) \left[\varphi(y - x\rho_1) \; 1\!\!1_{\{|x| < \vartheta\}} + \varphi(y - x\rho_2) \; 1\!\!1_{\{|x| \geq \vartheta\}} \right] \mathrm{d}x.
\end{aligned}
$$

The maximum likelihood estimator (MLE) $\hat{\vartheta}_n$ is defined by the equation

$$\sup_{\vartheta\in\Theta} L\left(\vartheta, X^n\right) = \max\left[L\left(\hat{\vartheta}_n+, X^n\right), L\left(\hat{\vartheta}_n-, X^n\right)\right].$$

$L\left(\vartheta, X^n\right)$ is piece-wise constant function and this equation has many (random interval) solutions. We show that the middle point of this interval minimizes the limit variance of the MLE.

The MLE coincides with the Least Squares Estimator $\vartheta_n^\star$ defined by the equation

$$\vartheta_n^\star = \arg\inf_{\vartheta\in\Theta} \sum_{j=0}^{n-1} \left(X_{j+1} - \rho_1\, X_j\, 1\!\!1_{\{|X_j|\leq\vartheta\}} - \rho_2\, X_j\, 1\!\!1_{\{|X_j|>\vartheta\}}\right)^2 \Big\}$$

and studied by Chan (1993), who showed the consistency and the limit distribution of $n\left(\vartheta_n^\star - \vartheta\right)$.

To apply the Bayesian approach, suppose that the unknown parameter is a random variable with known prior density $p(\theta), \theta \in \Theta$, which is continuous and positive. Using the quadratic loss function, the Bayesian estimator (which minimizes the mean square error) is the conditional mathematical expectation

$$\tilde{\vartheta}_n = \int_\alpha^\beta \theta\, p(\theta)\, L(\vartheta, X^n)\ \mathrm{d}\theta = \frac{\int_\alpha^\beta \theta\, p(\theta)\, L(\vartheta, X^n)\ \mathrm{d}\theta}{\int_\alpha^\beta p(\theta)\, L(\vartheta, X^n)\ \mathrm{d}\theta}.$$

Note that the function $L(\vartheta, X^n), \vartheta \in \Theta$ has jumps at the points

$$\vartheta_l = |X_j| \in \Theta, \qquad l = 1, \ldots, L,$$

where $L \leq n$. Clearly, if $\Theta = R$, then $L = n$. We can restrict observation window by interval $[\alpha, \beta]$ without loss of limit variance.

Introduce the stochastic process

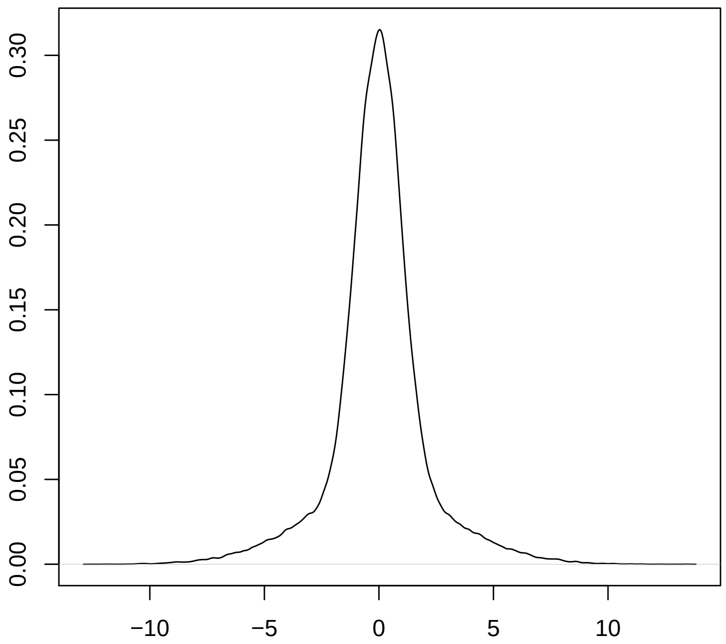
$$Z(u) = \begin{cases} \exp\left\{-\frac{\rho^2\vartheta^2}{2\sigma^2}\; N_+(u) - \frac{\rho\,\vartheta}{\sigma^2}\;\sum_{l=0}^{N_+(u)} \varepsilon_l^+\right\}, & u \geq 0, \\ \exp\left\{-\frac{\rho^2\vartheta^2}{2\sigma^2}\; N_-(-u) - \frac{\rho\,\vartheta}{\sigma^2}\;\sum_{l=0}^{N_-(-u)} \varepsilon_l^-\right\}, & u \leq 0, \end{cases}$$

where $N_+(\cdot)$ and $N_-(\cdot)$ are two independent Poisson processes of intensities $\lambda_+ = \lambda_- = 2f(\vartheta, \vartheta)$, $\rho = \rho_2 - \rho_1$ and $\varepsilon_r^+, \varepsilon_l^-$ are independent Gaussian $\mathcal{N}\left(0, \sigma^2\right)$ random variables.

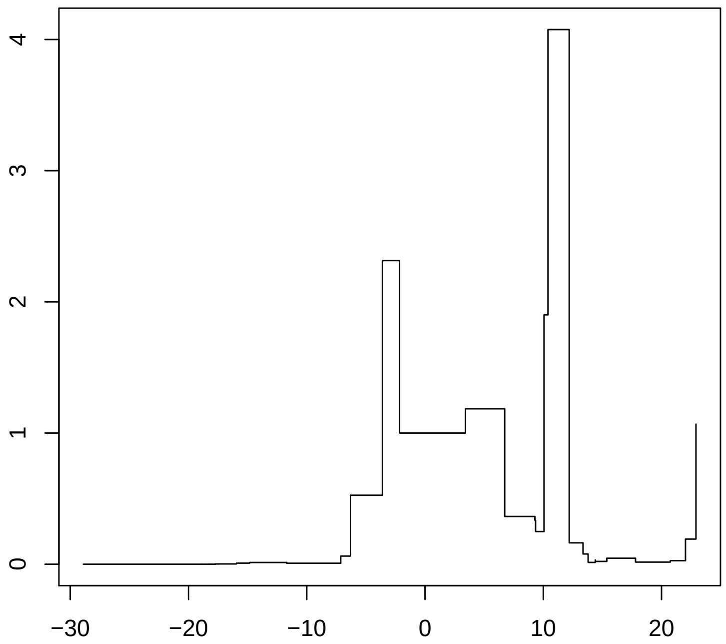
Note that

$$Y_+(u) = \sum_{l=0}^{N_+(u)} \left[\rho^2\,\vartheta^2 + 2\rho\,\vartheta\,\varepsilon_l^+\right], \qquad u \geq 0$$

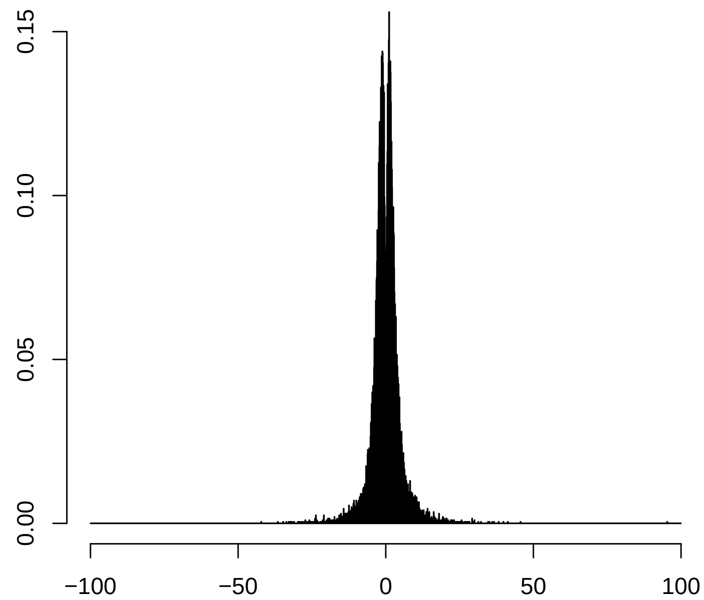
is compound Poisson process.

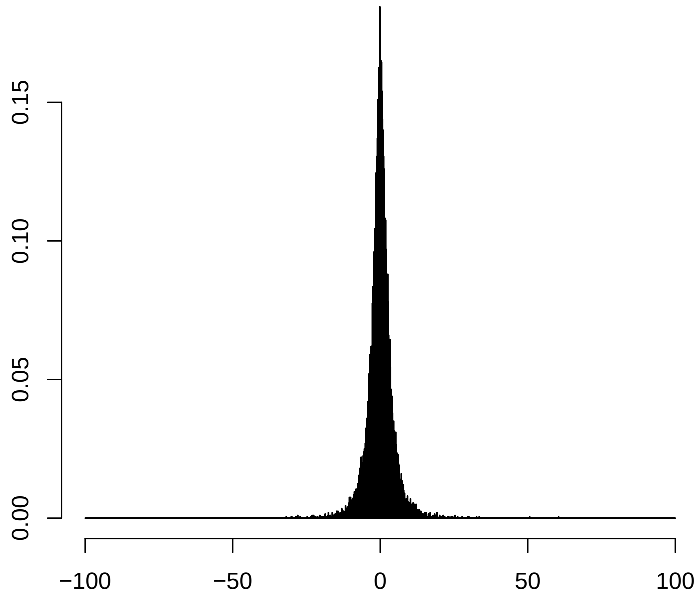


0.00
0.05
0.10
0.15
0.20
0.25
0.30
−10
−5
0
5
10



0
1
2
3
4
−30
−20
−10
0
10
20


0.00
0.05
0.10
0.15
−100
−50
0
50
100


0.00
0.05
0.10
0.15
−100
−50
0
50
100

The random process $Z(\cdot)$ is piecewise constant and as a result, the points u^* of the maximum of the process $Z(\cdot)$ is defined by

$$\sup_u Z(u) = Z(u^*) \ ,$$

where $\hat{u}_m < u^* < \hat{u}_M$. Here $\hat{u}_m$ and $\hat{u}_M$ are two adjacent events of the process $N_+(\cdot)$, or of the process $N_-(\cdot)$, or they are respectively the first event of $N_+(\cdot)$ and $N_-(\cdot)$.

The center of gravity of the interval is given by the point

$$\hat{u} = \frac{u_m + u_M}{2}.$$

It follows from K.S. Chan (1993) that

$$n\left(\hat{\vartheta}_n - \vartheta\right) \Longrightarrow \hat{u}.$$

Introduce the random variable

$$\tilde{u} = \frac{\int_R u\, Z\left(u\right)\, \mathrm{d}u}{\int_R Z\left(u\right)\, \mathrm{d}u}.$$

Remind that the Bayesian estimators are asymptotically efficient in this case too. For all estimators $\bar{\vartheta}_n$

$$\varliminf_{\delta\to 0}\varliminf_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta} n^2\,\mathbf{E}_\vartheta\left(\bar{\vartheta}_n-\vartheta\right)^2 \geq \mathbf{E}_{\vartheta_0}\tilde{u}^2$$

and for Bayesian estimators we have

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta} n^2\,\mathbf{E}_\vartheta\left(\tilde{\vartheta}_n-\vartheta\right)^2 = \mathbf{E}_{\vartheta_0}\tilde{u}^2.$$

It will be interesting to calculate $\mathbf{E}_\vartheta\hat{u}^2$ and $\mathbf{E}_\vartheta\tilde{u}^2$ to compare. Surely $\mathbf{E}_\vartheta\hat{u}^2 > \mathbf{E}_\vartheta\tilde{u}^2$. Simulation?

We have the following theorem.

Theorem 1 (N.H. Chan and Y.K.) *The Bayesian estimator* $\tilde{\vartheta}_n$ *constructed by the observations* X^n *of the threshold autoregressive process is consistent,*

$$n\left(\tilde{\vartheta}_n - \vartheta\right) \Longrightarrow \tilde{u}$$

and for any $p > 0$

$$\lim_{n\to\infty} \mathbf{E}_\vartheta \left|n\left(\tilde{\vartheta}_n - \vartheta\right)\right|^p = \mathbf{E}_\vartheta \left|\tilde{u}\right|^p .$$

The proof of this theorem is based on the general result by Ibragimov and Khasminskii (1981), Theorem 1.10.2. To apply it we study the normalized likelihood ratio process

$$Z_n\left(u\right) = \frac{L\left(\vartheta + \frac{u}{n}, X^n\right)}{L\left(\vartheta, X^n\right)}, \qquad u \in \mathbb{U}_n = \left[n\left(\alpha - \vartheta\right), n\left(\beta - \vartheta\right)\right],$$

where ϑ is the true value.

the main steps. Write the Bayesian estimator $\left(\theta_u = \vartheta + \frac{u}{n}\right)$ as

$$\tilde{\vartheta}_n = \frac{\int_\alpha^\beta \theta \, p(\theta) \, L(\theta, X^n) \, \mathrm{d}\theta}{\int_\alpha^\beta p(\theta) \, L(\theta, X^n) \, \mathrm{d}\theta} = \vartheta + \frac{1}{n} \frac{\int_{\mathbb{U}_n} u \, p(\theta_u) \, L(\theta_u, X^n) \, \mathrm{d}u}{\int_{\mathbb{U}_n} p(\theta_u) \, L(\theta_u, X^n) \, \mathrm{d}u}$$

$$= \vartheta + \frac{1}{n} \frac{\int_{\mathbb{U}_n} u \, p(\theta_u) \, \frac{L(\theta_u, X^n)}{L(\vartheta, X^n)} \, \mathrm{d}u}{\int_{\mathbb{U}_n} p(\theta_u) \, \frac{L(\theta_u, X^n)}{L(\vartheta, X^n)} \, \mathrm{d}u} = \vartheta + \frac{1}{n} \frac{\int_{\mathbb{U}_n} u \, p(\theta_u) \, Z_n(u) \, \mathrm{d}u}{\int_{\mathbb{U}_n} p(\theta_u) \, Z_n(u) \, \mathrm{d}u} .$$

Then

$$n\left(\tilde{\vartheta}_n - \vartheta\right) = \frac{\int_{\mathbb{U}_n} u \, p(\theta_u) \, Z_n(u) \, \mathrm{d}u}{\int_{\mathbb{U}_n} p(\theta_u) \, Z_n(u) \, \mathrm{d}u} \Longrightarrow \frac{\int_R u \, Z(u) \, \mathrm{d}u}{\int_R Z(u) \, \mathrm{d}u}$$

We have to prove

- The convergence of the finite dimensional distributions of $Z_n(\cdot)$ to the finite dimensional distributions of $Z(\cdot)$
- To establish the following two estimates

$$\mathbf{E}_\vartheta \left[Z_n^{1/2}(u_2) - Z_n^{1/2}(u_1) \right]^2 \leq C \left| u_2 - u_1 \right| ,$$

and for any $M > 0$

$$\mathbf{E}_\vartheta Z_n^{1/2}(u) \leq \frac{C_M}{\left| u \right|^M} .$$

Put $Z_n(u) = \exp\left\{-\frac{1}{2\sigma^2} Y_n(u)\right\}$ and study the process $Y_n(u)$ for $u \geq 0$

$$Y_n(u) = \sum_{j=0}^{n-1} \left[\rho^2\, X_j^2 + 2\rho\, X_j\, \varepsilon_{j+1}\right]\; 1\!\!1_{\left\{\vartheta < |X_j| < \vartheta + \frac{u}{n}\right\}}.$$

This process can be replaced by

$$Y_n^\circ(u) = \sum_{j=0}^{n-1} \left[\rho^2\, \vartheta^2 + 2\rho\, \vartheta\; \mathrm{sgn}\left(X_j\right)\; \varepsilon_{j+1}\right]\; 1\!\!1_{\left\{\vartheta < |X_j| < \vartheta + \frac{u}{n}\right\}}$$

and we have to verify the convergence

$$Y_n^\circ(u) \Longrightarrow Y_+(u) = \sum_{l=0}^{N_+(u)} \left[\rho^2\, \vartheta^2 + 2\rho\, \vartheta\, \varepsilon_l^+\right]$$

First note that the characteristic function

$$\Phi\left(v\right)=\mathbf{E}_{\vartheta}e^{\mathrm{i}vY_{+}\left(u\right)}=\exp\left\{u\left(e^{\mathrm{i}v\rho^{2}\,\vartheta^{2}-2v^{2}\rho^{2}\,\vartheta^{2}\,\sigma^{2}}-1\right)2f\left(\vartheta,\vartheta\right)\right\},$$

Fix $u>0$, then as $n\rightarrow\infty$ the band $\left[\vartheta,\vartheta+\frac{u}{n}\right]$ becomes more and more narrow and the events, when $|X_{j_l}|\in\left[\vartheta,\vartheta+\frac{u}{n}\right]$ become more rare. This means that the distance between two adjacent events $|X_{j_l}|\in\left[\vartheta,\vartheta+\frac{u}{n}\right]$ and $\left|X_{j_{l+1}}\right|\in\left[\vartheta,\vartheta+\frac{u}{n}\right]$ tends to infinity. As the process $\left(X_j\right)_{j\geq1}$ is geometrically mixing, these events become asymptotically independent.

Under such circumstances, the characteristic function

$$\Phi_n(v) = \mathbf{E}_\vartheta e^{\mathrm{i}vY_n^\circ(u)} \sim \left(\mathbf{E}_\vartheta \exp\left\{\left(\mathrm{i}v\rho^2\vartheta^2 - 2v^2\rho^2\,\vartheta^2\,\sigma^2\right)\ \mathbb{I}_{\{\mathbb{B}_1(u)\}}\right\}\right)^n .$$

Further,

$$\begin{aligned}\mathbf{E}_\vartheta \exp&\left\{\left(\mathrm{i}v\rho^2\vartheta^2 - 2v^2\rho^2\,\vartheta^2\,\sigma^2\right)\ \mathbb{I}_{\{\mathbb{B}_1(u)\}}\right\}\\ &= 1 - 2\frac{u}{n}f(\vartheta,\vartheta) + 2\frac{u}{n}e^{\mathrm{i}v\rho^2\vartheta^2 - 2v^2\rho^2\,\vartheta^2\,\sigma^2} f(\vartheta,\vartheta) + o\left(\frac{u}{n}\right).\end{aligned}$$

Hence

$$\begin{aligned}\ln\Phi_n(v) &= n\ln\left(1 + \frac{u}{n}\left(e^{\mathrm{i}v\rho^2\vartheta^2 - 2v^2\rho^2\,\vartheta^2\,\sigma^2} - 1\right)2f(\vartheta,\vartheta) + o\left(\frac{u}{n}\right)\right)\\ &\longrightarrow u\left(e^{\mathrm{i}v\rho^2\vartheta^2 - 2v^2\rho^2\,\vartheta^2\,\sigma^2} - 1\right)2f(\vartheta,\vartheta) = \ln\Phi(v)\ .\end{aligned}$$

Goodness of Fit Testing

Classical GoF Tests

If we observe n i.i.d. r. v.'s $(X_1, \ldots, X_n) = X^n$ with distribution function $F(x)$ and the basic hypothesis is simple

$$\mathscr{H}_0, \qquad F(x) \equiv F_*(x), \qquad x \in \mathbb{R},$$

then the Cramér-von Mises and Kolmogorov-Smirnov statistics are

$$W_n^2 = n \int \left[\hat{F}_n(x) - F_*(x)\right]^2 \mathrm{d}F_*(x), \quad D_n = \sup_x \sqrt{n} \left|\hat{F}_n(x) - F_*(x)\right|$$

respectively. Here

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{X_j < x\}}$$

is the empirical distribution function.

We have the convergence

$$W_n^2 \Longrightarrow \int_0^1 W_0(s)^2 \,\mathrm{d}s, \quad D_n \Longrightarrow \sup_{0\leq s\leq 1} |W_0(s)|,$$

where $W_0(\cdot)$ is Brownian bridge.

Let us denote by $\mathcal{K}_\alpha$ the class of tests of asymptotic size α, i.e.;

$$\mathcal{K}_\alpha = \left\{\bar\psi \quad : \quad \mathbf{E}_0\, \bar\psi = \alpha + o(1)\right\}.$$

If we take the constants c_α, d_α from the equations

$$\mathbf{P}\left\{\|W_0(\cdot)\|^2 > c_\alpha\right\} = \alpha, \quad \mathbf{P}\left\{\|W_0(\cdot)\|_\infty > d_\alpha\right\} = \alpha$$

Then the *Cramér-von Mises* and *Kolmogorov-Smirnov* tests

$$\psi_n(X^n) = 1_{\{W_n^2 > c_\alpha\}} \in \mathcal{K}_\alpha, \qquad \phi_n(X^n) = 1_{\{D_n > d_\alpha\}} \in \mathcal{K}_\alpha.$$

These tests are *distribution-free* .

These tests are uniformly consistent against any alternative

$$\mathscr{H}_\rho = \left\{ F(\cdot) : \|F(\cdot) - F_*(\cdot)\| \geq \rho \right\}, \qquad \rho > 0$$

If we define the (absolutely continuous) alternative as

$$\hat{\mathscr{H}}_\rho : \qquad F(x) = F_*(x) + \frac{1}{\sqrt{n}} \int_{-\infty}^{x} u(F_*(y))\,\mathrm{d}F_*(y), \quad \|u(\cdot)\| \geq \rho$$

then for any fixed $u(\cdot)$ the limit of the test C-vM statistic is

$$W_n^2 \Longrightarrow \int_0^1 \left[W_0(s) + \int_0^s u(v)\,\mathrm{d}v \right]^2 \mathrm{d}s,$$

but this alternative became *invisible* in minimax sense. Indeed it contains the functions $u_k(x) = \sqrt{2\rho}\cos(2\pi kx)$, $\quad x \in [0,1]$ and the power function of any test ϕ_n of size α satisfies the equality

$$\inf_{f \in \hat{\mathscr{H}}_\rho} \beta(\phi_n, f) \leq \alpha.$$

Threshold O-U process

Simple hypothesis. Suppose that the basic hypothesis is simple:

$$\mathscr{H}_0 \qquad : \qquad \textit{the observed process } X^T \textit{ is } TOU\left(\vartheta_0\right)$$

i.e., the observations $X^T = (X_t, 0 \leq t \leq T)$ come from the equation

$$\mathrm{d}X_t = -\rho_1 X_t \, 1\!\!1_{\{|X_t| < \vartheta_0\}} \mathrm{d}t - \rho_2 X_t \, 1\!\!1_{\{|X_t| \geq \vartheta_0\}} \mathrm{d}t + \sigma \mathrm{d}W_t, \;\; 0 \leq t \leq T$$

with known ϑ_0 and we have to test this hypothesis. Our goal is to find tests of C-vM and K-S types of asymptotic size α.

Let us denote $g(x,\vartheta) = -\rho_1 x\, 1\!\!\mathrm{I}_{\{|x|<\vartheta_0\}} - \rho_2 x\, 1\!\!\mathrm{I}_{\{|x|\geq\vartheta_0\}}$ and introduce the statistics

$$W_T^2 = \frac{1}{\sigma^2 T^2}\int_0^T \left[X_t - X_0 - \int_0^t g(X_s,\vartheta_0)\,\mathrm{d}s\right]^2 \mathrm{d}t,$$

and

$$D_T = \frac{1}{\sigma\sqrt{T}} \sup_{0\leq t\leq T}\left|X_t - X_0 - \int_0^t g(X_s,\vartheta_0)\,\mathrm{d}s\right|$$

It is easy to see that under $\mathscr{H}_0$ (= in distribution)

$$W_T^2 = \int_0^1 W(s)^2\,\mathrm{d}s, \qquad D_T = \sup_{0\leq s\leq 1}|W(s)|,$$

where $W(\cdot)$ is Wiener process. Hence the tests

$$\psi_T\left(X^T\right) = 1\!\!\mathrm{I}_{\left\{W_T^2>c_\alpha\right\}}, \qquad \phi_T\left(X^T\right) = 1\!\!\mathrm{I}_{\{D_T>d_\alpha\}}$$

are distribution free. The both tests are consistent.

Composite hypothesis. Suppose that the basic hypothesis is composite:

$$\mathscr{H}_0 \quad : \quad \textit{the observed process } X^T \textit{ is } TOU\left(\vartheta\right), \vartheta \in \Theta$$

Let us introduce the statistics

$$W_T^2 = \frac{1}{\sigma^2 T^2} \int_0^T \left[X_t - X_0 - \int_0^t g\left(X_s, \tilde{\vartheta}_T\right) \mathrm{d}s\right]^2 \mathrm{d}t,$$

and

$$D_T = \frac{1}{\sigma\sqrt{T}} \sup_{0\leq t\leq T} \left|X_t - X_0 - \int_0^t g\left(X_s, \tilde{\vartheta}_T\right) \mathrm{d}s\right|$$

where $\tilde{\vartheta}_T$ is bayesian estimator. Remind, that $\tilde{\vartheta}_T = \vartheta + \frac{\tilde{u}_T}{T}$.

It can be shown that under $\mathscr{H}_0$

$$W_T^2 \Longrightarrow \int_0^1 W\left(s\right)^2 \mathrm{d}s, \qquad\qquad D_T \Longrightarrow \sup_{0\leq s\leq 1} \left|W\left(s\right)\right|.$$

Hence the tests

$$\psi_T\left(X^T\right) = \mathbb{1}_{\left\{W_T^2 > c_\alpha\right\}}, \qquad\qquad \phi_T\left(X^T\right) = \mathbb{1}_{\left\{D_T > d_\alpha\right\}}$$

are asymptotically distribution free. The tests are consistent against any fixed alternative.