On Identification of Threshold Models for Time Series and Diffusion Processes

Yu.A. Kutoyants

Laboratoire de Statistique & Processus,

Université du Maine, FRANCE

Journée ISFA-LM, Mai 3, 2012

Threshold Autoregressive Time Series

Usual autoregressive (AR) time series

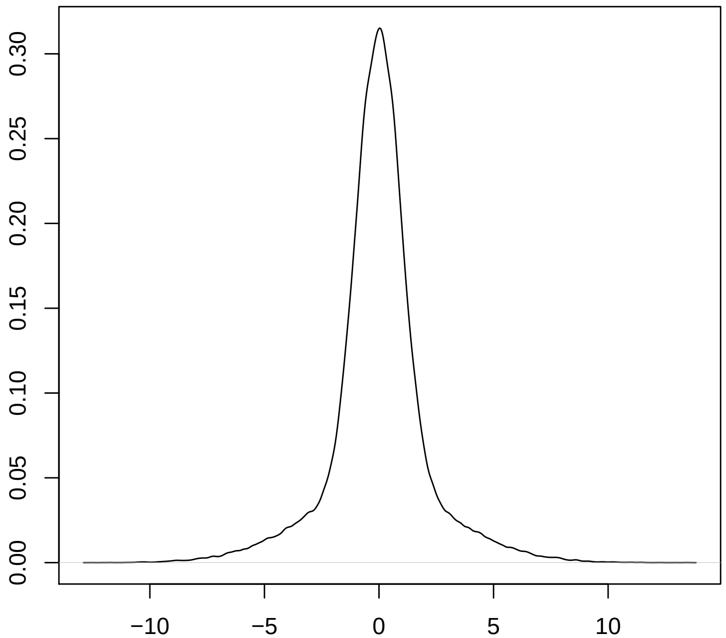
$$X_{j+1} = \rho\, X_j \; + \varepsilon_{j+1}, \qquad j = 0, \ldots, n-1,$$

where ε_j are i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$ and $|\rho| < 1$. We are interested by the time series of the different type

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

where $\rho_1 \neq \rho_2$ and $|\rho_i| < 1$. 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).



Invariant density.

Threshold Ornstein-Uhlenbeck Process

Let the observed process be

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

where W_t is Wiener process, $\rho_1 \neq \rho_2$ and $\rho_i > 0$. 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 ergodic diffusion with the invariant density

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

1 Threshold Models

Consider the following threshold time series model

$$X_{j+1} = h\left(X_j\right)\, \mathbb{I}_{\{X_j<\vartheta\}} + g\left(X_j\right)\, \mathbb{I}_{\{X_j\geq\vartheta\}} + \varepsilon_{j+1}, \qquad j = 0, \dots, n-1,$$

where $\vartheta \in \Theta = (\alpha, \beta)$, ε_j are i.i.d. r.v.'s with smooth density function $f(\cdot)$, $h(x) \neq g(x)\,, x \in \Theta$ and X_0 is an initial value which is indepedent of ε_j and ϑ. Suppose that the functions $h(\cdot)\,, g(\cdot)$ and $f(\cdot)$ are known and the parameter ϑ is unknown. Moreover, suppose that the functions $h(\cdot)\,, g(\cdot)$ and $f(\cdot)$ are such that the time series $\left(X_j\right)_{j\geq 1}$ is exponentially mixing and therefore has ergodic properties with invariant density $\varphi(\vartheta, x)$. Our main problem is to estimate ϑ from the data $X^n = (X_0, X_1, \dots, X_n)$ and examine the asymptotic properties of the estimators as $n \to \infty$. This model is commonly known as *threshold autoregression* (TAR) time series.

Recall that the likelihood function is

$$L(\vartheta, X^n) = f_0(X_0) \prod_{j=0}^{n-1} f\left(X_{j+1} - h(X_j)\, 1\!\!1_{\{X_j<\vartheta\}} - g(X_j)\, 1\!\!1_{\{X_j\geq\vartheta\}}\right).$$

Here $f_0(\cdot)$ is the density function of the initial value X_0.

Using this likelihood ratio, define the maximum likelihood estimator (MLE) $\hat{\vartheta}_n$ by the equation

$$\sup_{\vartheta\in\Theta} L(\vartheta, X^n) = L\left(\hat{\vartheta}_n, X^n\right). \tag{1}$$

As usually done, if this equation has more than one solution, (our case) then the MLE is taken to be any one.

On the other hand, suppose that the unknown parameter ϑ is a random variable with density function $p(\vartheta), \vartheta \in \Theta$. Then the Bayesian estimator (BE), which corresponds to the quadratic loss function, is calculated via the following formula

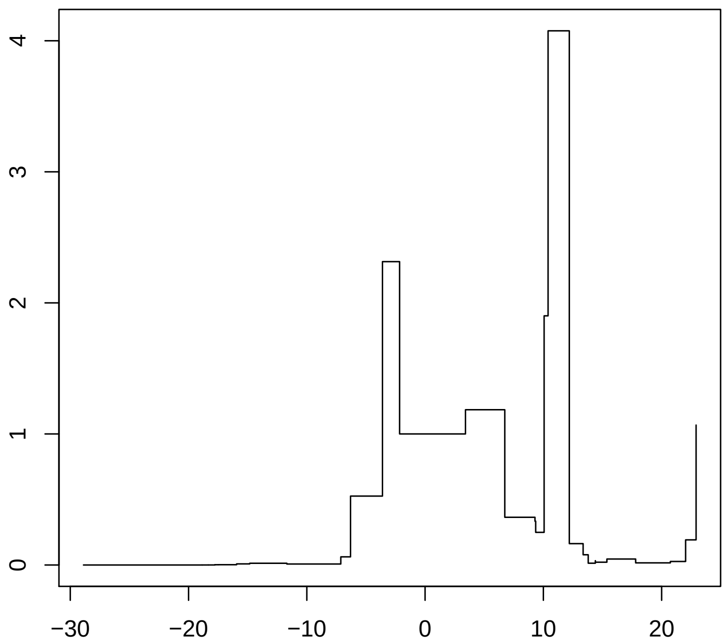
$$\tilde{\vartheta}_n = \int_\alpha^\beta \vartheta \, p(\vartheta | X^n) \, \mathrm{d}\vartheta = \frac{\int_\alpha^\beta \vartheta \, p(\vartheta) \, L(\vartheta, X^n) \, \mathrm{d}\vartheta}{\int_\alpha^\beta p(\vartheta) \, L(\vartheta, X^n) \, \mathrm{d}\vartheta}. \tag{2}$$

Here $p(\vartheta | X^n)$ is the conditional density defined by the last equality. It is always assumed that the density $p(\vartheta), \alpha < \vartheta < \beta$ is a continuous positive function in this paper.

1.1 Calculation of Estimators

It is easy to see that due to the presence of the indicator functions, the likelihood ratio is piece-wise constant. Let $X_{(j)}, j = 1, \ldots, M$ be the ordered values of $X_j \in \Theta$, i.e., if $X_j \in \Theta$, then $\alpha \equiv X_{(0)} < X_{(1)} < X_{(2)} < \ldots < X_{(M)} < \beta \equiv X_{(M+1)}$. Obviously, $M \leq n$ and if $\Theta = R$, then $M = n$. Therefore the function $L(\vartheta, X^n)$ is right-continuous with left-hand limits (càdlàg) and takes constant values (denoted as L_m) on the intervals $[X_{(m)}, X_{(m+1)})$. Hence, the values $\hat{\vartheta}_n$ satisfying the equation (1) constitute a random interval.

To calculate the MLE $\hat{\vartheta}_n$, one proceeds as follows. Having observations X^n, select the values X_j satisfying the condition $X_j \in \Theta$, calculate $X_{(1)}, \ldots, X_{(M)}$ and $L_m, m = 1, \ldots, M+1$ and define $\hat{m}$ by the equation: $\max_m L_m = L_{\hat{m}}$. Then take $\hat{\vartheta}_n$ as any point in the interval $[X_{(\hat{m})} X_{(\hat{m}+1)})$.



Likelihood ratio.

To calculate the BE, note that the integrals in (2) can be written as the following sums

$$\int_\alpha^\beta \vartheta\, p(\vartheta)\, L(\vartheta, X^n)\ \mathrm{d}\vartheta = \sum_{m=0}^{M+1} L_m\, h_m, \qquad h_m = \int_{X_{(m)}}^{X_{(m+1)}} \vartheta p(\vartheta)\ \mathrm{d}\vartheta,$$

$$\int_\alpha^\beta p(\vartheta)\, L(\vartheta, X^n)\ \mathrm{d}\vartheta = \sum_{m=0}^{M+1} L_m\, g_m, \qquad g_m = \int_{X_{(m)}}^{X_{(m+1)}} p(\vartheta)\ \mathrm{d}\vartheta.$$

Therefore calculating the BE is tantamount to evaluating the sum

$$\tilde{\vartheta}_n = \frac{\sum_{m=0}^{M+1} L_m\, h_m}{\sum_{m=0}^{M+1} L_m\, g_m}.$$

In the particular case that the prior distribution is uniform, then

$$\tilde{\vartheta}_n = \frac{\sum_{m=0}^{M} L_m \ \left(X_{(m+1)}^2 - X_{(m)}^2\right)}{2\sum_{m=0}^{M} L_m \ \left(X_{(m+1)} - X_{(m)}\right)}.$$

Let $M = n$ and put

$$q_m = \frac{L_m \left(X_{(m+1)} - X_{(m)}\right)}{\sum_{i=0}^{n+1} L_i \left(X_{(i+1)} - X_{(i)}\right)}, \quad q_m > 0, \quad \sum_{m=0}^{n+1} q_m = 1.$$

Then the BE admits the simple representation

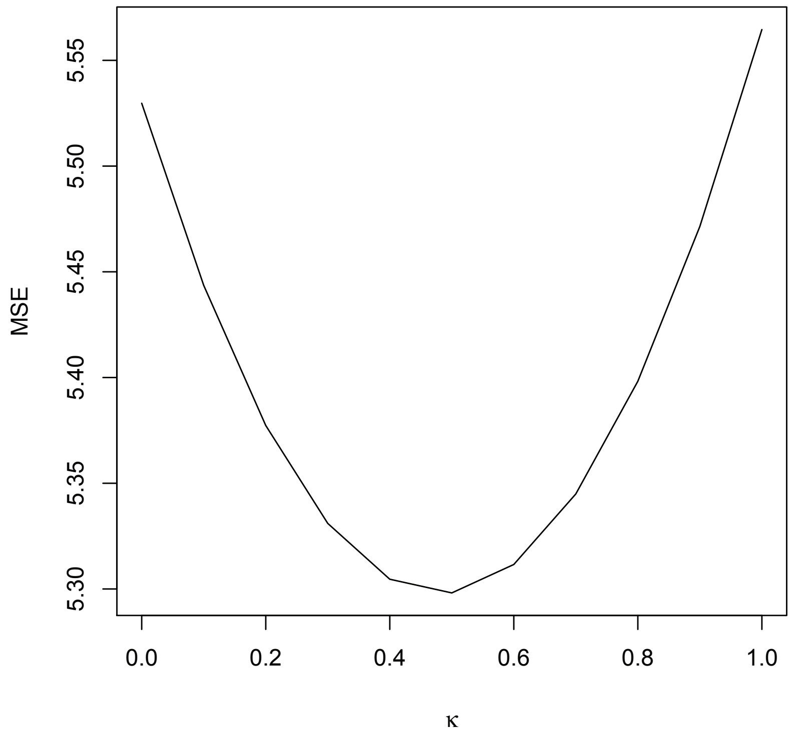
$$\tilde{\vartheta}_n = \sum_{m=1}^{n} \frac{X_{(m)} + X_{(m+1)}}{2} \ q_m. \tag{3}$$

As a result, the representation (3) is asymptotically valid for all Bayesian estimators.

As far as the MLE is concerned, one still faces the issue of choosing a point in the random interval $[X_{(\hat{m})}, X_{(\hat{m}+1)})$. One possibility is choosing the middle point of this interval as the MLE

$$\hat{\vartheta}_n = \frac{X_{(\hat{m})} + X_{(\hat{m}+1)}}{2}. \tag{4}$$

With this choice, representations (3) and (4) indicate that these estimators can be calculated conveniently. For example, in calculating the MLE, we only need to keep track of the maximal values of the likelihood ratio (LR) and to compare it with the new values of likelihood ratio. Moreover, if the value $L_{\hat{m}}(X_{(\hat{m}+1)} - X_{(\hat{m})})$ is essentially larger than all other values, then the patterns of BE and MLE are similar because the weight $q_{\hat{m}}$ in (3) is almost 1.



Limit Variance of the MLE.

2 Asymptotic Efficiency

To describe the properties of the estimators as $n \to \infty$, note that the likelihood ratio $L(\vartheta, X^n), \vartheta \in \Theta$ is a discontinuous function of ϑ. In such singular estimation problems, the MLE and the BE usually have a *singular* rate of convergence, i.e., convergence in distribution as follows:

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

The limit variance of the BE usually does not vary with the prior density $p(\cdot)$. To compare these estimators, the limit mean-squared errors $\mathbf{E}_\vartheta \hat{u}_\vartheta^2$ and $\mathbf{E}_\vartheta \tilde{u}_\vartheta^2$ are compared.

We establish the following minimax lower bound for the mean-squared risk of all estimators:
For any $\vartheta_0 \in \Theta$ *and any estimator* $\bar{\vartheta}_n$*, we have*

$$\varliminf_{\delta \to 0} \varliminf_{n \to \infty} \sup_{|\vartheta - \vartheta_0| < \delta} n^2 \mathbf{E}_\vartheta \left(\bar{\vartheta}_n - \vartheta\right)^2 \geq \mathbf{E}_{\vartheta_0} \tilde{u}_{\vartheta_0}^2. \qquad (5)$$

After establishing this lower bound, the BE is asymptotically more efficient because the MLE has a larger limit variance than the BE. This is one of the reasons that a main focus of this work is on the asymptotic behavior of the Bayesian estimators for various TAR models.

3 Gaussian TAR

To illustrate the asymptotic properties of the estimators, first start with the simple TAR process studied in Chan (1993) and Chan and Kutoyants (2009). Suppose that

$$X_{j+1} = \rho_1 \, X_j \, 1\!\!1_{\{X_j<\vartheta\}} + \rho_2 \, X_j \, 1\!\!1_{\{X_j\geq\vartheta\}} + \varepsilon_{j+1}, \qquad j = 0, \ldots, n-1, \tag{6}$$

where ε_j are i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$, $\rho_1 \neq \rho_2$ and $|\rho_2| < 1$. This time series is exponentially mixing and let $\varphi\left(x, \vartheta\right)$ be its invariant density function. The likelihood ratio has the following form

$$\ln L\left(\vartheta, X^n\right) = \ln f_0\left(X_0\right) - \frac{n}{2} \ln\left(2\pi\sigma^2\right)$$
$$- \frac{1}{2\sigma^2} \sum_{j=0}^{n-1} \left(X_{j+1} - \rho_1 \, X_j \, 1\!\!1_{\{X_j<\vartheta\}} - \rho_2 \, X_j \, 1\!\!1_{\{X_j\geq\vartheta\}}\right)^2 .$$

We now study the normalized likelihood ratio (ϑ_0 is the true value)

$$Z_n(u) = \frac{L\left(\vartheta_0 + \frac{u}{n}, X^n\right)}{L\left(\vartheta_0, X^n\right)}, \qquad u \in \mathbb{U}_n = \left[n\left(\alpha - \vartheta_0\right), n\left(\beta - \vartheta_0\right)\right].$$

Therefore ($\rho = \rho_2 - \rho_1$ and $u > 0$)

$$\log Z_n(u) = -\frac{1}{2\sigma^2} \sum_{j=0}^{n-1} \left[\rho^2 X_j^2 + 2\rho X_j \varepsilon_{j+1}\right] \; 1\!\!\mathrm{I}_{\left\{\vartheta \le X_j < \vartheta + \frac{u}{n}\right\}}.$$

The argument goes as follows. Fix u and study the limit of this sum as $n \to \infty$. Note that for large n there are only a few non-zero terms in this sum because the band $\left[\vartheta, \vartheta + \frac{u}{n}\right]$ becomes narrow and due to the boundedness of the invariant density $\varphi(x, \vartheta)$, $x \in \left[\vartheta, \vartheta + \frac{u}{n}\right]$, we have approximately $u\varphi(\vartheta, \vartheta)$ terms. Moreover, the values of X_j in these terms are close to ϑ and the distance between non-zero terms increases to infinity with n.

Due to the strong mixing feature of the time series, all non-zero terms will have asymptotically independent X_j and ε_{j+1}. These observations lead to the following limits

$$\sum_{j=0}^{n-1} X_j^2 \, \mathbb{1}_{\left\{\vartheta \leq X_j < \vartheta + \frac{u}{n}\right\}} \Longrightarrow \vartheta^2 \tilde{N}_+ (u) \,,$$

and

$$\sum_{j=0}^{n-1} X_j \, \varepsilon_{j+1} \, \mathbb{1}_{\left\{\vartheta \leq X_j < \vartheta + \frac{u}{n}\right\}} \Longrightarrow \sigma\vartheta \sum_{l=1}^{\tilde{N}_+(u)} \epsilon_l^+ \,,$$

where $\tilde{N}_+ (u) \,, u \geq 0$ is a homogeneous Poisson process with intensity function $\lambda = \varphi (\vartheta, \vartheta)$ and ϵ_l^+ are independent standard normal random variables that are independent of $\tilde{N}_+ (u)$.

Next, define the stochastic process

$$Z_0(u) = \begin{cases} \exp\left\{\gamma \sum_{l=1}^{N_+(u)} \left(\varepsilon_l^+ - \frac{\gamma}{2}\right)\right\}, & u \geq 0, \\ \exp\left\{\gamma \sum_{l=1}^{N_-(-u)} \left(\varepsilon_l^- - \frac{\gamma}{2}\right)\right\}, & u < 0, \end{cases} \tag{7}$$

where $\gamma = \frac{\rho\vartheta}{\sigma}$, $N_+(\cdot)$ and $N_-(\cdot)$ are two independent Poisson processes with intensities 1 and $\varepsilon_l^+, \varepsilon_l^-$ are independent standard normal random variables. Write $Z(u) = Z_0(\lambda u)$, where $\lambda = \varphi(\vartheta, \vartheta)$.

Introduce two random variables $\hat{u}_\vartheta$ and $\tilde{u}_\vartheta$ for the MLE and the BE respectively by the relations

$$\hat{u}_\vartheta = \arg\sup_u Z(u), \quad \tilde{u}_\vartheta = \frac{\int_{\mathcal{R}} u\, Z(u)\, \mathrm{d}u}{\int_{\mathcal{R}} Z(u)\, \mathrm{d}u}$$

From the result of Chan (1993), it follows that the MLE satisfies

$$n\left(\hat{\vartheta}_n - \vartheta\right) \Longrightarrow \hat{u}_\vartheta = \frac{u_{(\hat{m})} + u_{(\hat{m}+1)}}{2},$$

where $\left[u_{(\hat{m})}, u_{(\hat{m}+1)}\right]$ is the interval of the maximum of the process $Z\left(\cdot\right)$.

Our result:

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

we have convergence of moments

$$n^p \mathbf{E}_\vartheta \left|\tilde{\vartheta}_n - \vartheta\right|^p \to \mathbf{E}_\vartheta \left|\tilde{u}_\vartheta\right|^p$$

and BE are asymptotically efficient (Chan and K, 2009).

Example. Consider the TAR model with $\varepsilon_j \sim \mathcal{N}(0,1)$. Then numerical simulations give rise to the values

$$\mathbf{E}_\vartheta \hat{u}_\vartheta^2 \approx 22.83, \qquad \mathbf{E}_\vartheta \tilde{u}_\vartheta^2 \approx 17.$$

For the MLE ($\vartheta = \vartheta_0 + u/n$):

$$\begin{aligned}
&\mathbf{P}_{\vartheta_0}\left\{n\left(\hat\vartheta_n - \vartheta_0\right) < x\right\} \\
&\quad = \mathbf{P}_{\vartheta_0}\left\{\sup_{n(\vartheta-\vartheta_0)<x} L\left(\vartheta, X^n\right) > \sup_{n(\vartheta-\vartheta_0)\geq x} L\left(\vartheta, X^n\right)\right\} \\
&\quad = \mathbf{P}_{\vartheta_0}\left\{\sup_{n(\vartheta-\vartheta_0)<x} \frac{L\left(\vartheta, X^n\right)}{L\left(\vartheta_0, X^n\right)} > \sup_{n(\vartheta-\vartheta_0)\geq x} \frac{L\left(\vartheta, X^n\right)}{L\left(\vartheta_0, X^n\right)}\right\} \\
&\quad = \mathbf{P}_{\vartheta_0}\left\{\sup_{u<x} \frac{L\left(\vartheta_0 + \frac{u}{n}, X^n\right)}{L\left(\vartheta_0, X^n\right)} > \sup_{u\geq x} \frac{L\left(\vartheta_0 + \frac{u}{n}, X^n\right)}{L\left(\vartheta_0, X^n\right)}\right\} \\
&\quad = \mathbf{P}_{\vartheta_0}\left\{\sup_{u<x} Z_n\left(u\right) > \sup_{u\geq x} Z_n\left(u\right)\right\} \\
&\quad \longrightarrow \mathbf{P}_{\vartheta_0}\left\{\sup_{u<x} Z\left(u\right) > \sup_{u\geq x} Z\left(u\right)\right\} = \mathbf{P}_{\vartheta_0}\left(\hat u_{\vartheta_0} < x\right),
\end{aligned}$$

For the BE, change the variable $\vartheta = \vartheta_0 + u/n \equiv \vartheta_u$

$$
\begin{aligned}
\tilde{\vartheta}_n &= \frac{\int_\alpha^\beta \vartheta p\left(\vartheta\right) L\left(\vartheta, X^n\right) \mathrm{d}\vartheta}{\int_\alpha^\beta p\left(\vartheta\right) L\left(\vartheta, X^n\right) \mathrm{d}\vartheta} \\
&= \vartheta_0 + \frac{1}{n} \frac{\int_{\mathbb{U}_n} up\left(\vartheta_u\right) L\left(\vartheta_u, X^n\right) \mathrm{d}u}{\int_{\mathbb{U}_n} p\left(\vartheta_u\right) L\left(\vartheta_u, X^n\right) \mathrm{d}u} \\
&= \vartheta_0 + \frac{1}{n} \frac{\int_{\mathbb{U}_n} up\left(\vartheta_u\right) \frac{L\left(\vartheta_0 + \frac{u}{n}, X^n\right)}{L\left(\vartheta_0, X^n\right)} \mathrm{d}u}{\int_{\mathbb{U}_n} p\left(\vartheta_u\right) \frac{L\left(\vartheta_0 + \frac{u}{n}, X^n\right)}{L\left(\vartheta_0, X^n\right)} \mathrm{d}u} \\
&= \vartheta_0 + \frac{1}{n} \frac{\int_{\mathbb{U}_n} u\, p\left(\vartheta_u\right) Z_n\left(u\right)\ \mathrm{d}u}{\int_{\mathbb{U}_n} p\left(\vartheta_u\right) Z_n\left(u\right)\ \mathrm{d}u}\ .
\end{aligned}
$$

Then using the convergence $p\left(\vartheta_u\right) \rightarrow p\left(\vartheta_0\right) > 0$, one obtains

$$\begin{aligned}
&\mathbf{P}_{\vartheta_0}\left\{n\left(\tilde{\vartheta}_n - \vartheta_0\right) < x\right\} \\
&\qquad = \mathbf{P}_{\vartheta_0}\left\{\frac{\int_{\mathbb{U}_n} u\, p\left(\vartheta_u\right) Z_n\left(u\right)\ \mathrm{d}u}{\int_{\mathbb{U}_n} p\left(\vartheta_u\right) Z_n\left(u\right)\ \mathrm{d}u} < x\right\} \\
&\qquad \longrightarrow \mathbf{P}_{\vartheta_0}\left\{\frac{\int_R u\, Z\left(u\right)\ \mathrm{d}u}{\int_R Z\left(u\right)\ \mathrm{d}u} < x\right\} = \mathbf{P}_{\vartheta_0}\left(\tilde{u}_{\vartheta_0} < x\right) .
\end{aligned}$$

Hence

$$n\left(\tilde{\vartheta}_n - \vartheta_0\right) \Longrightarrow \tilde{u}_{\vartheta_0}$$

3.1 Three unknown parameters

To account for the other unknown parameters, the following generalization is also given in Chan and Kutoyants (2009). Suppose that all three parameters $(\vartheta, \rho_1, \rho_2)$ in the model

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

are unknown, then under identifiability condition, it is shown that the BE of these parameters $\left(\tilde{\vartheta}_n, \tilde{\rho}_{1,n}, \tilde{\rho}_{2,n}\right)$ are consistent.

We have the convergence

$$n\left(\tilde{\vartheta}_n-\vartheta_0\right)\Longrightarrow \tilde{u}_{\vartheta_0},\qquad \sqrt{n}\left(\tilde{\rho}_{1.n}-\rho_{1,0}\right)\Longrightarrow \frac{\zeta_1}{\mathrm{I}_1},$$

$$\sqrt{n}\left(\tilde{\rho}_{2,n}-\rho_{2,0}\right)\Longrightarrow \frac{\zeta_2}{\mathrm{I}_2},$$

where $\mathrm{I}_1=\sigma^{-1}\mathbf{E}\left[\xi^2 1\!\!1_{\{\xi<\vartheta_0\}}\right], \mathrm{I}_2=\sigma^{-1}\mathbf{E}\left[\xi^2 1\!\!1_{\{\xi\geq\vartheta_0\}}\right]$ (ξ is a random variable with density function $\varphi\left(x,\vartheta_0\right)$). The three random variables $\zeta_1\sim\mathcal{N}\left(0,1\right)$, $\zeta_2\sim\mathcal{N}\left(0,1\right)$ and $\tilde{u}_{\vartheta_0}$ are independent. Note that if it is further supposed that the variance σ^2 is also unknown, then a simple modification of the proof allows one to establish the asymptotic normality of the estimator $\tilde{\sigma}_n^2$, say, $\sqrt{n}\left(\tilde{\sigma}_n^2-\sigma_0^2\right)\Rightarrow\eta\sim\mathcal{N}\left(0,2\sigma^4\right)$. In this situation, the random variable η will be correlated with ζ_1 and ζ_2, but independent of $\tilde{u}_{\vartheta_0}$.

3.2 Misspecification

It is interesting to observe that the singular estimation and testing problems are in some sense more robust than the regular ones. This phenomenon can be illustrated as follows. Suppose that the observed process is

$$X_{j+1} = \rho_1 \, X_j \, 1\!\!1_{\{X_j<\vartheta\}} + \rho_2 \, X_j \, 1\!\!1_{\{X_j\geq\vartheta\}} + h\left(X_j\right) + \varepsilon_{j+1}, \; j = 0, \ldots, n-1,$$

where $\vartheta \in (\alpha, \beta)\,, \alpha > 0$, ε_j are i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$, $\rho_1 \neq \rho_2$ and $h\left(x\right)$ is such that the time series has the appropriate mixing properties with the invariant density $\varphi_h\left(x, \vartheta\right)$. In this setting, the function $h\left(x\right)$ is unknown. But one uses this model (with $h\left(\cdot\right) \equiv 0$) to estimate ϑ instead.

In singular estimation problems, consistent estimation is however possible for a large class of functions $h\left(\cdot\right)$.

If we assume that

$$h\left(x\right) > \frac{\rho_1 - \rho_2}{2}\, x, \qquad \vartheta_0 < x \leq \beta,$$

and

$$h\left(x\right) < \frac{\rho_2 - \rho_1}{2}\, x, \qquad \alpha \leq x < \vartheta_0\,,$$

then the MLE is consistent.

Note that similar results were also obtained for ergodic diffusion processes and inhomogeneous Poisson processes.

3.3 Windows

To illustrate the effect of windows, consider again the "toy model"

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

where $\vartheta \in (\alpha, \beta)\,, \alpha > 0$, ε_j are i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$, $\rho_1 \neq \rho_2$. In practice, it may sometimes be difficult to observe the values of X_j outside a certain range (for example, censorship or measuring devices with bounded scales). Using observations X_j restricted to a chosen narrow window $\mathbb{W}_n$ with the width of this window tending to zero, it is shown that it is possible to construct estimators based on these restricted observations that are consistent or even asymptotically efficient for the class of estimators constructed from complete observations.

This result was first established for diffusion processes, but it is of general nature and can be applied to many other models with singular situations such as nonlinear time series.

To see how this works, suppose there exist some consistent and asymptotically normal estimators $\bar{\vartheta}_{N_0}$ constructed by the first $N_0 = [\sqrt{n}]$ observations $X_1, \ldots, X_{N_0}$, i.e., $\sqrt{N_0}\left(\bar{\vartheta}_{N_0} - \vartheta\right) \Rightarrow \mathcal{N}\left(0, r\left(\vartheta\right)^2\right)$ and suppose further that converegence of moments is established. For example, one can use the method of moments to obtain $\bar{\vartheta}_{N_0}$.

Next, introduce the windows

$$\mathbb{W}_n = \left[\bar{\vartheta}_{N_0} - n^{-1/8}, \bar{\vartheta}_{N_0} + n^{-1/8}\right], \qquad n = 1, 2, \ldots$$

and the *pseudo likelihood ratio*

$$\begin{aligned} &\ln \tilde{L}_n\left(\vartheta, X^n\right) \\ &= -\sum_{j=N_0+1}^{n-1} \left(X_{j+1} - \rho_1\, X_j\, 1\!\!1_{\{X_j<\vartheta\}} - \rho_2\, X_j\, 1\!\!1_{\{X_j\geq\vartheta\}}\right)^2\, 1\!\!1_{\{X_j\in\mathbb{W}_n\}}. \end{aligned}$$

Keep all values X_l that fall into the window $\mathbb{W}_n$ and keep the corresponding succeeding values X_{l+1}. This window has width $2n^{-1/8} \to 0$ and obviously there are very few observations available.

But the study of the asymptotic behavior of the likelihood ratio shows that only these observations contribute to the construction of the limit process $Z(\cdot)$ Now construct the "pseudo-MLE" $\hat{\vartheta}_n$ and "pseudo-BE" $\tilde{\vartheta}_n$ as follows:

$$\tilde{L}_n\left(\hat{\vartheta}_n, Y^n\right) = \sup_{\vartheta\in\Theta} \tilde{L}_n\left(\vartheta, Y^n\right), \qquad \tilde{\vartheta}_n = \frac{\int_\Theta \vartheta\, p\left(\vartheta\right)\ \tilde{L}_n\left(\vartheta, Y^n\right) \mathrm{d}\vartheta}{\int_\Theta\ p\left(\vartheta\right)\ \tilde{L}_n\left(\vartheta, Y^n\right) \mathrm{d}\vartheta},$$

where $p(\cdot)$ is continuous positive on Θ density. To study these estimators, introduce the normalized likelihood ratio

$$\tilde{Z}_n\left(u\right) = \frac{\tilde{L}_n\left(\vartheta_0 + \frac{u}{n}, X^n\right)}{\tilde{L}_n\left(\vartheta_0, X^n\right)}, \qquad u \in \mathbb{U}_n.$$

The LR $\tilde{Z}_n(u) = \exp\left\{-\frac{1}{2\sigma^2}\tilde{Y}_n(u)\right\}$, where

$$\tilde{Y}_n(u) = \sum_{j=N_0+1}^{n} \left[\rho^2 X_j^2 + 2\rho X_j\,\varepsilon_{j+1}\right] \mathbb{1}_{\left\{\vartheta_0 \le X_j < \vartheta_0 + \frac{u}{n}\right\}} \mathbb{1}_{\{X_j \in \mathbb{W}_n\}}.$$

Here, u is assumed to be positive and $\rho = \rho_2 - \rho_1$. Note that with a large probability, one has

$$\mathbb{1}_{\left\{\vartheta_0 \le X_j < \vartheta_0 + \frac{u}{n}\right\}} \mathbb{1}_{\{X_j \in \mathbb{W}_n\}} = \mathbb{1}_{\left\{\vartheta_0 \le X_j < \vartheta_0 + \frac{u}{n}\right\}}.$$

Therefore the limit of $\tilde{Y}_n(u)$ is the same as the limit of $Y_n(u)$ obtained before. Once again we obtain the consistency, convergence in distribution

$$n\left(\tilde{\vartheta}_n - \vartheta_0\right) \Longrightarrow \tilde{u}_{\vartheta_0} \qquad (8)$$

and the convergence of moments. Note that such windows were already used in a similar problem with threshold ergodic diffusion processes and inhomogeneous Poisson processes.

4 Multi-Thresholds

Consider a TAR time series with multiple thresholds

$$X_{j+1} = \sum_{l=0}^{L} \rho_l X_j \mathbb{1}_{\{\vartheta_l < X_j \le \vartheta_{l+1}\}} + \varepsilon_{j+1}, \quad j = 1, \ldots, n-1, \tag{9}$$

where $\varepsilon_j \sim \mathcal{N}\left(0, \sigma^2\right)$. Suppose that all thresholds are unknown $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_L)$ and one has to estimate $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Here $\boldsymbol{\Theta} = (\alpha_1, \beta_1) \times \ldots \times (\alpha_L, \beta_L)$ and $\beta_l < \alpha_{l+1}$. Let $\vartheta_0 = -\infty$ and $\vartheta_{L+1} = \infty$. The likelihood function is

$$L\left(\boldsymbol{\vartheta}, X^n\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}}$$

$$\exp\left\{-\frac{1}{2\sigma^2} \sum_{j=0}^{n-1} \left[X_{j+1} - \sum_{l=0}^{L} \rho_l X_j \mathbb{1}_{\{\vartheta_l < X_j \le \vartheta_{l+1}\}}\right]^2\right\}$$

Define the BE $\tilde{\boldsymbol{\vartheta}}_n$, which corresponds to the quadratic loss function by the formula

$$\tilde{\boldsymbol{\vartheta}}_n = \int_{\boldsymbol{\Theta}} \boldsymbol{\vartheta} p\left(\boldsymbol{\vartheta} | X^n\right) \, \mathrm{d}\boldsymbol{\vartheta} = \frac{\int_{\boldsymbol{\Theta}} \boldsymbol{\vartheta} p\left(\boldsymbol{\vartheta}\right) L\left(\boldsymbol{\vartheta}, X^n\right) \, \mathrm{d}\boldsymbol{\vartheta}}{\int_{\boldsymbol{\Theta}} p\left(\boldsymbol{\vartheta}\right) L\left(\boldsymbol{\vartheta}, X^n\right) \, \mathrm{d}\boldsymbol{\vartheta}}.$$

Here $p\left(\cdot\right)$ is continuous positive prior density. To study the asymptotic properties of these estimators, introduce the normalized likelihood ratio

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

Let

$$\mathbb{B}_{l,j}\left(u_l\right) = \left\{\omega : \quad \vartheta_l < X_j \leq \vartheta_l + \frac{u_l}{n}\right\}.$$

Further, the following weak convergence also holds:

$$\sum_{j=0}^{n-1} X_j^2 \, 1\!\!\mathrm{I}_{\{\mathbb{B}_{l,j}(u_l)\}} \Longrightarrow \vartheta_l^2 N_l\left(u_l\right),$$

$$\sum_{j=0}^{n-1} X_j \, \varepsilon_{j+1} \, 1\!\!\mathrm{I}_{\{\mathbb{B}_{l,j}(u_l)\}} \Longrightarrow \vartheta_l \sigma \sum_{i=1}^{N_l(u_l)} \epsilon_{l,i+1},$$

where $N_l\left(\cdot\right)$ are independent Poisson processes of intensities $\lambda_l = \varphi\left(\vartheta_l, \boldsymbol{\vartheta}\right)$ and $\epsilon_{l,i}$ are independent standard normal random variables. Let $\gamma_l = \frac{\rho_{l+1}-\rho_l}{\sigma}\vartheta_l$. Then the following weak convergence holds

$$Z_n\left(\boldsymbol{u}\right) \Longrightarrow Z\left(\boldsymbol{u}\right) = \prod_{l=1}^{L} Z_l\left(u_l\right) = \prod_{l=1}^{L} \exp\left\{\gamma_l \sum_{i=1}^{N_l(u_l)} \left(\epsilon_{l,i+1} - \frac{\gamma_l}{2}\right),\right\},$$

i.e., the limit likelihood ratio is the product of L independent likelihood ratios.

This form of $Z(\cdot)$ yields the limit distributions of the Bayesian estimators also. From this argument, it can be seen that the estimators of the different thresholds have asymptotically independent distributions

$$n\left(\tilde{\vartheta}_{1,n}-\vartheta_1\right)\Longrightarrow \tilde{u}_{1,\boldsymbol{\vartheta}},\quad \ldots,\quad n\left(\tilde{\vartheta}_{L,n}-\vartheta_L\right)\Longrightarrow \tilde{u}_{L,\boldsymbol{\vartheta}},$$

where

$$\tilde{u}_{l,\boldsymbol{\vartheta}}=\frac{\int_{\mathcal{R}} u\, Z_l\left(u\right)\mathrm{d}u}{\int_{\mathcal{R}} Z_l\left(u\right)\mathrm{d}u},\qquad l=1,\ldots,L.$$

5 Higher-Order TAR(p) Model

Let

$$X_{j+1} = \sum_{i=0, i\neq k}^{p} \rho_i X_{j-i} + \rho_k^- X_{j-k} \mathbb{1}_{\{X_{j-k}<\vartheta\}} + \rho_k^+ X_{j-k} \mathbb{1}_{\{X_{j-k}\geq\vartheta\}} + \varepsilon_{j+1} \,,$$

where $0 \leq k \leq p$ is known, $\rho_k^- \neq \rho_k^+$, ε_j are independent $\mathcal{N}\left(0, \sigma^2\right)$ and $j = p, \ldots, n-1$. Suppose that the coefficients ρ_i are such that the time series is strongly mixing with invariant density $\varphi\left(x, \vartheta\right)$. To estimate the parameter ϑ, consider the normalized likelihood ratio

$$Z_n\left(u\right) = \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=p}^{n-1} \left(\rho_k^2 X_{j-k}^2 + 2\rho_k\, X_{j-k}\, \varepsilon_{j+1}\right) \mathbb{1}_{\left\{\vartheta \leq X_{j-k} < \vartheta + \frac{u}{n}\right\}}\right\}$$

where $\rho_k = \rho_k^+ - \rho_k^-$.

We have (for $u > 0$)

$$Z_n(u) \Longrightarrow Z(u) = \exp\left\{\gamma_k \sum_{l=1}^{N_+(u)} \left(\epsilon_l - \frac{\gamma_k}{2}\right)\right\},$$

where $\gamma_k = \frac{\rho_k \vartheta}{\sigma}$ and $\tilde{N}_+(u)$ is a homogeneous Poisson process with intensity function $\lambda = \varphi(\vartheta, \vartheta)$. A similar result can also be obtained for $u < 0$. The random variables ϵ_l are usually standard normal, which are independent of the Poisson process. The limit distribution of the BE $n\left(\tilde{\vartheta}_n - \vartheta\right) \Rightarrow \tilde{u}_\vartheta$ is defined as in the preceding result by this limit process. It is interesting to note that if the observed process is

$$X_{j+1} = \sum_{l=0, l\neq k}^{p} \rho_l X_{j-l} + \rho_k^- X_{j-k} 1\!\!1_{\{X_{j-m} < \vartheta\}} + \rho_k^+ X_{j-k} 1\!\!1_{\{X_{j-m} \geq \vartheta\}} + \varepsilon_{j+1}$$

where $m \neq k$, then the situation completely changes.

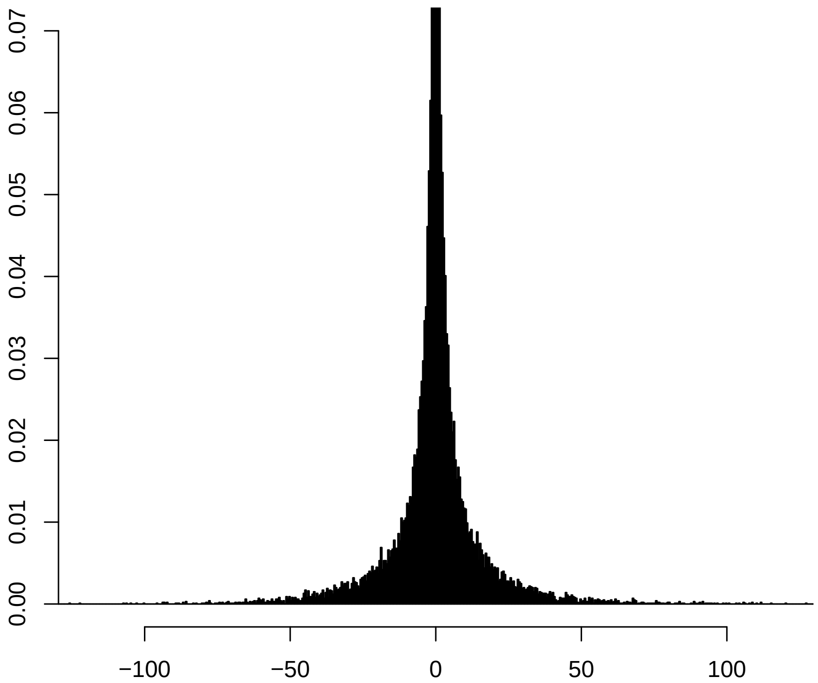
In this case, the normalized likelihood ratio becomes

$$Z_n(u) = \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=p}^{n-1} \left(\rho_k^2 X_{j-k}^2 + 2\rho_k\, X_{j-k}\varepsilon_{j+1} \right) \mathbb{I}_{\left\{\vartheta \le X_{j-m} < \vartheta + \frac{u}{n}\right\}} \right\}$$

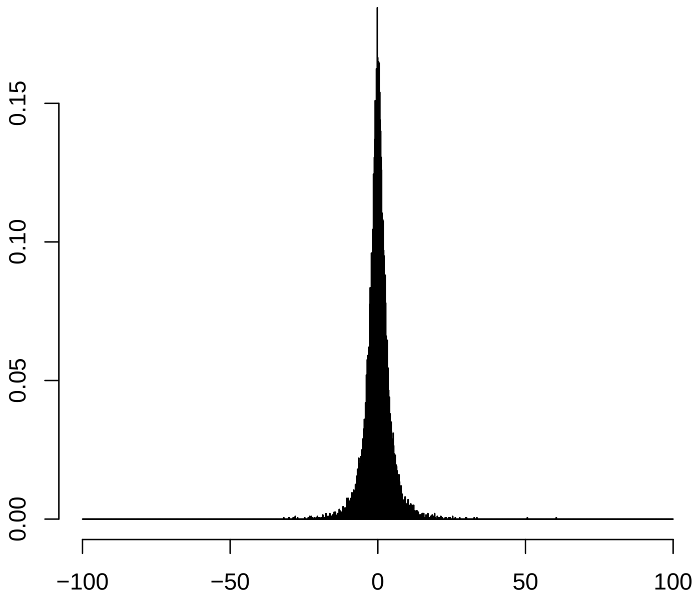
It can be shown that the limit of such a process is $(u > 0)$

$$Z(u) = \exp\left\{ \sum_{l=1}^{N_+(u)} \left(\epsilon_l - \frac{\xi_l \rho_k}{2\sigma} \right) \frac{\xi_l \rho_k}{\sigma} \right\},$$

where ξ_l are independent random variables with density $\varphi(x, \vartheta)$ (the invariant distribution). The random variables ϵ_l are standard normal, independent of ξ_l, but the Poisson process $N_+(\cdot)$ depends on ξ_l.



MLE.



BE.

6 Related Problems

6.1 Partially Observed Systems

Let us consider the time series

$$X_{j+1} = \rho_1 X_j 1\!\!1_{\{X_j < \theta\}} + \rho_2 X_j 1\!\!1_{\{X_j \geq \theta\}} + \xi_{j+1} + \epsilon_{j+1}, \quad j \geq 0$$

where $(\epsilon_j)_{j\geq 0}$ is a random process with known distribution, ρ_1 and ρ_2 are known constants and θ is the unknown threshold parameter to be estimated from the sample $X^n := (X_1, ..., X_n)$. The driving noises $(\varepsilon_j)_{j\geq 0}$ and $(\xi_j)_{j\geq 0}$ are independent: the *white noise* component (ε_j) is a sequence of i.i.d. $N(0,1)$ random variables, the *colored noise* (ξ_j) is the Gaussian process, generated by the linear recursion

$$\xi_j = a\xi_{j-1} + \zeta_j, \quad j \geq 1,$$

where $(\zeta_j)_{j\geq 0}$ are i.i.d. $N(0,1)$ random variables and a is a known constant $|a| < 1$, controlling the bandwidth of the noise. The process $\widehat{\xi}_j(\theta) := \mathbf{E}_\theta\big(\xi_j | F_j^X\big)$ and the deterministic sequence $\gamma_j := \mathbf{E}_\theta\big(\xi_j(\theta) - \widehat{\xi}_j\big)^2$ satisfy the Kalman filter equations

$$\widehat{\xi}_j(\theta) = a\widehat{\xi}_{j-1}(\theta) + \frac{a\gamma_{j-1}}{1+\gamma_{j-1}}\Big(X_j - f(X_{j-1},\theta) - \widehat{\xi}_{j-1}(\theta)\Big)$$

$$\gamma_j = a^2\gamma_{j-1} + 1 - \frac{a^2\gamma_{j-1}^2}{1+\gamma_{j-1}}, \quad \gamma_0 = \mathrm{var}(\xi_0)$$

subject to $\widehat{\xi}_0 = 0$. Here $f(x,\theta) := \big(\rho_2 \mathbb{1}_{\{x\geq\theta\}} + \rho_1 \mathbb{1}_{\{x<\theta\}}\big)x$.

The limit log-LR is (Chigansky, K., Liptser, 2011)

$$\ln Z(u) = \begin{cases} \sum_{j=1}^{N^+(u)} \left(\beta\varepsilon_j^+ - \frac{1}{2}\beta^2\right) & u \geq 0 \\ \\ \sum_{j=1}^{N^-(|u|)} \left(\beta\varepsilon_j^- - \frac{1}{2}\beta^2\right) & u < 0 \end{cases} \tag{10}$$

where

$$\beta^2 = \theta_0^2(\rho_2 - \rho_1)^2 \frac{1+\gamma^3}{(1+\gamma)(1+\gamma^2)},$$

and the Bayes estimators have the convergence

$$n\left(\tilde{\vartheta}_n - \vartheta\right) \Longrightarrow \tilde{u}, \qquad \mathbf{E}_\vartheta \left|n\left(\tilde{\vartheta}_n - \vartheta\right)\right|^p \longrightarrow \mathbf{E}_\vartheta \left|\tilde{u}\right|^p$$

Another problem (not yet solved)

Consider a two-dimensional time series

$$X_{j+1} = \rho_1 \, Y_j \, \mathbb{I}_{\{X_j < \vartheta\}} + \rho_2 \, Y_j \, \mathbb{I}_{\{X_j \geq \vartheta\}} + \varepsilon_{j+1},$$

$$Y_{j+1} = a \, Y_j + \eta_{j+1}, \qquad\qquad j = 0, 1, \ldots, n-1,$$

where $\left(\varepsilon_j\right)_{j \geq 1}$ and $\left(\eta_j\right)_{j \geq 1}$ are two independent sequences of random variables such that $\varepsilon_j \sim \mathcal{N}\left(0, \sigma^2\right)$ and $\eta_j \sim \mathcal{N}\left(0, b^2\right)$ and $X_0 = Y_0 = 0$. Suppose that $\mathbf{P}\left(Y_0 \middle| X_0\right) \sim \mathcal{N}\left(0, d^2\right)$, $|a| < 1$ and only the values $X_0, X_1, \ldots, X_n$ are observed.

This model can be labelled as the threshold Kalman-Bucy Filter (TK-BF). It is a particular case of the so-called conditionally Gaussian partially observed systems discussed in Chapter 13 of Liptser and Shiryayev.

Introduce the notation

$$A(x,\vartheta) = \rho_1 \, 1\!\!\mathrm{I}_{\{X_j<\vartheta\}} + \rho_2 \, 1\!\!\mathrm{I}_{\{X_j\geq\vartheta\}}, \qquad m_j(\vartheta) = \mathbf{E}_\vartheta \left(Y_j \middle| \mathcal{F}_j\right),$$

$$\gamma_j(\vartheta) = \mathbf{E}_\vartheta \left(\left[Y_j - m_j(\vartheta)\right]^2 \middle| \mathcal{F}_j\right).$$

Here the σ algebra $\mathcal{F}_j = \sigma(X_0, X_1, \ldots, X_j)$. Then we can write the equations of optimal filtration for conditionally Gaussian time series

$$m_{j+1}(\vartheta) = a\, m_j(\vartheta) + \frac{aA(X_j,\vartheta)\,\gamma_j(\vartheta)}{\sigma^2 + A(X_j,\vartheta)^2\,\gamma_j(\vartheta)}\left[X_{j+1} - A(X_j,\vartheta)\, m_j(\vartheta)\right]$$

$$\gamma_{j+1}(\vartheta) = a^2\gamma_j(\vartheta) + b^2 - \frac{\left[a\,A(X_j,\vartheta)\ \gamma_j(\vartheta)\right]^2}{\sigma^2 + A(X_j,\vartheta)^2\,\gamma_j(\vartheta)}, \gamma_0 = \mathbf{E}\left(Y_0 - m_0\right)^2,$$

Moreover, the random variables X_j admit the representation (innovation)

$$X_{j+1} = A\left(X_j, \vartheta\right)\, m_j\left(\vartheta\right) + \left[\sigma^2 + A\left(X_j, \vartheta\right)^2 \gamma_j\left(\vartheta\right)\right]^{1/2} \bar{\varepsilon}_{j+1},$$

where $\bar{\varepsilon}_1, \bar{\varepsilon}_2, \ldots, \bar{\varepsilon}_n$ are i.i.d. $\mathcal{N}\left(0,1\right)$. The likelihood function can be written down explicitely but we can simplify the problem and to study the leasst squares estimator using the relation

$$K\left(\vartheta_n^*, X^n\right) = \inf_{\vartheta \in \Theta} K\left(\vartheta, X^n\right)$$

where

$$K\left(\vartheta, X^n\right) = \sum_{j=0}^{n-1} \left[X_{j+1} - A\left(X_j, \vartheta\right) m_j\left(\vartheta\right)\right]^2$$

6.2 Nonlinear TAR with Non-Gaussian Noise

Returning to a general nonlinear and non-Gaussian time series, consider the general nonlinear model

$$X_{j+1} = h\left(X_j\right) \mathbb{1}_{\{X_j<\vartheta\}} + g\left(X_j\right) \mathbb{1}_{\{X_j\geq\vartheta\}} + \varepsilon_{j+1}, \qquad j = 0, \ldots, n-1,$$

where $\vartheta \in (\alpha, \beta)$, $\alpha > 0$, $h(x) \neq g(x)$, $x \in [\alpha, \beta]$, and ε_j are i.i.d. with a known *smooth* density function $f(x) > 0, x \in \mathbb{R}$. Suppose that the functions $h(\cdot)$, $g(\cdot)$ and $f(\cdot)$ are such that the time series is strong mixing with invariant density $\varphi(\cdot, \vartheta)$. The likelihood function is

$$L(\vartheta, X^n) = f_0(X_0) \prod_{j=0}^{n-1} f\left(X_{j+1} - h\left(X_j\right) \mathbb{1}_{\{X_j<\vartheta\}} - g\left(X_j\right) \mathbb{1}_{\{X_j\geq\vartheta\}}\right)$$

The normalized likelihood ratio $(u > 0, \delta(x) = g(x) - h(x))$

$$Z_n(u) = \frac{L\left(\vartheta_0 + \frac{u}{n}, X^n\right)}{L\left(\vartheta_0, X^n\right)}$$

$$= \prod_{j=0}^{n-1} \frac{f\left(X_{j+1} - h(X_j)\ \mathbb{1}_{\left\{X_j < \vartheta_0 + \frac{u}{n}\right\}} - g(X_j)\ \mathbb{1}_{\left\{X_j \geq \vartheta_0 + \frac{u}{n}\right\}}\right)}{f\left(X_{j+1} - h(X_j)\ \mathbb{1}_{\{X_j < \vartheta_0\}} - g(X_j)\ \mathbb{1}_{\{X_j \geq \vartheta_0\}}\right)}$$

$$= \prod_{j=0}^{n-1} \frac{f\left(\delta(X_j)\ \mathbb{1}_{\left\{\vartheta_0 \leq X_j < \vartheta_0 + \frac{u}{n}\right\}} + \varepsilon_{j+1}\right)}{f\left(\varepsilon_{j+1}\right)}.$$

Hence,

$$\log Z_n(u) = \sum_{j=0}^{n-1} \log \frac{f\left(\delta(X_j)\ \mathbb{1}_{\left\{\vartheta_0 \leq X_j < \vartheta_0 + \frac{u}{n}\right\}} + \varepsilon_{j+1}\right)}{f\left(\varepsilon_{j+1}\right)}$$

$$= \sum_{j=0}^{n-1} \log \left[\frac{f\left(\delta(X_j) + \varepsilon_{j+1}\right)}{f\left(\varepsilon_{j+1}\right)}\right] \mathbb{1}_{\left\{\vartheta_0 \leq X_j < \vartheta_0 + \frac{u}{n}\right\}}$$

Due to the strong mixing property, the terms in this sum are asymptotically independent and X_j is independent of ε_{j+1}. Therefore, for the characteristic function of $\log Z_n(u)$, one deduces (asymptotically) the following equality

$$\mathbf{E}_{\vartheta_0} e^{\mathrm{i}\lambda \ln Z_n(u)}$$

$$= \prod_{j=0}^{n-1} \mathbf{E}_{\vartheta_0} \exp\left\{\mathrm{i}\lambda \ln\left[\frac{f\left(\delta\left(X_j\right)+\varepsilon_{j+1}\right)}{f\left(\varepsilon_{j+1}\right)}\right] 1\!\!\mathrm{I}_{\left\{\vartheta_0 \le X_j < \vartheta_0 + \frac{u}{n}\right\}}\right\}(1+o(1))$$

$$\to \exp\left\{u\varphi\left(\vartheta_0, \vartheta_0\right) \int_{-\infty}^{\infty} \left[\exp\left\{\mathrm{i}\lambda \ln\left[\frac{f\left(\delta\left(\vartheta_0\right)+y\right)}{f\left(y\right)}\right]\right\} - 1\right] f\left(y\right) \mathrm{d}y\right\}$$

$$= \mathbf{E}_{\vartheta_0} e^{\mathrm{i}\lambda \ln Z(u)}$$

Here

$$Z(u) = \begin{cases} \exp\left\{\sum_{l=1}^{N_+(\ u\ )} \ln \frac{f(\varepsilon_l+\delta(\vartheta_0))}{f(\varepsilon_l)}\right\}, & u \geq 0, \\ \exp\left\{\sum_{l=1}^{N_-(-u)} \ln \frac{f(\varepsilon_l-\delta(\vartheta_0))}{f(\varepsilon_l)}\right\}, & u < 0, \end{cases}$$

where $N_+(u)$ is a homogeneous Poisson process with intensity $\varphi(\vartheta_0, \vartheta_0)$ and independent of ζ_l. The corresponding $\hat{u}$ and $\tilde{u}$ are

$$Z(\hat{u}) = \sup_u Z(u), \qquad \tilde{u} = \frac{\int uZ(u)\,\mathrm{d}u}{\int Z(u)\,\mathrm{d}u}$$

We (Chigansky, K., 2011) prove that BE are as. efficient and

$$n\left(\tilde{\vartheta}_n - \vartheta\right) \Longrightarrow \tilde{u}, \qquad \mathbf{E}_\vartheta \left|n\left(\tilde{\vartheta}_n - \vartheta\right)\right|^p \longrightarrow \mathbf{E}_\vartheta \left|\tilde{u}\right|^p$$

Threshold Diffusion Processes

Consider

$$\mathrm{d}X_t = \sum_{j=1}^{k+1} S_j\left(X_t\right)\, 1\!\!\mathrm{I}_{\{\vartheta_{j-1} < X_t \leq \vartheta_j\}}\mathrm{d}t + \sigma\left(X_t\right)\mathrm{d}W_t,$$

where $\vartheta_0 = -\infty$, $\vartheta_j \in \Theta_j = \left(\alpha_j, \beta_j\right), j = 1, \ldots, k$, $\vartheta_{k+1} = \infty$, $\beta_j < \alpha_{j+1}$. The functions $S_j(x)$ and $\sigma\left(x\right)$ are such that the process X_t is ergodic with invariant density $f\left(\boldsymbol{\vartheta}, x\right)$.

Problem: *how to estimate* $\boldsymbol{\vartheta}$ *by observations* $X^T = \left(X_t, 0 \leq t \leq T\right)$ *and what are the properties of estimators as* $T \rightarrow \infty$*?*

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 $\rho_i > 0$. 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 ergodic diffusion with the invariant density

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

The log-likelihood ratio function is

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

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

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

Introduce the normalized likelihood ratio

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

This LR $Z_T(u)$ 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 the function

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

We need two r.v.'s : $\hat{u}_\vartheta$ and $\tilde{u}_\vartheta$ are defined by

$$Z(\hat{u}_\vartheta) = \sup_u Z(u), \qquad\qquad \tilde{u}_\vartheta = \frac{\int_R u\, Z(u)\, \mathrm{d}u}{\int_R Z(u)\, \mathrm{d}u}.$$

Properties of estimators

The MLE and BE are

- *Uniformly consistent*
- *Have different limit distributions*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\varliminf_{\delta\to0}\varliminf_{T\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta} T^2\,\mathbf{E}_\vartheta\left(\bar\vartheta_T-\vartheta\right)^2 \geq \mathbf{E}\tilde u_{\vartheta_0}^2$$

and for bayesian estimators we have

$$\lim_{\delta\to0}\lim_{T\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta} T^2\,\mathbf{E}_\vartheta\left(\tilde\vartheta_T-\vartheta\right)^2 = \mathbf{E}\tilde u_{\vartheta_0}^2.$$

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

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

$$\mathbf{E}\hat u_\vartheta^2=\frac{26}{\Gamma_\vartheta^2}\quad > \quad \mathbf{E}\tilde u_\vartheta^2=\frac{16\zeta\left(3\right)}{\Gamma_\vartheta^2}$$

where $\zeta\left(3\right)$ is Riemann zeta function and $16\zeta\left(3\right)\sim 19,3$

Misspecification

Suppose that the observed process is

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

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 $\rho_2 > \rho_1$. If the function $h\left(\cdot\right)$ satisfies the condition

$$\left|h\left(y\right)\right| < \frac{y}{2}\left(\rho_2 - \rho_1\right), \quad \alpha < y < \beta$$

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, are asymptotically independent and

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

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

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

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

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

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

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

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

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

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

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 $L\left(\vartheta, X^T_{\sqrt{T}}\right)$

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

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,

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

Nonlinear Multy Threshold Model. Suppose that

$$\mathrm{d}X_t = \sum_{j=1}^{k+1} S_j\left(X_t\right)\ 1\!\!\mathrm{I}_{\left\{\vartheta_{j-1}<X_t\leq\vartheta_j\right\}}\mathrm{d}t + \sigma\left(X_t\right)\mathrm{d}W_t,$$

where $\vartheta_0 = -\infty$, $\vartheta_j \in \Theta_j = \left(\alpha_j, \beta_j\right), j = 1, \ldots, k$, $\vartheta_{k+1} = \infty$, $\beta_j < \alpha_{j+1}$. The functions $S_j(x)$ and $\sigma\left(x\right)$ are such that the process X_t is ergodic with invariant density $f\left(\boldsymbol{\vartheta}, x\right)$. Then under the "natural" conditions the LR

$$Z_T\left(u\right) = \frac{L\left(\boldsymbol{\vartheta} + \frac{\mathbf{u}}{T}, X^T\right)}{L\left(\boldsymbol{\vartheta}, X^T\right)} \Longrightarrow Z\left(\mathbf{u}\right)$$

where

$$Z\left(\mathbf{u}\right) = \exp\left\{\sum_{j=1}^{k}\left[\gamma_j\left(\boldsymbol{\vartheta}\right)\ W_j\left(u_j\right) - \frac{\left|u_j\right|}{2}\ \gamma_j\left(\boldsymbol{\vartheta}\right)^2\right]\right\},$$

$$\gamma_j\left(\boldsymbol{\vartheta}\right)^2 = \left(S_{j+1}\left(\vartheta_j\right) - S_j\left(\vartheta_j\right)\right)^2\sigma\left(\vartheta_j\right)^{-2}\ f\left(\boldsymbol{\vartheta}, \vartheta_j\right)$$

the MLE $\widehat{\boldsymbol{\vartheta}}_T$ and bayessian estimator $\widetilde{\boldsymbol{\vartheta}}_T$ are consistent, have the following limit distributions:

$$T\left(\widehat{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}\right)\Longrightarrow \hat{\boldsymbol{u}}_{\boldsymbol{\vartheta}},\qquad\qquad T\left(\widetilde{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}\right)\Longrightarrow \tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}},$$

where $\tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}=(\hat{u}_1,\ldots,\hat{u}_k)$ with independent components and the moments converge : for any $p>0$

$$\lim_{T\to\infty} T^p\,\mathbf{E}_\vartheta\left|\widehat{\boldsymbol{\vartheta}}_T-\vartheta\right|^p=\mathbf{E}\left|\hat{\boldsymbol{u}}_{\boldsymbol{\vartheta}}\right|^p,\qquad \lim_{T\to\infty} T^p\,\mathbf{E}_\vartheta\left|\tilde{\vartheta}_T-\vartheta\right|^p=\mathbf{E}\left|\tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}\right|^p.$$

The BE are asymptotically efficient.

References

[1] Chan, K.S. (1993). Consistency and limiting distribution of the LSE of a TAR, *Ann. Statist.* **21**, 520–533.

[2] Chan, N.H., Kutoyants Yu. A., (2012) *On parameter estimations of threshold autoregressive models*, Statist. Infer. Stoch. Processes, 48, 1, 81-104.

[3] Chan, N.H., Kutoyants Yu. A., (2010) *Recent developments of threshold estimation for nonlinear time series*, J. Japan Statist. Soc., 40, 2, 277-308.

[4] Chigansky, P. Kutoyants, Yu.A. (2012) *On nonlinear TAR processes and threshold estimation*, submitted.

[5] Chigansky, P. Kutoyants, Yu.A. (2012) *Estimation of the threshold autoregressive models with correlated innovations* , submitted.

[6] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

[7] Hansen, B.E. (2000). Sample splitting and threshold estimation, *Econometrica* **68**, 575–603.

[8] Kutoyants, Yu. A. (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer, London.

[9] Kutoyants, Yu. A. (2012). On identification of the threshold diffusion processes. *Annals Inst. Statist. Math.*, 64, 2, 383-413.

[10] Tong, H. (1990). *Non-linear Time Series: A Dynamical Systems Approach.* Oxford University Press, Oxford.