Prediction of time series and nonstationary times, Maison des Sciences Economiques, Paris, February 10-11, 2012

Conditional Least Squares Estimation in nonlinear and nonstationary stochastic regression models

Christine Jacob
Applied Mathematics and Informatic unit, INRA, Jouy-en-Josas, France
christine.jacob@jouy.inra.fr

Model

- $\{Z_n\}$: real stochastic process, may depend on an external process $\{U_n\}$,

$$
\begin{aligned}
E(Z_n|F_{n-1}) \; &= \; \underbrace{g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})}_{\text{Model, Lipschitz in } \theta} = \underbrace{g^{(1)}(\boldsymbol{\theta}_0, F_{n-1})}_{\text{Approximate model}} + \underbrace{g^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})}_{\text{nuisance "negligible" part}} \overset{a.s.}{<} \infty \\
Var(Z_n|F_{n-1}) \; &= \; \sigma^2(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) \overset{a.s.}{<} \infty,
\end{aligned}
$$

where $F_{n-1} := \{Z_{n-1}, Z_{n-2}, \ldots, U_n, U_{n-1}, \ldots\}$: observations

$\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0 \subset \mathbb{R}^p$, $p < \infty$: parameter of interest
$\boldsymbol{\nu}_0 \in \mathbb{R}^q$, $q \leq \infty$: nuisance parameter
If $q = \infty$, $\nu_n := g^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})$, if $q = 0$, $g^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) = 0$

Goal

- Estimate $\boldsymbol{\theta}_0$ from one observed trajectory of $\{Z_n, U_n\}$ by Conditional Least Squares:

$$
\begin{aligned}
Z_n &= g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) + e_n,\ \ E(e_n|F_{n-1}) \stackrel{a.s.}{=} 0, E(e_n^2|F_{n-1}) := \sigma^2(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})) \\
\widehat{\boldsymbol{\theta}}_n &:= \arg\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\nu}}),\ \ S_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\nu}}) := \sum_{k=1}^{n} (Z_k - g(\boldsymbol{\theta}, \widehat{\boldsymbol{\nu}}, F_{k-1}))^2 \lambda(F_{k-1})
\end{aligned}
$$

$$
\text{or, if } q < \infty,\ (\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\nu}}_n) := \arg\min_{\boldsymbol{\theta},\boldsymbol{\nu}} S_n(\boldsymbol{\theta}, \boldsymbol{\nu}),\ \ S_n(\boldsymbol{\theta}, \boldsymbol{\nu}) := \sum_{k=1}^{n} (Z_k - g(\boldsymbol{\theta}, \boldsymbol{\nu}, F_{k-1}))^2 \lambda(F_{k-1})
$$

- Asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$, as $n \to \infty$: strong consistency, asymptotic distribution

Difficulties: stochasticity, nonstationarity, nonlinearity, no explicit expression of $\widehat{\boldsymbol{\theta}}_n$!

Literature: for the strong consistency, existence of sufficient conditions forbidding many types of nonstationarity, asymptotic distribution in very particular cases

- Optimal properties of $\widehat{\boldsymbol{\theta}}_n$ for a finite n

For $q = 0$ (no nuisance parameter), optimality if $\lambda(F_{k-1}) \propto \sigma^{-2}(\boldsymbol{\theta}_0, F_{k-1})$ (Heyde, 1997):

Ex. $p = 1 \Longrightarrow$ optimality if $\dot{S}_n(\boldsymbol{\theta}_0)$ minimum with $\ddot{S}_n(\boldsymbol{\theta}_0)$ maximum

$$\Longrightarrow \ \Big(E(\dot{S}_n^2(\boldsymbol{\theta}_0))\Big)\Big(E(\ddot{S}_n(\boldsymbol{\theta}_0))\Big)^{-2} \text{ minimum for } \lambda(F_{k-1}) \propto \sigma^{-2}(\boldsymbol{\theta}_0, F_{k-1})$$

$\Longrightarrow$ Take $\lambda(F_{k-1}) \propto \widehat{\sigma}^{-2}(\boldsymbol{\theta}_0, F_{k-1})$

Note: If $\lambda(F_{k-1})$ depends on $\boldsymbol{\theta}$, then the consistency of $\widehat{\boldsymbol{\theta}}_n$ is generally not ensured

Usefulness of $g^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) := \underbrace{g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})}_{E(Z_n|F_{n-1})} - \underbrace{g^{(1)}(\boldsymbol{\theta}_0, F_{n-1})}_{\textbf{Approximate model}}$

Assumption: $g^{(2)}(\boldsymbol{\theta}, \widehat{\boldsymbol{\nu}}, F_{n-1})$ is A.N. (Asymptotically Negligible):

$$\forall \delta > 0, \quad \overline{\lim}_n \frac{\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} |g^{(2)}(\boldsymbol{\theta}, \widehat{\boldsymbol{\nu}}, F_{n-1}) - g^{(2)}(\boldsymbol{\theta}_0, \nu_0, F_{n-1})|}{\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} |g^{(1)}(\boldsymbol{\theta}, F_{n-1}) - g^{(1)}(\boldsymbol{\theta}_0, F_{n-1})|} \stackrel{a.s.}{=} 0$$

• $g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})$ is more realistic than the simpler model $g^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})$,
but $\widehat{\nu}_n$ may have no asymptotic properties $\Longrightarrow$ study $\widehat{\theta}_n$

• Z_n is observed with an observation error η_n, $E(\eta_n|F_{n-1}) = 0$
$\Longrightarrow Z_n^\eta = Z_n + \eta_n$, $\Longrightarrow E(Z_n^\eta|F_{n-1}) = E(Z_n|F_{n-1}) = g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})$

$$\begin{aligned} g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) &= g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F^\eta_{n-1}) + g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) - g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F^\eta_{n-1}) \\ &= \underbrace{g^{(1)}(\boldsymbol{\theta}_0, F^\eta_{n-1})}_{g^{\eta(1)}(\boldsymbol{\theta}_0, F^\eta_{n-1})} + \underbrace{g^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F^\eta_{n-1}) + g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) - g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F^\eta_{n-1})}_{g^{\eta(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}, F^\eta_{n-1}) =: \boldsymbol{\nu}_n} \end{aligned}$$

- Z_n depends on the past errors $e_{n-1}, e_{n-2}, \ldots$

$$
\begin{aligned}
Ex.\ ARMA(p,q):\ Z_n &= \sum_{j=1}^{p} \alpha_j Z_{n-j} + \sum_{j=1}^{q} \beta_j e_{n-j} + e_n \Longleftrightarrow A(L)Z_n = B(L)e_n, \\
A(L) &:= 1 - \sum_{j=1}^{p} \alpha_j L^j,\ B(L) := 1 + \sum_{j=1}^{q} \beta_j L^j,\ LZ_n = Z_{n-1} \\
\Longrightarrow Z_n &= \sum_{j=1}^{p} \alpha_j Z_{n-j} + (B(L)-1)L^{-1} \underbrace{e_{n-1}}_{(B(L])^{-1}A(L)Z_{n-1}} + e_n \\
&= \underbrace{\sum_{j=0}^{n-1} \gamma_{n-j} Z_j}_{\text{Approximate model } g^{(1)}(.)} + \underbrace{\sum_{j=-1}^{-n_0} \gamma_{n-j} Z_j}_{\text{Nuisance part } g^{(2)}(.)} + e_n
\end{aligned}
$$

where γ_l is function of $\{\alpha_j\}$ and $\{\beta_j\}$ (for $p = 1$, $q = 1$, $\gamma_j = (-1)^{j-1}(\alpha_1 + \beta_1)\beta_1^{j-1}$),
$Z_n = 0$, $n < -n_0 \leq 0$

Extension: $S_n(\boldsymbol{\theta},\boldsymbol{\nu}) := \sum_{k=1}^{n}\Big(\Psi_k(Z_k - g(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))\Big)^2\lambda(F_{k-1})$,
$E(\Psi_k(Z_k - g(\boldsymbol{\theta}_0,\boldsymbol{\nu}_0,F_{k-1}))|F_{k-1}) = 0$

- Missing data

$\Longrightarrow \Psi_k(Z_k - g(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1})) = 1_{\{Z_k\in I_k^{obs}\}}(Z_k - g(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))$

- $E(Z_n|F_{n-1}) \leq \infty$, $E(Z_n^2|F_{n-1}) \leq \infty$

$\Longrightarrow \Psi_k(Z_k - g(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1})) = 1_{\{Z_k\in I_k\}}(Z_k - g(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))$, $\lim_k I_k \overset{a.s.}{=} \infty$

- Existence of outliers (Heyde, 1997)

$\Longrightarrow \lambda(F_{k-1}) = \sum_{h=1}^{k}\lambda_{h,k}\Psi_{k-h}(Z_{k-h} - g(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-h-1}))$,
$\Psi_k(x) = \Psi(x) = x$, if $|x| < m$, $\Psi(x) = m\,sign(x)$, if $|x| \geq m$ (Huber's function)

Extension: $S_n(\boldsymbol{\theta},\boldsymbol{\nu}) := \sum_{k=1}^{n}\Big(\Psi_k(Z_k - g(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))\Big)^2\lambda(F_{k-1}) + pen(\boldsymbol{\theta},p,F_{n-1})$**,** $E(\Psi_k(Z_k - g(\boldsymbol{\theta}_0,\boldsymbol{\nu}_0,F_{k-1}))|F_{k-1}) = 0$**, where** p **is the nber of** $\theta_i \neq 0$

$\Longrightarrow$ Penalize large values of p

Extension: Multivariate stochastic regression model:
$S_n(\boldsymbol{\theta},\boldsymbol{\nu}) = \sum_{k=1}^{n}(\mathbf{Z}_k - \mathbf{g}(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{Z}_k - \mathbf{g}(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))$**,** $\boldsymbol{\Sigma}_k^{-1}$ F_{k-1}**-measurable**

$$
\begin{aligned}
\Longrightarrow S_n(\boldsymbol{\theta},\boldsymbol{\nu}) &= \sum_{k=1}^{n}(\mathbf{Z}_k - \mathbf{g}(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))^T\mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{U}_k^{-1}(\mathbf{Z}_k - \mathbf{g}(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1})),\ \ \mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{U}_k^{-1} = \boldsymbol{\Sigma}_k^{-1},\ \ \boldsymbol{\Sigma}_k^{-1}\,F_{k-1} \\
&= \sum_{k=1}^{n}\Big(\boldsymbol{\Lambda}_k^{1/2}\mathbf{U}_k^{-1}(\mathbf{Z}_k - \mathbf{g}(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))\Big)^T\underbrace{\Big(\boldsymbol{\Lambda}_k^{1/2}\mathbf{U}_k^{-1}(\mathbf{Z}_k - \mathbf{g}(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))\Big)}_{\text{denoted by } \mathbf{Y}_k - \mathbf{f}_k(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1})} \\
&= \sum_{j=1}^{d}\sum_{k=1}^{n}(Y_{k,j} - f_{k,j}(\boldsymbol{\theta},\boldsymbol{\nu},F_{k-1}))^2,
\end{aligned}
$$

Examples of models

- Classical Regression with stochastic regressors
- Time series: $ARMAX(p, q, b)$ model, nonlinear time series
- Financial models: $GARCH(p, q)$ model, nonlinear financial models

$$\xi_n = s_n(\boldsymbol{\theta}_0)U_n, \ \{U_n\} \text{ i.i.d. } (0,1) \Longrightarrow Z_n := \xi_n^2 = s_n^2(\boldsymbol{\theta}_0) + \underbrace{s_n^2(\boldsymbol{\theta}_0)(U_n^2 - 1)}_{e_n}$$

$$s_n^2(\boldsymbol{\theta}) = \alpha_0 + \sum_{j=1}^{p} \beta_j s_{n-j}^2(\boldsymbol{\theta}) + \sum_{j=1}^{q} \alpha_j \xi_{n-j}^2 \ (volatility),$$

- Branching processes

$$N_n = \sum_{i=1}^{N_{n-1}} X_{n,i}, \ \{X_{n,i}\}_i | F_{n-1} \text{ i.i.d. } (m_{\boldsymbol{\theta}_0,\boldsymbol{\nu}_0}(F_{n-1}), \sigma^2_{\boldsymbol{\theta}_0,\boldsymbol{\nu}_0}(F_{n-1}))$$

$$\Longrightarrow Z_n := N_n = m_{\boldsymbol{\theta}_0,\boldsymbol{\nu}_0}(F_{n-1})N_{n-1} + \sqrt{N_{n-1}}e_n, \ E(e_n^2|F_{n-1}) = \sigma^2_{\boldsymbol{\theta}_0,\boldsymbol{\nu}_0}(F_{n-1})$$

Asymptotic properties for $q = 0$ (no nuisance parameter)

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}),\ \ S_n(\boldsymbol{\theta}) := \sum_{k=1}^{n} (Z_k - g(\boldsymbol{\theta}, F_{k-1}))^2 \lambda(F_{k-1}) =: \sum_{k=1}^{n} (Y_k - f(\boldsymbol{\theta}, F_{k-1}))^2$$

Let $e_k := Y_k - f(\boldsymbol{\theta}_0, F_{k-1})$, $\sigma_k^2 := Var(e_k | F_{k-1})$. Assume

$$A1:\ \ \forall \delta > 0,\ \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} \sum_{k=1}^{\infty} \frac{\sigma_k^2 d_k^2(\theta)}{\left(\sum_{h=1}^{k} d_h^2(\boldsymbol{\theta})\right)^2} \overset{a.s.}{<} \infty,\ \ d_k(\boldsymbol{\theta}) := f(\boldsymbol{\theta}, F_{k-1}) - f(\boldsymbol{\theta}_0, F_{k-1})$$

- Strong consistency (Jacob, 2010)

Let $D_n(\boldsymbol{\theta}) = \sum_{k=1}^{n} \Big(f(\boldsymbol{\theta}, F_{k-1}) - f(\boldsymbol{\theta}_0, F_{k-1}) \Big)^2$ (identifiability criterion). Assume $A1$. Then

$$A2: \;\; \forall \delta > 0, \;\; \underline{\lim}_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} D_n(\boldsymbol{\theta}) \overset{a.s.}{=} \infty \Longrightarrow \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s.}{=} \boldsymbol{\theta}_0$$

Linear case: $f(\boldsymbol{\theta}, F_{n-1}) = \theta_0^T \mathbf{W}_{n-1} \Longrightarrow \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = \Big(\sum_{k=1}^{n} \mathbf{W}_{k-1} \mathbf{W}_{k-1}^T \Big)^{-1} \sum_{k=1}^{n} e_k \mathbf{W}_{k-1}$,
$D_n(\boldsymbol{\theta}) = \Big(\boldsymbol{\theta} - \boldsymbol{\theta}_0 \Big)^T \mathbf{W}_{n-1} \mathbf{W}_{n-1}^T \Big(\boldsymbol{\theta} - \boldsymbol{\theta}_0 \Big)$

$$\Longrightarrow \underline{\lim}_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} D_n(\boldsymbol{\theta}) \overset{a.s.}{=} \infty \Longleftrightarrow \lim_n \lambda_{min} \Big(\sum_{k=1}^{n} \mathbf{W}_{k-1} \mathbf{W}_{k-1}^T \Big) \overset{a.s.}{=} \infty$$

Note. In the literature, existence of additional sufficient conditions:
$D_n(\boldsymbol{\theta})$ has to tend to ∞ at some *rate*
Ex. Linear case $f(\boldsymbol{\theta}, F_{n-1}) = \theta_0^T \mathbf{W}_{n-1}$, the additional condition is
$\lim_n [\ln(\lambda_{max}\{\sum_{k=1}^{n} \mathbf{W}_k \mathbf{W}_k^T\}]^{\rho}][\lambda_{min}\{\sum_{k=1}^{n} \mathbf{W}_k \mathbf{W}_k^T\}]^{-1} = 0$ (Lai & Wei, 1982)

- Proof

1. Wu's Lemma (1981):

$$\underline{\lim}_n \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} (S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0)) \overset{a.s.}{>} 0, \forall\delta > 0 \Longrightarrow \lim_n \widehat{\boldsymbol{\theta}}_n \overset{a.s.}{=} \boldsymbol{\theta}_0.$$

2. $S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) = D_n(\boldsymbol{\theta}) + 2L_n(\boldsymbol{\theta})$

$$\Longrightarrow \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) \;\geq\; \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} D_n(\boldsymbol{\theta})\Big(1 - 2 \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} \frac{|L_n(\boldsymbol{\theta})|}{D_n(\boldsymbol{\theta})}\Big)$$

$$\frac{L_n(\boldsymbol{\theta})}{D_n(\boldsymbol{\theta})} \;:=\; \frac{\sum_{k=1}^n e_k \Big(f(\boldsymbol{\theta}, F_{k-1}) - f(\boldsymbol{\theta}_0, F_{k-1})\Big)}{\sum_{k=1}^n \Big(f(\boldsymbol{\theta}, F_{k-1}) - f(\boldsymbol{\theta}_0, F_{k-1})\Big)^2}$$

3. Prove that $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} |L_n(\boldsymbol{\theta})| \Big(D_n(\boldsymbol{\theta})\Big)^{-1}$ converges a.s. to 0

Linear case $f(\boldsymbol{\theta}, F_{n-1}) = \theta_0^T \mathbf{W}_{n-1}$

$$\Longrightarrow \frac{|L_n(\boldsymbol{\theta})|}{D_n(\boldsymbol{\theta})} = \frac{|(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \sum_k e_k \mathbf{W}_{k-1}|}{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \sum_{k=1}^n \mathbf{W}_{k-1}\mathbf{W}_{k-1}^T(\boldsymbol{\theta}-\boldsymbol{\theta}_0)}$$

$$\Longrightarrow \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} \frac{|L_n(\boldsymbol{\theta})|}{D_n(\boldsymbol{\theta})} = \frac{|L_n(\boldsymbol{\theta}_n)|}{D_n(\boldsymbol{\theta}_n)},\ \boldsymbol{\theta}_n \text{ depends on } e_n, F_{n-1}$$

$$p = 1 \Longrightarrow \frac{|L_n(\theta_n)|}{D_n(\theta_n)} \leq |(\theta_n - \theta_0)^{-1}|| \frac{\sum_k e_k W_{k-1}}{\sum_{k=1}^n W_{k-1}^2}| \Longrightarrow \text{SLLNM (Hall \& Heyde, 1980)}$$

General nonlinear case with $p \geq 1$? $\sum_{k=1}^n e_k \frac{\left(f(\boldsymbol{\theta}_n, F_{k-1}) - f(\boldsymbol{\theta}_0, F_{k-1})\right)}{\sum_{h=1}^k \left(f(\boldsymbol{\theta}_n, F_{h-1}) - f(\boldsymbol{\theta}_0, F_{h-1})\right)^2}$ is not a martingale!

Strong Law of Large Numbers for SubMartingales (Jacob, 2010)

$d_k(\boldsymbol{\theta}) := f(\boldsymbol{\theta}, F_{k-1}) - f(\boldsymbol{\theta}_0, F_{k-1})$ F_{k-1}-measurable and Lipschitz in $\boldsymbol{\theta}$,
$E(e_k|F_{k-1}) = 0$, $E(e_k^2|F_{k-1}) =: \sigma_k^2$

$$\lim_n \inf_{\boldsymbol{\theta}} \sum_{k=1}^{n} d_k^2(\boldsymbol{\theta}) \overset{a.s;}{=} \infty \text{ and } \sup_{\boldsymbol{\theta}} \sum_{k=1}^{\infty} \frac{\sigma_k^2 d_k^2(\theta)}{\Big(\sum_{h=1}^{k} d_h^2(\boldsymbol{\theta})\Big)^2} \overset{a.s.}{<} \infty \implies \lim_n \sup_{\boldsymbol{\theta}} \Big|\frac{\sum_{k=1}^{n} e_k d_k(\boldsymbol{\theta})}{\sum_{k=1}^{n} d_k^2(\boldsymbol{\theta})}\Big| \overset{a.s.}{=} 0$$

Proof: $\sup_{\boldsymbol{\theta}} |\sum_{k=1}^{n} e_k d_k(\boldsymbol{\theta}) \Big(\sum_{h=1}^{k} d_h^2(\boldsymbol{\theta})\Big)^{-1}|$ is a submartingale
$\Longrightarrow$ use submartingale properties (Hall & Heyde, 1980), and analytical lemma's

Note. SLLNM: $\lim_n \frac{\sum_{k=1}^{n} e_k d_k(\boldsymbol{\theta})}{\sum_{k=1}^{n} d_k^2(\boldsymbol{\theta})} \overset{a.s.}{=} 0$

Markov's theorem: $\lim_n \sum_{k=1}^{n} \frac{\sigma_k^2 d_k^2(\theta)}{n^2} \overset{a.s.}{=} 0 \Longrightarrow \lim_n \frac{\sum_{k=1}^{n} e_k d_k(\boldsymbol{\theta})}{n} \overset{P}{=} 0$

- Asymptotic distribution (Jacob, 2010)

1. Taylor's expansion of $\dot{\mathrm{S}}_n(\widehat{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}_0$

$$\Longrightarrow \dot{\mathrm{S}}_n(\widehat{\boldsymbol{\theta}}_n) = \dot{\mathrm{S}}_n(\boldsymbol{\theta}_0) + \ddot{\mathrm{S}}_n(\boldsymbol{\theta}_n)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \Longrightarrow (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\Big(\ddot{\boldsymbol{S}}_n(\boldsymbol{\theta}_n)\Big)^{-1}\dot{\mathrm{S}}_n(\boldsymbol{\theta}_0)$$

2. Find a $p \times p$ deterministic matrix $\boldsymbol{\Phi}_n$ such that

$$\boldsymbol{\Phi}_n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\boldsymbol{\Phi}_n^{-1/2}\Big(\ddot{\mathrm{S}}_n(\boldsymbol{\theta}_n)\boldsymbol{\Phi}_n^{-1}\Big)^{-1}\dot{\mathrm{S}}_n(\boldsymbol{\theta}_0)$$

with $\lim_n \ddot{\mathrm{S}}_n(\boldsymbol{\theta}_n)\boldsymbol{\Phi}_n^{-1} \overset{P}{=} \mathbf{I}$, and $\lim_n \boldsymbol{\Phi}_n^{-1/2}\dot{\mathrm{S}}_n(\boldsymbol{\theta}_0)$ exists in distribution

$$\dot{\boldsymbol{S}}_n(\boldsymbol{\theta}_0) \;=\; -2\sum_{k=1}^{n} e_k \dot{\mathbf{f}}(\boldsymbol{\theta}_0, F_{k-1}) \Longrightarrow \boldsymbol{\Phi}_n^{-1/2} = O\Big(\sum_{k=1}^{n} \dot{\mathbf{f}}(\boldsymbol{\theta}_0, F_{k-1})\dot{\mathbf{f}}^T(\boldsymbol{\theta}_0, F_{k-1})\Big)^{-1/2}$$

$$\ddot{\mathrm{S}}_n(\boldsymbol{\theta}_n) \;=\; \underbrace{2\sum_{k=1}^{n} \dot{\mathbf{f}}(\boldsymbol{\theta}_n, F_{k-1})\dot{\mathbf{f}}^T(\boldsymbol{\theta}_n, F_{k-1})}_{\sim \boldsymbol{\Phi}_n} - \underbrace{2\sum_{k=1}^{n} e_k \ddot{\mathbf{f}}(\boldsymbol{\theta}_n, F_{k-1})}_{\neq 0 \text{ in the nonlinear case}}$$

Use the SLLNSM for proving that $\lim_n \sum_{k=1}^{n} e_k \ddot{\mathbf{f}}(\boldsymbol{\theta}_n, F_{k-1})\boldsymbol{\Phi}_n^{-1} \overset{a.s.}{=} 0$

Examples

- Polymerase Chain Reaction: replication *in vitro* of a population of N_0 DNA

(Lalam, Jacob & Jagers, 2004)

$$N_n = \sum_{i=1}^{N_{n-1}} (1 + X_{n,i}), \;\; \{X_{n,i}\}_i | F_{n-1} \text{ i.i.d. } Ber(p_{\boldsymbol{\theta}_0}(N_{n-1}))$$

$$p_{\boldsymbol{\theta}_0}(N_{n-1}) := P(X_{n,i} = 1 | N_{n-1}) = \frac{K_0}{K_0 + N_{S_0,n-1}} \frac{\Big(1 + \exp(-C_0(S_0^{-1} N_{S_0,n-1} - 1))\Big)}{2}$$

where $N_{S_0,n-1} = S_0$, if $N_{n-1} < S_0$, and $N_{S_0,n-1} = N_{n-1}$, if $N_{n-1} \geq S_0$

N_n increases exponentially when $N_{n-1} < S_0$ (BGW branching process),
N_n increases linearly ($\lim_n N_n n^{-1} \overset{a.s.}{=} K_0/2$) when $N_{n-1} \geq S_0$

$$
\begin{aligned}
Z_n = N_n + \eta_n \quad &= \quad \left(1 + \frac{K_0}{K_0 + N_{S_0,n-1}} \frac{\Big(1 + \exp(-C_0(S_0^{-1} N_{S_0,n-1} - 1))\Big)}{2}\right) N_{n-1} + e_n + \eta_n \\
&\overset{n\ large}{=} \quad \underbrace{Z_{n-1} + \frac{K_0 Z_{n-1}}{2(K_0 + Z_{n-1})}}_{g^{(1)}(\boldsymbol{\theta}_0, Z_{n-1})} + \underbrace{O\Big(\exp(-C_0(S_0^{-1} Z_{n-1} - 1))\Big)}_{g^{(2)}(\boldsymbol{\theta}_0, Z_{n-1}, \eta_n)} + e_n
\end{aligned}
$$

Non asymptotic identifiability of (K, C, S) because of C, S

Strong identifiability of K given $(\widehat{C}_n, \widehat{S}_n)$

$\Longrightarrow \lim_n \widehat{K}_n | (\widehat{C}_n, \widehat{S}_n) \overset{a.s.}{=} K_0$, $\Phi_n^{1/2} (\widehat{K}_n - K_0) | (\widehat{C}_n, \widehat{S}_n) \overset{D}{=} \mathcal{N}(0, K/2)$

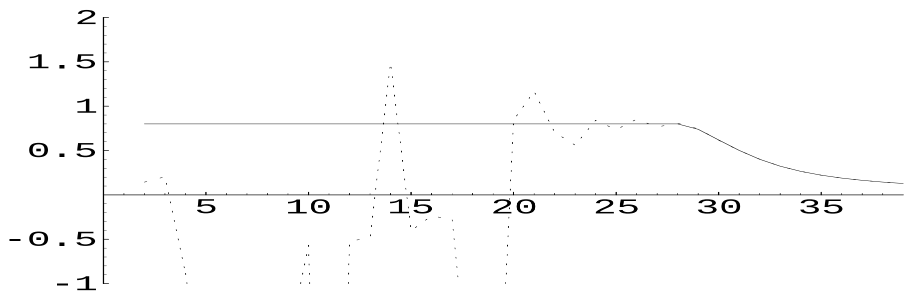

Figure 1: Efficiency $\{p_{\boldsymbol{\theta}_0}(N_{k-1})\}_{k\leq n}$ calculated from a simulated trajectory of the branching process ($K_0 = 4.00311.10^{10}$, $S_0 = 10^{10}$, $C_0 = 0$)
In dashed line: $\overline{p}(Z_{k-1}) = Z_k Z_{k-1}^{-1} - 1$, $k \leq n$ (empirical efficiency), in continuous line: $p_{\widehat{\boldsymbol{\theta}}_n}(Z_{k-1})$, $k \leq n$ (estimated efficiency)

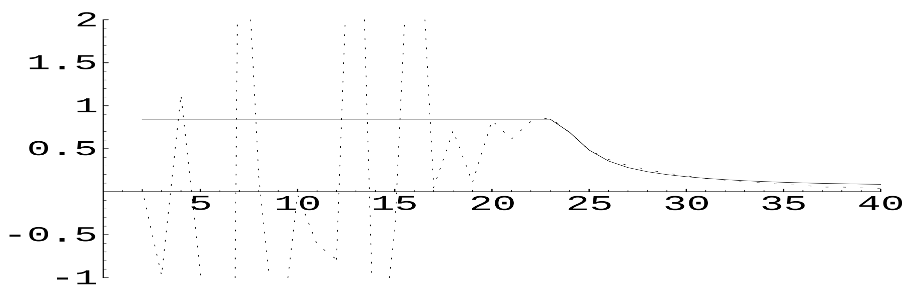

Figure 2: Real-time PCR, well 21 of data set 1, efficiency $\{p_{\boldsymbol{\theta}_0}(N_{k-1})\}_{k\leq n}$. In dashed line: $\{\overline{p}(Z_{k-1}) = Z_k Z_{k-1}^{-1} - 1\}_{k\leq n}$ (empirical efficiency), in continuous line: $\{p_{\widehat{\boldsymbol{\theta}}_n}(Z_{k-1})\}_{k\leq n}$ (estimated efficiency) with $\widehat{n}_s = 23$ (saturation threshold cycle), $\widehat{K}_{h,n} = 0.38055$, $\widehat{S}_{h,n} = 0.070553$, $\widehat{C}_{h,n} = 0.6$

PCR: another model (Jacob, 2010)

$$
\begin{aligned}
N_n &= \sum_{i=1}^{N_{n-1}} (1 + X_{n,i}) \\
P(X_{n,i} = 2 | N_{n-1}) &= \left(\frac{K_0}{K_0 + N_{S_0,n-1}}\right) \frac{(1 + S_0^{\alpha_0} N_{S_0,n-1}^{-\alpha_0})}{2}, \ \alpha > 0 \\
&\overset{n\ large}{=} Z_{n-1} + \frac{K_0 Z_{n-1}}{2(K_0 + Z_{n-1})} + \frac{K_0 S_0^{\alpha_0} Z_{n-1}^{1-\alpha_0}}{2(K_0 + Z_{n-1})} + O\Big(\eta_n\Big) + e_n
\end{aligned}
$$

(K, S, α) non asymptotically identifiable due to $\alpha \Longrightarrow$ assume α_0 known with $0 \leq 2\alpha_0 \leq 1$
Let $\boldsymbol{\theta} = (K, S^{\alpha_0})$. Then

$$
\lim_n \mathbf{\Phi}_n^{1/2} (\widehat{\theta}_n - \theta_0) \overset{d}{=} \mathcal{N}(0, (K/2)\mathbf{I}), \ \mathbf{\Phi}_n = \frac{1}{4} \begin{pmatrix} n & K n^{1-\alpha_0} \\ K n^{1-\alpha_0} & K^2 a_n(\alpha_0) \end{pmatrix}
$$

- $GARCH(1,1)$

$$
\begin{aligned}
Z_n &:= \xi_n^2 = \underbrace{s_n^2(\boldsymbol{\theta}_0)}_{g(\boldsymbol{\theta}_0,\boldsymbol{\nu}_0,F_{n-1})} + \underbrace{s_n^2(\boldsymbol{\theta}_0)(U_n^2-1)}_{e_n}, \; E(U_n^2|F_{n-1}) = 1 \\
s_n^2(\boldsymbol{\theta}) &= \alpha_0 + \alpha_1 \xi_{n-1}^2 + \beta_1 s_{n-1}^2(\boldsymbol{\theta}) \qquad (1) \\
\Longrightarrow \; s_n^2(\boldsymbol{\theta}) &= \alpha_0 (\sum_{l=0}^{n-1} \beta_1^l) + \alpha_1 \sum_{l=0}^{n-1} \beta_1^{l-1} \xi_{n-1-l}^2) + \beta_1^n s_0^2 \\
&= \underbrace{\alpha_0 (\sum_{l=0}^{\infty} \beta_1^l) + \alpha_1 \sum_{l=0}^{n-1} \beta_1^{l-1} \xi_{n-1-l}^2}_{g^{(1)}(\boldsymbol{\theta},\boldsymbol{\nu},F_{n-1})} + \underbrace{\beta_1^n (s_0^2 - \alpha_0 \sum_{l=0}^{\infty} \beta_1^l)}_{g^{(2)}(\boldsymbol{\theta},\boldsymbol{\nu},F_{n-1}) =: \nu_n}
\end{aligned}
$$

A.N. of $g^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) = 0$, for $\beta_{10} < 1 \Longrightarrow$ take $\widehat{\nu}_n := \widehat{g^{(2)}}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) = 0$

(1) $\Longrightarrow \Longrightarrow E(s_n^2(\boldsymbol{\theta}_0))\gamma_0^{-n} = \alpha_{00} \sum_{k=1}^n \gamma_0^{-k} + s_0^2(\boldsymbol{\theta}_0)$, $\gamma_0 := \alpha_{10} + \beta_{10}$
$\Longrightarrow \gamma_0 < 1 \Longrightarrow \lim_n E(s_n^2(\boldsymbol{\theta}_0)) = \alpha_{00}(1-\gamma_0)^{-1}$
$\Longrightarrow \gamma_0 > 1 \Longrightarrow \lim_n s_n^2(\boldsymbol{\theta}_0)\gamma_0^{-n} \stackrel{a.s.}{=} W$, $E(W) < \infty$
$\Longrightarrow \gamma_0 = 1 \Longrightarrow E(s_n^2(\boldsymbol{\theta}_0)) = n\alpha_{00} + s_0^2(\boldsymbol{\theta}_0)$

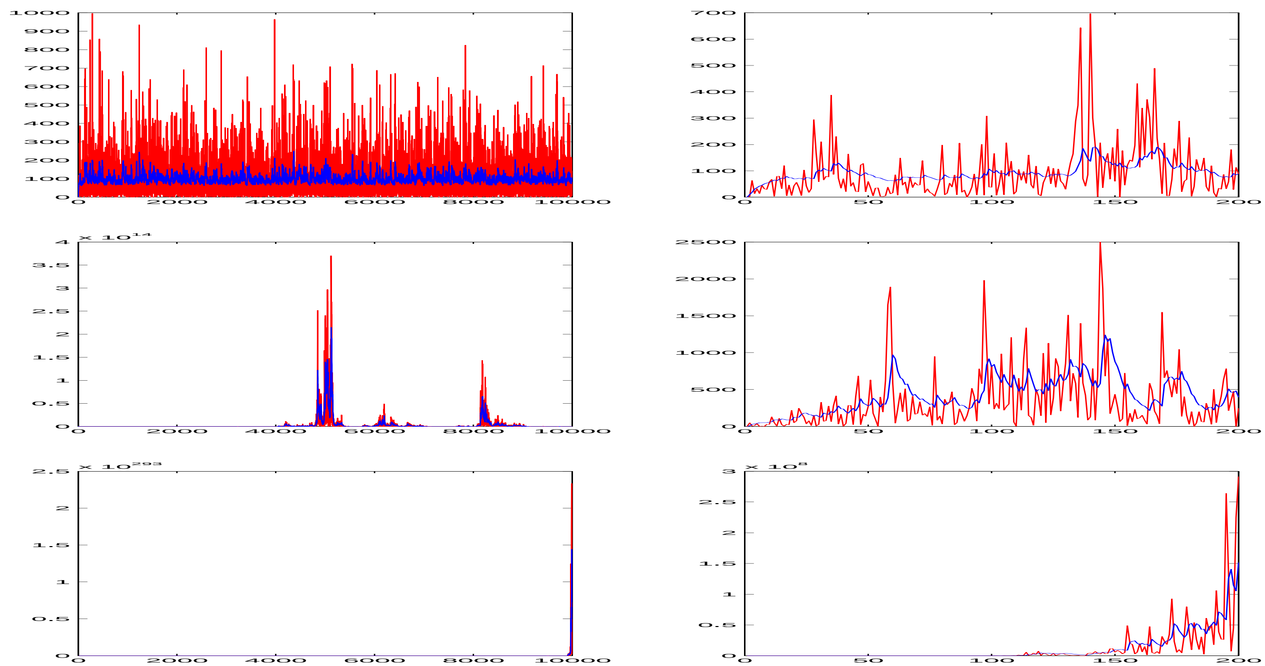

Figure 3: Simulations with $\{U_n^2\}$ i.i.d. $\exp(1)$. Red line: $\{\xi_n^2\}$, blue line: $\{s_n^2(\theta)\}$
On the first line, $\theta_0 = (\alpha_{00}, \alpha_{10}, \beta_{10}) = (10, 0.1, 0.8)$; on the second line, $\theta_0 = (10, 0.22, 0.8)$; on the third line, $\theta_0 = (10, 0.3, 0.8)$

Conditional Least Squares estimator of $\boldsymbol{\theta}_0 := (C_0, \alpha_{10}, \beta_{10})$, $\beta_{10} < 1$, $C_0 := \alpha_{00}(1-\beta_{10})^{-1}$

$$\begin{aligned}
Z_n &:= \xi_n^2 = \underbrace{s_n^2(\boldsymbol{\theta}_0)}_{g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})} + \underbrace{s_n^2(\boldsymbol{\theta}_0)(U_n^2 - 1)}_{e_n}, \;\; E(U_n^2|F_{n-1}) = 1 \\
&=: g(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1}) + e_n \\
&= \underbrace{C_0 + \alpha_{10} \sum_{l=0}^{n-1} \beta_{10}^l \xi_{n-1-l}^2}_{g^{(1)}(\boldsymbol{\theta}, F_{k-1})} + \underbrace{g^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0, F_{n-1})}_{\nu_n} + e_n, \;\; E(e_n^2|F_{n-1}) \propto s_n^4(\boldsymbol{\theta}_0) \\
\widehat{\boldsymbol{\theta}}_n | \widehat{\boldsymbol{\nu}} = \mathbf{0} &:= \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\theta}} S_n(\boldsymbol{\theta}, \mathbf{0}), \;\; S_n(\boldsymbol{\theta}, \mathbf{0}) := \sum_{k=1}^{n} (Z_k - g^{(1)}(\boldsymbol{\theta}, F_{k-1}))^2 \lambda(F_{k-1})
\end{aligned}$$

Optimality if $\lambda(F_{n-1}) \propto (Var(Z_n|F_{n-1}))^{-1} \propto s_n^{-4}(\boldsymbol{\theta}_0)$

$\Longrightarrow$ take $\lambda(F_{n-1}) = (g^{(1)}(\boldsymbol{\theta}_{*n}, F_{n-1}))^{-2} = \left(1 + \alpha_* \sum_{l=0}^{n-1} \beta_*^l \xi_{n-1-l}^2\right)^{-2}$, $\beta_* < 1$

Strong Consistency of $\widehat{\boldsymbol{\theta}}_n$ if $A1$ and $A2$ checked

- $A1$ checked for all $\alpha_* > 0$, $0 < \beta_* < 1$
- For $\boldsymbol{\theta} = (C, \alpha_{10}, \beta_{10})$, $A2 \Longleftrightarrow \sum_{k=1}^{\infty} \lambda(F_{k-1}) \overset{a.s.}{=} \infty$

$$\sum_{k=1}^{\infty} \lambda(F_{k-1}) = \sum_{k=1}^{\infty} \left(1 + \alpha_* \sum_{l=0}^{k-1} \beta_*^{k-1-l} \xi_l^2\right)^{-2} \geq \sum_{k=1}^{\infty} \left(1 + \alpha_*(1-\beta_*)^{-1} M_{k-1}^{\xi}\right)^{-2}, \; M_{k-1}^{\xi} := \sup_{l \leq k-1} \{\xi_l^2\}$$

$\Longrightarrow \sum_{k=1}^{\infty} \lambda(F_{k-1}) \geq \sum_m (L_{m+1} - L_m)\left(1 + \alpha_*(1-\beta_*)^{-1} \xi_{L_m}^2\right)^{-2}$, $\xi_{L_m}^2 := m^{th}$ record of $\{\xi_n^2\}$

$\Longrightarrow \sum_{k=1}^{\infty} \lambda(F_{k-1}) = \infty$ if $(L_{m+1} - L_m)(\xi_{L_m}^2)^{-2}$ does not tend to 0 too quickly ($\gamma_0 < 1$)

$\Longrightarrow$ for $\gamma_0 > 1$, $\sum_{k=1}^{\infty} \lambda(F_{k-1}) \simeq \sum_{k=1}^{\infty} \left(1 + \alpha_* W \gamma_0^{k-1} \sum_{l=0}^{k-1} (\beta_*/\gamma_0)^{k-1-l} U_l^2\right)^{-2} < \infty$, a.s.

- For $\boldsymbol{\theta} = (C_0, \alpha_1, \beta_{10})$ or $\boldsymbol{\theta} = (C_0, \alpha_{10}, \beta_1)$, $A2$ checked for all γ_0

Asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n$, $\boldsymbol{\theta} = (\alpha_1, \beta_1)$, $\gamma_0 > 1$:

$\lim_n \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathcal{N}(0, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ dependent on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_*$, $Var(\widehat{\alpha}_n)$ and $Var(\widehat{\beta}_n)$ minimum for $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$

Simulations of $\{\xi_n^2\}_{n=1}^N$, $\{U_n^2\}$ i.i.d. $exp(1)$, $s_1^2(\boldsymbol{\theta}_0) = 0$, calculus of $\widehat{\theta}_N$, for different values of $\boldsymbol{\theta}_0 = (\alpha_{00}, \alpha_{10}, \beta_{10})$ and N

For each value of $\boldsymbol{\theta}_0$ and of N, two graphics are given.

The first one represents $\xi_1^2, \dots, \xi_N^2$ (erratic line) with $s_1^2(\boldsymbol{\theta}_0), \dots, s_N^2(\boldsymbol{\theta}_0)$ ("smooth" line).

The second one represents $S_n(\boldsymbol{\theta}, 0)$ calculated with $\boldsymbol{\theta}_* = (1, 1, 0.999)$, $\boldsymbol{\theta} \in [\alpha_{00} - 0.1, \alpha_{00} + 0.05] \times [\alpha_{10} - 0.1, \alpha_{10} + 0.05] \times [\beta_{10} - 0.1, \beta_{10} + 0.05]$.

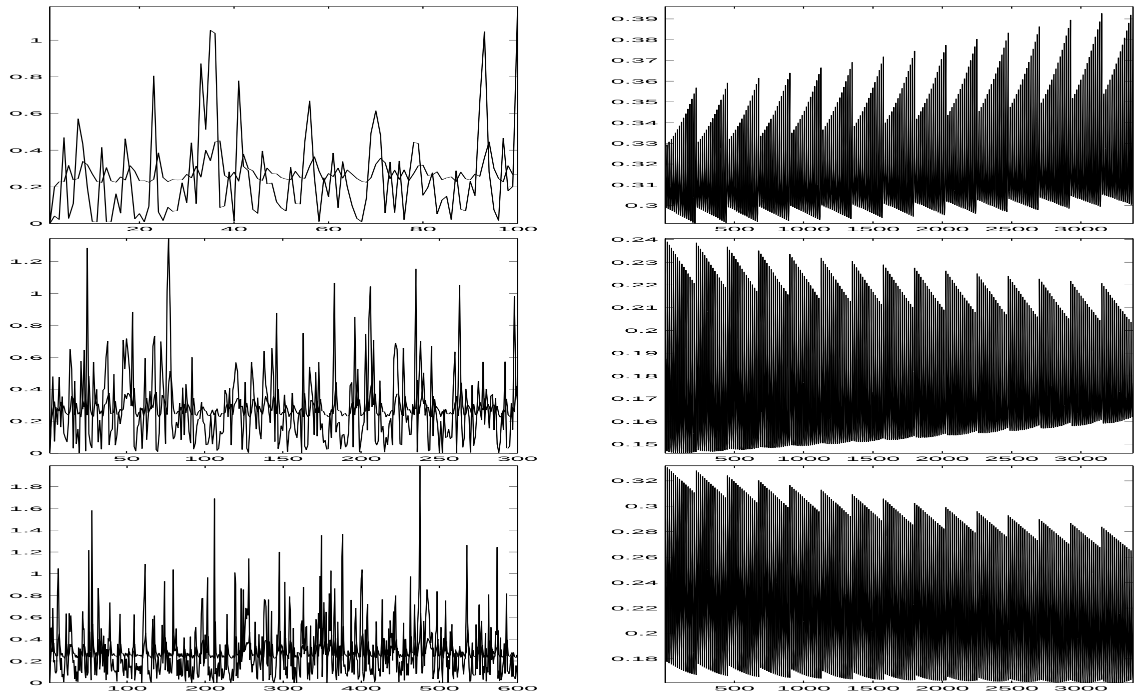

Figure 4: $\boldsymbol{\theta}_0 = (0.2, 0.2, 0.1)$.
1rst line, $N = 100$: $\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}) = S_n(0.16, 0.11, 0.15) = 0.2913$, and $S_n(\boldsymbol{\theta}_0) = 0.3129$.
2nd line, $N = 300$: $\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}) = S_n(0.24, 0.11, 0.09) = 0.1460$, and $S_n(\boldsymbol{\theta}_0) = 0.1564$.
3rd line, $N = 600$, $\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}) = S_n(0.25, 0.25, 0.07) = 0.1612$, and $S_n(\boldsymbol{\theta}_0) = 0.1889$.

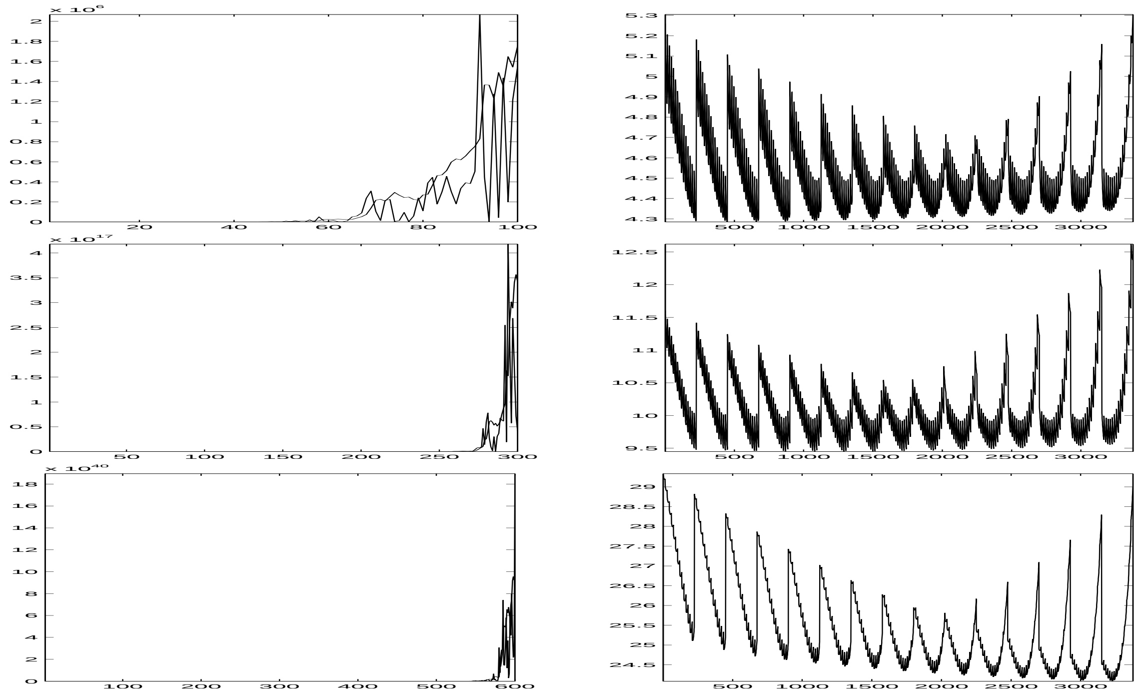

Figure 5: $\boldsymbol{\theta}_0 = (0.2, 0.3, 0.9)$.
1rst line, $N = 100$: $\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}) = S_n(0.23, 0.22, 0.95) = 4.2846$, and $S_n(\boldsymbol{\theta}_0) = 4.3297$
2nd line, $N = 300$: $\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}) = S_n(0.25, 0.25, 0.92) = 9.4485$, and $S_n(\boldsymbol{\theta}_0) = 9.6351$.
3rd line, $N = 600$, $\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}) = S_n(0.12, 0.35, 0.86) = 24.0845$, and $S_n(\boldsymbol{\theta}_0) = 24.3542$.

General comments:
For each value of $\boldsymbol{\theta}_0$ and of N, the minimum value of $S_n(\boldsymbol{\theta})$ is quite close to $S_n(\boldsymbol{\theta}_0)$
Assuming C_0 known or setting $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ does not improve significantly the results.

Conclusion

Indirect way of proof (Wu's Lemma) $+$ SLLNSM
$\Longrightarrow$ The difficulties (stochasticity, nonstationarity, nonlinearity, no explicit expression of $\widehat{\boldsymbol{\theta}}_n$) are removed

$\Longrightarrow$ Strong consistency, Asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n$

Thank you for your attention!

Main Reference:

JACOB, C. (2010) Conditional Least Squares Estimation in nonstationary nonlinear stochastic regression models.
Ann. Statist., **38(1)**, 566–597.