Testing for parameter change in a general class of integer-valued time series models

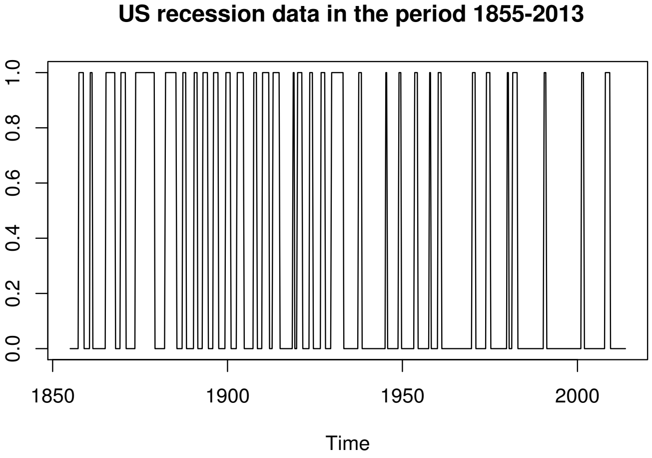
William KENGNE

Université de Cergy-Pontoise

Joint work with M.L. DIOP (Université Gaston Berger, Sénégal)

CIRM, "Processus"
February 19, 2016

An example



US recession data in the period 1855-2013
0.0
0.2
0.4
0.6
0.8
1.0
1850
1900
1950
2000
Time

Outline

Introduction

Exponential family autoregressive models

Test for change detection

Some numerical results

Outline

Counts data modeling

- Count data : $Y_1, \cdots, Y_n$;
- Integer-valued process $(Y_t)_{t \in \mathbb{Z}}$;
- Y_t may depends on $(Y_{t-i})_{i \geq 1}$.

How such data can be modeled?

Linear Poisson autoregression

$$Y_t/Y_{t-1},\cdots \sim \text{Poisson}(\lambda_t) \text{ with } \lambda_t = \alpha_0 + \alpha_1 \lambda_{t-1} + \beta_1 Y_{t-1}.$$

Properties

- Stationary : Ferland *et al.* (2006);
- Ergodicity, inference : Fokianos *et al.* (2009).

Nonlinear Poisson autoregression

$$Y_t / Y_{t-1}, \cdots \sim \text{Poisson}(\lambda_t) \text{ with } \lambda_t = f(\lambda_{t-1}, Y_{t-1}).$$

Properties

- Stationary : Neumann (2011);
- Inference in a semi-parametric setting : Fokianos and Tjøstheim (2012).

See also Doukhan *et al.* (2012) and Doukhan and Kengne (2015) for more general setting.

Outline

Exponential family autoregressive models

Davis and Liu (2012)

Consider a process $Y = (Y_t)_{t\in\mathbb{Z}}$ satisfying :

$$Y_t|\mathcal{F}_{t-1} \sim p(y|\eta_t) \quad \text{with} \quad X_t = f_{\theta^*}(X_{t-1}, Y_{t-1}) \tag{1}$$

$$X_t = \mathbb{E}(Y_t|\mathcal{F}_{t-1}) = A'(\eta_t)$$

with a discrete distribution that satisfied

$$p(y|\eta) = \exp\left\{\eta y - A(\eta)\right\} h(y)$$

$\theta^* \in \Theta \subset \mathbb{R}^d$; $\mathcal{F}_{t-1} = \sigma\left\{\eta_1, X_{t-1}, X_{t-2}, \cdots\right\}$

$$\sup_{\theta\in\Theta} \left|f_\theta(x,y) - f_\theta(x',y')\right| \le \delta_1 \left|x - x'\right| + \delta_2 \left|y - y'\right|.$$

Example 1

Negative binomial INGARCH(1,1)

$$Y_t|\mathcal{F}_{t-1} \sim \mathrm{NB}(r, p_t), \text{ with}$$
$$r\frac{(1-p_t)}{p_t} = \mathbb{E}(Y_t|\mathcal{F}_{t-1}) = X_t = \alpha_0^* + \alpha^* Y_{t-1} + \beta^* X_{t-1};$$

the true parameter $\theta_0 = (\alpha_0^*, \alpha^*, \beta^*)$ belongs to a compact set $\Theta \subset (0, +\infty) \times [0, +\infty)^2$ such that $\alpha + \beta < 1$.

NB(r, p) denotes the negative binomial distribution.

Particular case of (1) : $\eta_t = \log\left(\frac{X_t}{X_t+r}\right)$; $A(\eta_t) = r\log\left(\frac{r}{1-e^{\eta_t}}\right)$.

Example 2

Binary time series

Let $(Y_t)_{t\in\mathbb{Z}}$ be a binary time series satisfying :

$$Y_t|\mathcal{F}_{t-1} \sim \mathrm{B}(X_t) \text{ with } X_t = \alpha_0^* + \alpha^* Y_{t-1} + \beta^* X_{t-1};$$

the true parameter $\theta_0 = (\alpha_0^*, \alpha^*, \beta^*) \in \Theta$ where Θ is a compact subset of $(0, +\infty) \times [0, +\infty)^2$ such that $\alpha_0 + \alpha + \beta < 1$

Particular case of (1) : $\eta_t = \log\left(\frac{X_t}{1-X_t}\right)$; $A(\eta_t) = \log\left(1 + e^{\eta_t}\right)$.

See Fokianos *et al.* (2013b) for similar model with explanatory variables.

Likelihood estimator

Let $(Y_1, \ldots, Y_n)$ be a trajectory generated from the model (1), according to θ_0. The conditional log-likelihood is

$$L_n(\theta) = \log\left(\mathcal{L}(\theta | Y_1, \ldots, Y_n, \eta_1)\right) = \sum_{t=1}^{n} \ell_t(\theta)$$

$$\text{with} \;\; \ell_t(\theta) = \eta_t(\theta) Y_t - A(\eta_t(\theta)).$$

$\eta_t(\theta) = (A')^{-1}(X_t(\theta))$

The maximum likelihood estimator

$$\widehat{\theta}_n := \underset{\theta \in \Theta}{\operatorname{argmax}}(L_n(\theta)).$$

Consistency and asymptotic normally take place (Davis and Liu (2012)).

Outline

Change-point problem

Observations : $Y_1, \cdots, Y_n$.

H$_0$: $(Y_1, \ldots, Y_n)$ is a trajectory of $(Y_t)_{t \in \mathbb{Z}}$ solution of (1), depending on $\theta_0 \in \Theta$.

H$_1$: $\exists \theta_1^*, \theta_2^*$ with $\theta_1^* \neq \theta_2^*, 0 < t^* < n$ such that $(Y_1, \cdots, Y_{t^*})$ is a trajectory of $\{Y_t^{(1)}, t \in \mathbb{Z}\}$ and $(Y_{t^*+1}, \cdots, Y_n)$ a trajectory of $\{Y_t^{(2)}, t \in \mathbb{Z}\}$,
$\{Y_t^{(1)}, t \in \mathbb{Z}\}$ and $\{Y_t^{(2)}, t \in \mathbb{Z}\}$ are stationary solutions of (1) depending on θ_1^* and θ_2^*.

Change-point problem

General strategy

Construct a function φ and choose a constant $C > 0$.

- $\max_{1 \le k \le n} \|\varphi(Y_1, \cdots, X_k) - \varphi(X_1, \cdots, Y_n)\|_{Y_1, \cdots, Y_n} > C \Rightarrow$ change;
- $\max_{1 \le k \le n} \|\varphi(Y_1, \cdots, X_k) - \varphi(Y_1, \cdots, Y_n)\|_{Y_1, \cdots, Y_n} > C \Rightarrow$ change.

Question:

What are the suitable choice of φ and C?

Test statistic

Let $\widehat{\theta}_n(T_{k,k'})$ be the MLE computed on the observations $Y_k, Y_{k+1}, \cdots, Y_{k'}$.

The asymptotic covariance matrix of the estimator under H_0:

$$\widehat{\Omega}_n = \frac{1}{n}\sum_{t=1}^{n}\left(A''\left(\eta_t(\theta)\right)\left(\frac{\partial\eta_t(\theta)}{\partial\theta}\right)\left(\frac{\partial\eta_t(\theta)}{\partial\theta}\right)^T\right)\Bigg|_{\theta=\widehat{\theta}_n(T_{1,n})}$$

$\Rightarrow$ problem under H_1.

Test statistic

Let $(u_n)_{n\geq 1}$ be an integer number sequence satisfying $u_n \to +\infty$, $\frac{u_n}{n} \to 0$ as $n \to +\infty$.

$$\widehat{\Omega}_n(u_n) = \frac{1}{2}\left[\frac{1}{u_n}\sum_{t=1}^{u_n} A''\left(\eta_t(\theta)\right)\left(\frac{\partial\eta_t(\theta)}{\partial\theta}\right)\left(\frac{\partial\eta_t(\theta)}{\partial\theta}\right)^T\Big|_{\theta=\widehat{\theta}_n(T_{1,u_n})}\right.$$
$$\left.+\frac{1}{n-u_n}\sum_{t=u_n+1}^{n} A''\left(\eta_t(\theta)\right)\left(\frac{\partial\eta_t(\theta)}{\partial\theta}\right)\left(\frac{\partial\eta_t(\theta)}{\partial\theta}\right)^T\Big|_{\theta=\widehat{\theta}_n(T_{u_n+1,n})}\right].$$

Test statistic

Let $(v_n)_{n\geq 1}$ be an integer number sequence satisfying $v_n \to +\infty$, $\frac{v_n}{n} \to 0$ as $n \to +\infty$.

The test statistics:

$$\widehat{C}_n = \max_{v_n \leq k \leq n-v_n} \widehat{C}_{k,n} \text{ where}$$

$$\widehat{C}_{n,k} = \frac{1}{q^2\left(\frac{k}{n}\right)} \frac{k^2(n-k)^2}{n^3} \left(\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n})\right)' \widehat{\Omega}_n(u_n) \left(\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n})\right);$$

q : the weight satisfying

$$I_{0,1}(q,c) = \int_0^1 \frac{1}{t(1-t)} \exp\left(-\frac{cq^2(t)}{t(1-t)}\right) dt, \ c > 0.$$

Asymptotic behavior

Theorem

Under H_0 with the above assumptions, if $\exists c > 0$ such that $I(q, c) < \infty$, then

$$\widehat{C}_n \xrightarrow[n\to+\infty]{\mathcal{D}} \sup_{0<\tau<1} \frac{\|W_d(\tau)\|^2}{q^2(\tau)};$$

where W_d is a d-dimensional Brownian bridge.

Theorem

With the above assumptions. Under H_1, if $\theta_1^* \neq \theta_2^*$ then

$$\widehat{C}_n \xrightarrow[n\to+\infty]{P} +\infty.$$

Outline

Illustration for NB-INGARCH(1,1)

$$Y_t|\mathcal{F}_{t-1} \sim \mathrm{NB}(r, p_t), \text{ with } r\frac{(1-p_t)}{p_t} = X_t = \alpha_0^* + \alpha^* Y_{t-1} + \beta^* X_{t-1}$$

Under H_0 : $\theta_0 = (0.20, 0.30, 0.25)$; under H_1 : $\theta_1 = (0.70, 0.3, 0.25)$

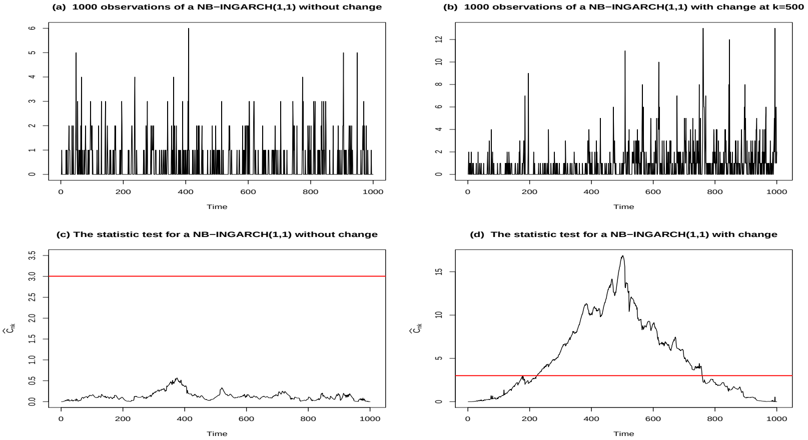
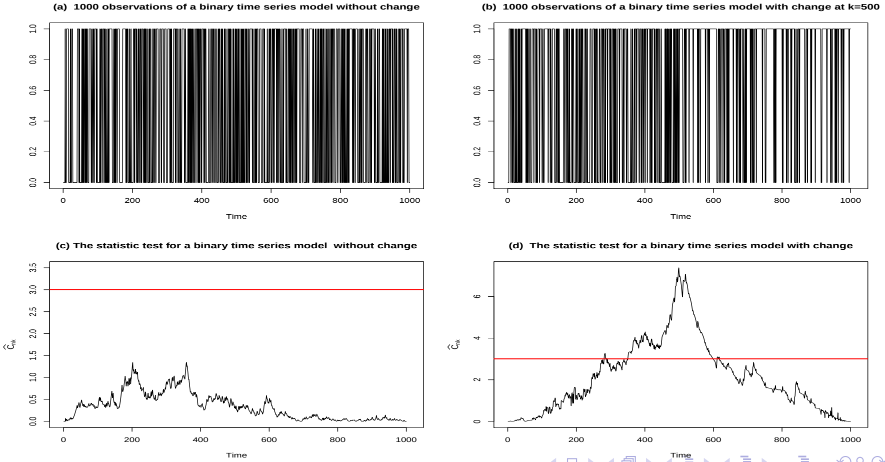


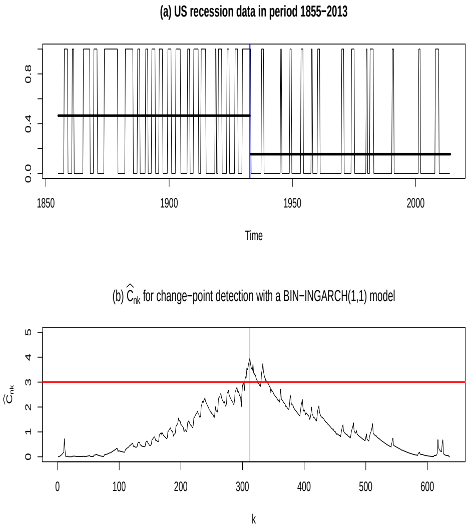
Illustration for binary time series

$$Y_t | \mathcal{F}_{t-1} \sim \mathrm{B}(X_t) \quad \text{with} \quad X_t = \alpha_0^* + \alpha^* Y_{t-1} + \beta^* X_{t-1}$$

Under $\mathrm{H}_0 : \theta_0 = (0.30, 0.15, 0.25)$; under $\mathrm{H}_1 : \theta_1 = (0.05, 0.15, 0.25)$



US recession data



THANK YOU
FOR YOUR
ATTENTION.