Detecting systematic anomalies affecting systems when inputs are stationary time series

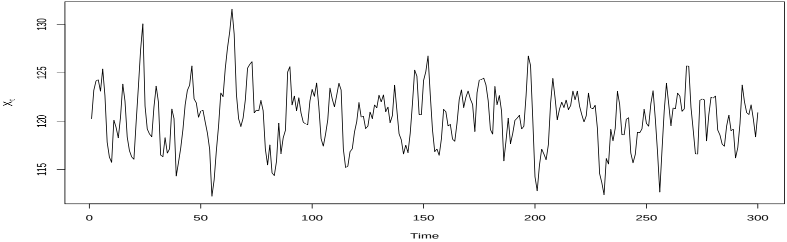
Ričardas Zitikis

School of Mathematical and Statistical Sciences
Western University, Ontario
&
Risk and Insurance Studies Centre
York University, Ontario

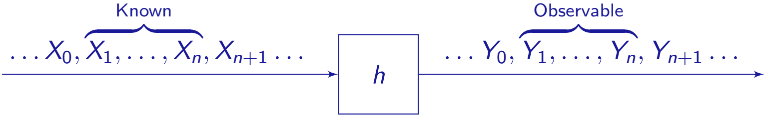
March 10, 2021

You are most welcome to e-mail: rzitikis@uwo.ca

Life as usual: ups and downs...

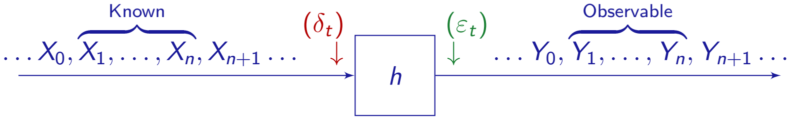


...or is it?



- Inputs $(X_t)_{t\in\mathbb{Z}}$
- Transfer function h
  - ↗ Ex 1: Automatic voltage regulator
  - ↘ Ex 2: Insurer processing claims
- Outputs $(Y_t)_{t\in\mathbb{Z}}$
- Observable pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$

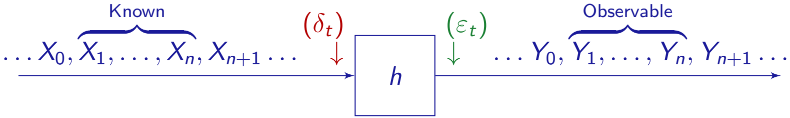
> Think of $(X_t)_{t\in\mathbb{Z}}$ as a stationary and causal time series



- Input risks $(\delta_t)_{t\in\mathbb{Z}}$
- Output risks $(\varepsilon_t)_{t\in\mathbb{Z}}$

Think of risks as ⟶ measurement errors / miscalculations / oversights

- $Y_t = \begin{cases} h(X_t+\delta_t) & \text{when only inputs are directly affected} \\ h(X_t)+\varepsilon_t & \text{when only outputs are affected} \\ h(X_t+\delta_t)+\varepsilon_t & \text{when inputs \& outputs are affected} \end{cases}$



If h were known, we could use, e.g.,

$$\frac{1}{n}\sum_{i=1}^{n}\big(h(X_i)-Y_i\big)^2 \quad \begin{cases} = 0 & \text{when the system is risk free} \\ > 0 & \text{when the system is risk affected} \end{cases}$$

But we only know that $h \in \mathcal{H}$ (model uncertainty)

Our philosophy

We check (for risks) only those systems that were in reasonable order when newly installed

What does "be in reasonable order" mean?

Whose definition to use?

- Manufacturer's definition? Maybe, but likely only indirectly
- Our definition? Yes, because it is aligned with our goals

...and it is on the next slide

Definition. The risk-free outputs $Y_t = h(X_t)$ are in reasonable order if

$$B_n^0 := \frac{1}{n^{1/2}} \sum_{i=2}^{n} |h(X_{i:n}) - h(X_{i-1:n})| = O_{\mathbb{P}}(1)$$

where $X_{1:n} \leq \cdots \leq X_{n:n}$ are the ordered inputs $X_1, \ldots, X_n$

Example (to work out intuition). If $h \in$ Lipschitz, then the risk-free outputs are in reasonable order because

$$B_n^0 \leq \frac{\|h\|_{\text{Lip}}}{n^{1/2}} \sum_{i=2}^{n} |X_{i:n} - X_{i-1:n}| = \|h\|_{\text{Lip}} \frac{X_{n:n} - X_{1:n}}{n^{1/2}}$$

$$= \|h\|_{\text{Lip}} \frac{\text{Range}(X_1, X_2, \ldots, X_n)}{n^{1/2}} = O_{\mathbb{P}}(1)$$

Suppose that a brand new system was in reasonable order

- If it is still risk free, then $I_n \to_{\mathbb{P}}$ somewhere $\neq 0.5$
- If it gets risk affected, then $I_n \to_{\mathbb{P}} 0.5$

What is this magical I_n?

The system-monitoring index (statistic)

$$I_n = \frac{\sum_{i=2}^{n} (Y_{i,n} - Y_{i-1,n})_+}{\sum_{i=2}^{n} |Y_{i,n} - Y_{i-1,n}|}$$

where $Y_{1,n}, \dots, Y_{n,n}$ are the concomitants of $X_1, \dots, X_n$

Example

(X_i, Y_i)	$(X_{i:n}, Y_{i,n})$
$(5, 3)$	$(1, 9)$
$(1, 9)$	$(2, 6)$
$(4, 2)$	$(4, 2)$
$(2, 6)$	$(5, 3)$

A mathematical insight into the meaning of I_n

$$I_n = \frac{\sum_{i=2}^{n}(Y_{i,n} - Y_{i-1,n})_+}{\sum_{i=2}^{n}|Y_{i,n} - Y_{i-1,n}|} = \frac{1}{2}\left(1 + \frac{Y_{n,n} - Y_{1,n}}{\sum_{i=2}^{n}|Y_{i,n} - Y_{i-1,n}|}\right)$$

because $x_+ = (|x| + x)/2$

$$= \frac{1}{2}\left(1 + \frac{\text{Pseudo Range}(Y_1, Y_2, \ldots, Y_n)}{\text{Total Variation}(Y_1, Y_2, \ldots, Y_n)}\right)$$

if there are no risks and n is large

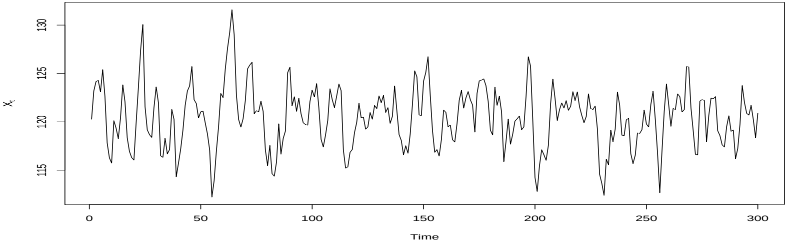
$$\approx \frac{1}{2}\left(1 + \frac{\int h'(x)\mathrm{d}x}{\int |h'(x)|\mathrm{d}x}\right)$$

Example. Let risk free (X_t) be ARMA$(1,1)$ and follow

$$(X_t - 120) = 0.6(X_{t-1} - 120) + \eta_t + 0.4\eta_{t-1}$$

with iid Gaussian innovations $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ where σ_η^2 is such that

$$X_t \sim \mathcal{N}(120, 9)$$



Automatic voltage regulators (AVR's) often keep voltage between

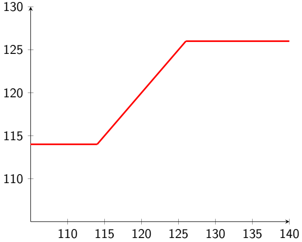
$$120 \pm 6 \text{ volts} \qquad (\pm 5\% \text{ of the nominal voltage})$$

The transfer function

$$h(x) = \begin{cases} x_{\min} & \text{when } x < x_{\min} \\ x & \text{when } x_{\min} \le x \le x_{\max} \\ x_{\max} & \text{when } x > x_{\max} \end{cases}$$

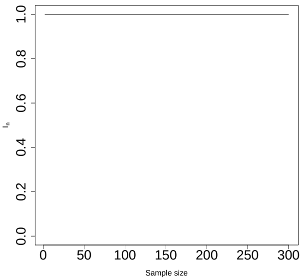
with

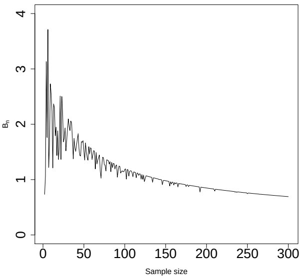
$$x_{\min} = 114 \quad \& \quad x_{\max} = 126$$



Note. Insurance layers often have similar transfer functions: deductible $x_{\min}$, policy limit $x_{\max}$, etc

When the risk-free AVR is in reasonable order, we see





$$I_n^0 = \frac{\sum_{i=2}^n (h(X_{i:n}) - h(X_{i-1:n}))_+}{\sum_{i=2}^n |h(X_{i:n}) - h(X_{i-1:n})|}$$

$$= 1 \neq 0.5 \quad \text{(risk free)}$$

$$B_n^0 = \frac{1}{n^{1/2}} \sum_{i=2}^n |h(X_{i:n}) - h(X_{i-1:n})|$$

$$= O_{\mathbb{P}}(1) \quad \text{(in reasonable order)}$$

Illustrative risk specifications: let (δ_t) and (ε_t) be

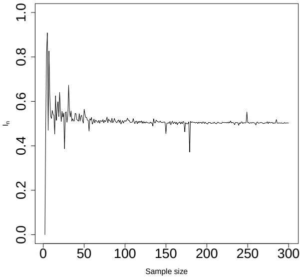
- independent of (X_t)
- independent of each other
- iid Lomax$(\alpha, 1)$ and thus have the means

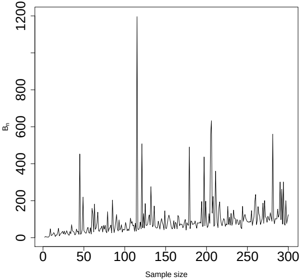
$$\mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = \frac{1}{\alpha - 1}$$

Examples: $\alpha = 1.2 \implies \mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = 5$

$$\alpha = 11 \implies \mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = 0.1$$

$\mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = 5$ (fast converging I_n and B_n)





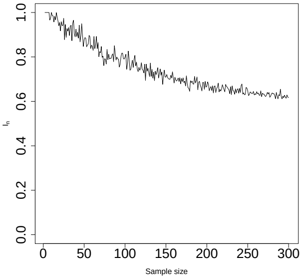
$$I_n = \frac{\sum_{i=2}^{n} (Y_{i,n} - Y_{i-1,n})_+}{\sum_{i=2}^{n} |Y_{i,n} - Y_{i-1,n}|}$$

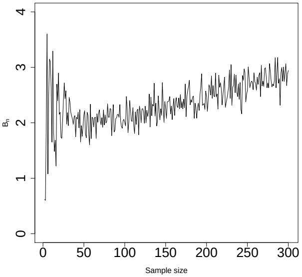
$\to_{\mathbb{P}} 0.5$ (risk affected)

$$B_n = \frac{1}{n^{1/2}} \sum_{i=2}^{n} |Y_{i,n} - Y_{i-1,n}|$$

$\to_{\mathbb{P}} \infty$ (out of reasonable order)

$\mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = 0.1$ (slowly converging I_n and B_n)





$$I_n = \frac{\sum_{i=2}^{n} (Y_{i,n} - Y_{i-1,n})_+}{\sum_{i=2}^{n} |Y_{i,n} - Y_{i-1,n}|}$$

$\to_{\mathbb{P}} 0.5$ (risk affected)

$$B_n = \frac{1}{n^{1/2}} \sum_{i=2}^{n} |Y_{i,n} - Y_{i-1,n}|$$

$\to_{\mathbb{P}} \infty$ (out of reasonable order)

The answer depends on the validity of

$$I_n = \frac{\sum_{i=2}^{n}(Y_{i,n} - Y_{i-1,n})_+}{\sum_{i=2}^{n}|Y_{i,n} - Y_{i-1,n}|} \to_{\mathbb{P}} I_\infty \begin{cases} = 0.5 & \text{when risk affected} \\ \neq 0.5 & \text{when risk free} \end{cases}$$

assuming

$$B_n^0 = \frac{1}{n^{1/2}} \sum_{i=2}^{n} |h(X_{i:n}) - h(X_{i-1:n})| = O_{\mathbb{P}}(1)$$

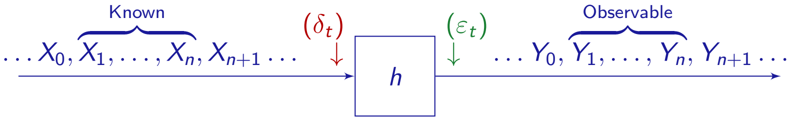
i.e., when risk-free outputs are in reasonable order

Definition. The outputs (Y_t) are out of reasonable order if

$$B_n := \frac{1}{n^{1/2}} \sum_{i=2}^{n} |Y_{i,n} - Y_{i-1,n}| \to_{\mathbb{P}} \infty$$

...and then necessarily $I_n \to_{\mathbb{P}} 0.5$

When do Y_t's become out of reasonable order?



$$h(X_t + \delta_t) + \varepsilon_t = Y_t$$

Roughly speaking, this happens when X_t's are stationary and at least one of the risks does not vanish from I_n, that is, **when δ_t's and ε_t's are**

- noticeable (e.g., if $h(x) = c$, then δ_t's vanish from I_n)
- non-degenerate (e.g., if $\varepsilon_t = c$, then they vanish from I_n)
- not disguised as X_t's (e.g., if $\delta_t = X_t$, $\varepsilon_t = X_t$, and $h(x) = x$, then $Y_t = 3X_t$ and the risk-identifying "3" vanishes from I_n)

In the risk-free scenario, when do we have

$$I_n \to_{\mathbb{P}} \text{ somewhere } \neq 0.5 \text{ ?}$$

- Let the stationary inputs (X_t) satisfy the Glivenko-Cantelli property and be temperately dependent (next two slides)
- Let h be almost everywhere differentiable with vanishing derivative outside an interval $[a, b]$ (recall the AVR function)

Then in the risk-free scenario we have

$$I_n \to_{\mathbb{P}} I_\infty := \frac{\int_a^b (h'(x))_+ \mathrm{d}x}{\int_a^b |h'(x)| \mathrm{d}x} = \frac{1}{2}\left(1 + \frac{\int_a^b h'(x)\mathrm{d}x}{\int_a^b \big|h'(x)\big|\mathrm{d}x}\right)$$

$\neq 0.5$ unless $h(b) = h(a)$

Definition. Inputs (X_t) with the same marginal cdf's F satisfy the Glivenko-Cantelli property if $X_1, \dots, X_n$ asymptotically identify F, that is,

$$\sup_{x \in \mathbb{R}} \left| F_n(x) - F(x) \right| \to_{\mathbb{P}} 0 \quad \text{when} \quad n \to \infty$$

where F_n is the empirical cdf based on $X_1, \dots, X_n$

- If the inputs (X_t) follow the stationary ARMA(p, q) model driven by iid innovations (η_n) with densities, then they satisfy the Glivenko-Cantelli property

Definition. Inputs (X_t) with the same marginal cdf's F are temperately dependent if

$$\mathbb{P}(X_{1:n} \geq x) \to 0 \quad \text{and} \quad \mathbb{P}(X_{n:n} \leq x) \to 0$$

for all x such that $F(x) \in (0, 1)$

Examples

- If (X_t) are iid, then they are temperately dependent
- If $X_t = X$ for all $t \in \mathbb{Z}$, then they are **not** temperately dependent (they are super-strongly dependent)
- If (X_t) are strictly stationary and α-mixing, then they are temperately dependent

The system-monitoring index (statistic)

$$I_n = \frac{\sum_{i=2}^{n}(Y_{i,n} - Y_{i-1,n})_+}{\sum_{i=2}^{n}|Y_{i,n} - Y_{i-1,n}|} \to_{\mathbb{P}} I_\infty \begin{cases} = 0.5 & \text{risk affected} \\ \neq 0.5 & \text{risk free} \end{cases}$$

- is simple to implement
- works as intended in practically plausible situations
- jointly with another index, helps to determine *when* the system gets affected by risks: at the input, output, or both stages

A "stopping" question

When to sound the alarm?

n=50? 100? ...