ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Univariate Self-Starting Shiryaev (U3S): A Bayesian Online Change Point Model in Short Runs

Konstantinos Bourazas[1] and Panagiotis Tsiamyrtzis[2]

[1]Department of Statistics, Athens University of Economics and Business
[2]Department of Mechanical Engineering, Politecnico di Milano

kbourazas@aueb.gr, panagiotis.tsiamyrtzis@polimi.it

September 14, 2021

A brief description

In this work the focus is placed on:

A brief description

In this work the focus is placed on:

- individual univariate short horizon data

A brief description

In this work the focus is placed on:

- individual univariate short horizon data
- detecting persistent shifts without phase I calibration (self-starting)

A brief description

In this work the focus is placed on:

- individual univariate short horizon data
- detecting persistent shifts without phase I calibration (self-starting)

Our proposal:

A brief description

In this work the focus is placed on:

- individual univariate short horizon data
- detecting persistent shifts without phase I calibration (self-starting)

Our proposal:

- is based on Shiryaev's process (1963)

A brief description

In this work the focus is placed on:

- individual univariate short horizon data
- detecting persistent shifts without phase I calibration (self-starting)

Our proposal:

- is based on Shiryaev's process (1963)
- relaxes the strict assumption of known parameters

A brief description

In this work the focus is placed on:

- individual univariate short horizon data
- detecting persistent shifts without phase I calibration (self-starting)

Our proposal:

- is based on Shiryaev's process (1963)
- relaxes the strict assumption of known parameters
- detects a potential change point (At Most One Change - AMOC scenario) and provides posterior inference for all parameters of interest.

Assuming the "At Most One Change (**AMOC**)" model we have:

Assuming the “At Most One Change (**AMOC**)” model we have:

- $\boldsymbol{x_n} = (x_1, x_2, \ldots, x_n)$ is a random sample of data, obtained sequentially

Self-starting Shiryaev (3S)

Assuming the "At Most One Change (**AMOC**)" model we have:

- $\boldsymbol{x_n} = (x_1, x_2, ..., x_n)$ is a random sample of data, obtained sequentially
- $\boldsymbol{\theta}$ is the vector of the In Control (IC) unknown parameter(s)

Assuming the "At Most One Change (**AMOC**)" model we have:

- $\boldsymbol{x_n} = (x_1, x_2, ..., x_n)$ is a random sample of data, obtained sequentially
- $\boldsymbol{\theta}$ is the vector of the In Control (IC) unknown parameter(s)
- ϕ is the vector of the Out Of Control (OOC) unknown parameter(s)

Assuming the "At Most One Change (**AMOC**)" model we have:

- $\boldsymbol{x_n} = (x_1, x_2, ..., x_n)$ is a random sample of data, obtained sequentially
- $\boldsymbol{\theta}$ is the vector of the In Control (IC) unknown parameter(s)
- ϕ is the vector of the Out Of Control (OOC) unknown parameter(s)
- $g(\boldsymbol{\theta}, \phi)$ is a known link function that represents the OOC scenario

Self-starting Shiryaev (3S)

Assuming the “At Most One Change (**AMOC**)” model we have:

- $\boldsymbol{x_n} = (x_1, x_2, ..., x_n)$ is a random sample of data, obtained sequentially
- $\boldsymbol{\theta}$ is the vector of the In Control (IC) unknown parameter(s)
- $\boldsymbol{\phi}$ is the vector of the Out Of Control (OOC) unknown parameter(s)
- $g(\boldsymbol{\theta}, \boldsymbol{\phi})$ is a known link function that represents the OOC scenario
- τ is the unknown change point

Assuming the "At Most One Change (**AMOC**)" model we have:

- $\boldsymbol{x_n} = (x_1, x_2, ..., x_n)$ is a random sample of data, obtained sequentially
- $\boldsymbol{\theta}$ is the vector of the In Control (IC) unknown parameter(s)
- $\boldsymbol{\phi}$ is the vector of the Out Of Control (OOC) unknown parameter(s)
- $g(\boldsymbol{\theta}, \boldsymbol{\phi})$ is a known link function that represents the OOC scenario
- τ is the unknown change point

The likelihood will be:

$$f\left(\boldsymbol{x_n}|\boldsymbol{\theta},\boldsymbol{\phi},\tau\right) = \begin{cases} f\left(\boldsymbol{x_n}|\boldsymbol{\theta},\boldsymbol{\phi},\tau \le n\right) = \prod_{i=1}^{\tau-1} f\left(x_i|\boldsymbol{\theta}\right) \prod_{i=\tau}^{n} f\left(x_i|g(\boldsymbol{\theta},\boldsymbol{\phi})\right) & \text{if } \tau \le n \\ f\left(\boldsymbol{x_n}|\boldsymbol{\theta},\tau > n\right) = \prod_{i=1}^{n} f\left(x_i|\boldsymbol{\theta}\right) & \text{if } \tau > n \end{cases}$$

The stopping time is based on the posterior marginal probability of a change point occurrence, which is:

$$
\begin{aligned}
p\left(\tau \leq n | \boldsymbol{x_n}\right) &= \frac{f\left(\boldsymbol{x_n} | \tau \leq n\right)\pi(\tau \leq n)}{f\left(\boldsymbol{x_n} | \tau \leq n\right)\pi(\tau \leq n) + f\left(\boldsymbol{x_n} | \tau > n\right)\pi(\tau > n)} \\
&= \frac{\displaystyle\sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)} \cdot BF_{k,n+}}{\displaystyle\sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)} \cdot BF_{k,n+} + 1}
\end{aligned}
$$

where $BF_{k,n+} = \dfrac{f(\boldsymbol{x_n} | \tau = k)}{f(\boldsymbol{x_n} | \tau > n)}$ (Bayes Factor), compares the evidence that the $k^{th} \leq n$ observation to be the change point against the evidence that all available n observations are IC.

- The marginal distributions involved in the computation are:

$$f(\boldsymbol{x_n}|\tau > n) = \int_{\Theta} f(\boldsymbol{x_n}|\boldsymbol{\theta}, \tau > n)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$f(\boldsymbol{x_n}|\tau \leq n) = \int_{\Phi}\int_{\Theta} f(\boldsymbol{x_n}|\boldsymbol{\theta}, \boldsymbol{\phi}, \tau \leq n)\pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi}$$

Self-starting Shiryaev (3S)

- The marginal distributions involved in the computation are:

$$f(\boldsymbol{x_n}|\tau > n) = \int_{\Theta} f(\boldsymbol{x_n}|\boldsymbol{\theta}, \tau > n)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$f(\boldsymbol{x_n}|\tau \leq n) = \int_{\Phi}\int_{\Theta} f(\boldsymbol{x_n}|\boldsymbol{\theta}, \boldsymbol{\phi}, \tau \leq n)\pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi}$$

- If the prior $\pi(\boldsymbol{\theta})$ is improper, we sacrifice the s first observations $\boldsymbol{x_{1:s}}$ needed to make the posterior $p(\boldsymbol{\theta}|\boldsymbol{x_{1:s}})$ proper and then replace the prior $\pi(\boldsymbol{\theta})$ by $p(\boldsymbol{\theta}|\boldsymbol{x_{1:s}})$.

Self-starting Shiryaev (3S)

- Stopping times $T(\cdot)$:

Self-starting Shiryaev (3S)

- Stopping times $T(\cdot)$:

Constant decision limit p^*

$$T(p^*) = inf\left\{n \geq 1 : p\left(\tau \leq n | \boldsymbol{x_n}\right) \geq p^*\right\}$$

Adapted decision limit p_n^*

$$T(p_n^*) = inf\left\{ n \geq 1 : p\left(\tau \leq n | \boldsymbol{x_n}\right) \geq p_n^* = \frac{K \cdot \sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)}}{K \cdot \sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)} + 1} \right\}$$

where p^* and K are chosen with respect to the false alarm tolerance.

- Stopping times $T(\cdot)$:

  Constant decision limit p^*
  $T(p^*) = inf \left\{ n \geq 1 : p\left(\tau \leq n | \boldsymbol{x_n}\right) \geq p^* \right\}$

  Adapted decision limit p_n^*

$$T(p_n^*) = inf \left\{ n \geq 1 : p\left(\tau \leq n | \boldsymbol{x_n}\right) \geq p_n^* = \frac{K \cdot \sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)}}{K \cdot \sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)} + 1} \right\}$$

  where p^* and K are chosen with respect to the false alarm tolerance.

- Apart from change point detection, we can also provide inference for the unknown parameters:

  - $$\begin{cases} p_{IC}\left(\boldsymbol{\theta} | \boldsymbol{x_n}\right) & \text{if a change point did not occur} \\ p_{OOC}\left(\boldsymbol{\theta}, \boldsymbol{\phi}, \tau | \boldsymbol{x_n}\right) & \text{if an alarm is raised} \end{cases}$$

Self-starting Shiryaev (3S)

IC scenario ($\tau > n$)

OOC scenario ($\tau \leq n$)

Self-starting Shiryaev (3S)

IC scenario ($\tau > n$)

↓

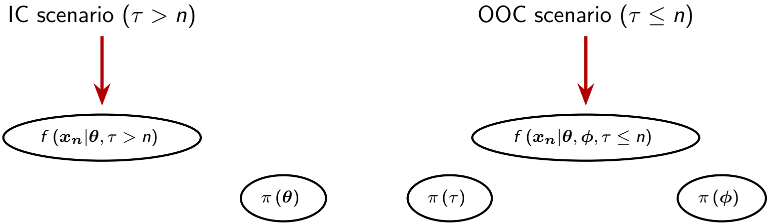
$f\left(\boldsymbol{x_n} | \boldsymbol{\theta}, \tau > n\right)$

OOC scenario ($\tau \leq n$)

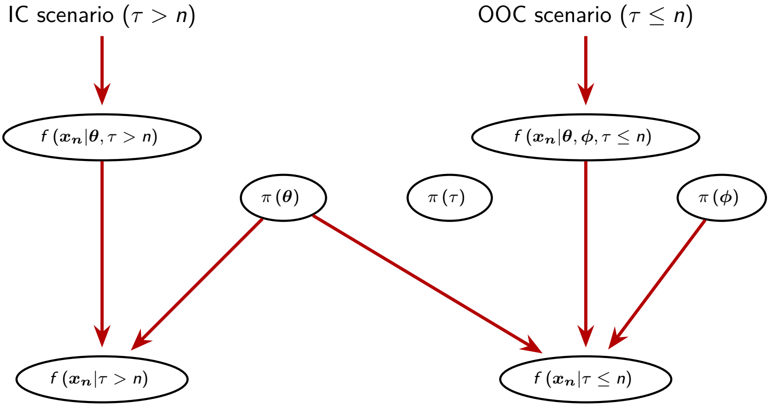
↓

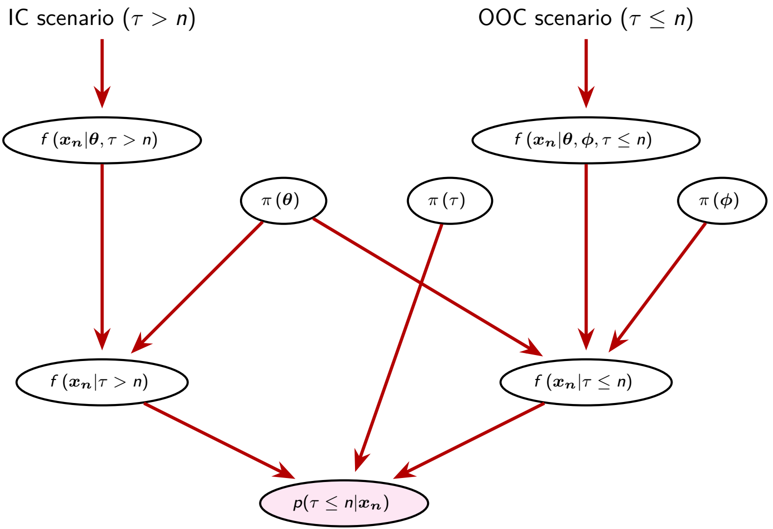
$f\left(\boldsymbol{x_n} | \boldsymbol{\theta}, \boldsymbol{\phi}, \tau \leq n\right)$

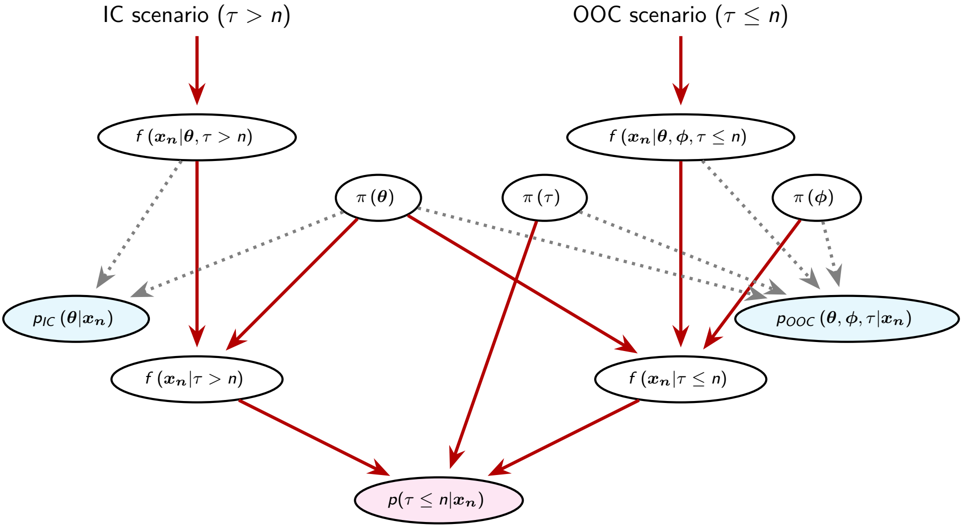
Self-starting Shiryaev (3S)



Self-starting Shiryaev (3S)




IC scenario $(\tau > n)$
OOC scenario $(\tau \leq n)$
$f(\boldsymbol{x_n}|\boldsymbol{\theta}, \tau > n)$
$f(\boldsymbol{x_n}|\boldsymbol{\theta}, \boldsymbol{\phi}, \tau \leq n)$
$\pi(\boldsymbol{\theta})$
$\pi(\tau)$
$\pi(\boldsymbol{\phi})$
$f(\boldsymbol{x_n}|\tau > n)$
$f(\boldsymbol{x_n}|\tau \leq n)$
$p(\tau \leq n|\boldsymbol{x_n})$


IC scenario $(\tau > n)$
OOC scenario $(\tau \leq n)$
$f(\boldsymbol{x_n}|\boldsymbol{\theta}, \tau > n)$
$f(\boldsymbol{x_n}|\boldsymbol{\theta}, \boldsymbol{\phi}, \tau \leq n)$
$\pi(\boldsymbol{\theta})$
$\pi(\tau)$
$\pi(\boldsymbol{\phi})$
$p_{IC}(\boldsymbol{\theta}|\boldsymbol{x_n})$
$p_{OOC}(\boldsymbol{\theta}, \boldsymbol{\phi}, \tau|\boldsymbol{x_n})$
$f(\boldsymbol{x_n}|\tau > n)$
$f(\boldsymbol{x_n}|\tau \leq n)$
$p(\tau \leq n|\boldsymbol{x_n})$

- **Model parameters**

U3S for the Normal mean

- **Model parameters**

$\boldsymbol{\theta} = \left(\theta_1, \theta_2^2\right)$: the mean and the variance of the data

$\boldsymbol{\phi} = \delta$: the magnitude of a mean step change

$g(\boldsymbol{\theta}, \boldsymbol{\phi}) = \theta_1 + \delta \cdot \theta_2$

U3S for the Normal mean

- **Model parameters**

  $\boldsymbol{\theta} = \left(\theta_1, \theta_2^2\right)$: the mean and the variance of the data

  $\phi = \delta$: the magnitude of a mean step change

  $g(\boldsymbol{\theta}, \phi) = \theta_1 + \delta \cdot \theta_2$

- **Model states**

U3S for the Normal mean

- **Model parameters**

  $\boldsymbol{\theta} = \left(\theta_1, \theta_2^2\right)$: the mean and the variance of the data

  $\phi = \delta$: the magnitude of a mean step change

  $g(\boldsymbol{\theta}, \phi) = \theta_1 + \delta \cdot \theta_2$

- **Model states**

  IC state: $x_i|\boldsymbol{\theta} \overset{iid}{\sim} N\left(\theta_1, \theta_2^2\right)$

  OOC state: $x_i|(\boldsymbol{\theta}, \phi) \overset{iid}{\sim} N\left(\theta_1 + \delta \cdot \theta_2, \theta_2^2\right)$

Prior setting

- $\pi(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta}|\boldsymbol{Y})^{\alpha_0}\,\pi_0(\boldsymbol{\theta})$ (power prior, Ibrahim 2000), where:
  $\boldsymbol{Y} = (y_1, ..., y_{n_0})$ is the vector of the historical data (if available),
  $0 \leq \alpha_0 \leq 1$ is fixed and controls the influence of the historical data,
  $\pi_0(\boldsymbol{\theta}) = NIG(\mu_0, \lambda, a, b)$ (*Normal-Inverse-Gamma*) the initial prior.

Prior setting

- $\pi(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta}|\boldsymbol{Y})^{\alpha_0}\, \pi_0(\boldsymbol{\theta})$ (power prior, Ibrahim 2000), where:
  $\boldsymbol{Y} = (y_1, \ldots, y_{n_0})$ is the vector of the historical data (if available),
  $0 \leq \alpha_0 \leq 1$ is fixed and controls the influence of the historical data,
  $\pi_0(\boldsymbol{\theta}) = NIG(\mu_0, \lambda, a, b)$ (*Normal-Inverse-Gamma*) the initial prior.
- $\delta = \gamma \cdot \delta_1 + (1-\gamma) \cdot \delta_2$ (mixture of shifts), where:
  $\delta_i \sim N\left(\mu_{\delta i}, {\sigma_{\delta i}}^2\right)$,
  $\gamma \sim Ber(\pi)$,
  π is the prior probability of the shift δ_1 in the mixture.

Prior setting

- $\pi(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta}|\boldsymbol{Y})^{\alpha_0}\,\pi_0(\boldsymbol{\theta})$ (power prior, Ibrahim 2000), where:
  $\boldsymbol{Y} = (y_1, \ldots, y_{n_0})$ is the vector of the historical data (if available),
  $0 \leq \alpha_0 \leq 1$ is fixed and controls the influence of the historical data,
  $\pi_0(\boldsymbol{\theta}) = NIG(\mu_0, \lambda, a, b)$ (*Normal-Inverse-Gamma*) the initial prior.
- $\delta = \gamma \cdot \delta_1 + (1-\gamma) \cdot \delta_2$ (mixture of shifts), where:
  $\delta_i \sim N\left(\mu_{\delta i}, {\sigma_{\delta i}}^2\right)$,
  $\gamma \sim Ber(\pi)$,
  π is the prior probability of the shift δ_1 in the mixture.
- $\tau \sim DW(p, \beta)$ (*Discrete Weibull*), where
  τ is the location of a potential change point,
  p is the probability for an observation to be OOC,
  β controls the hazard function,
  if $\beta = 1$ then $\tau \sim G(p)$ (*Geometric*)

Under the **IC scenario** the posterior distribution is:

- $(\theta_1, \theta_2^2)|(\tau > n, \boldsymbol{x_n}) \sim NIG\left(\frac{\lambda\mu_0 + X_{1:n}}{\lambda + n}, \lambda + n, a_p, b_p\right)$

where $X_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i$, $a_p = a + \frac{n}{2}$ and $b_p = b + \frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\lambda n}{\lambda + n}(\bar{x} - \mu_0)^2\right)$

Posterior distributions

Under the **IC scenario** the posterior distribution is:

- $(\theta_1, \theta_2^2)|(\tau > n, \boldsymbol{x_n}) \sim NIG\left(\frac{\lambda\mu_0 + X_{1:n}}{\lambda + n}, \lambda + n, a_p, b_p\right)$

where $X_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i$, $a_p = a + \frac{n}{2}$ and $b_p = b + \frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\lambda n}{\lambda + n}(\bar{x} - \mu_0)^2\right)$

Under the **OOC scenario** the posterior distributions are:

- $\theta_1|(\theta_2^2, \delta, \tau \leq n, \boldsymbol{x_n}) \sim N\left(\frac{\lambda\mu_0 + X_{1:n} - n_\tau\delta\theta_2}{\lambda + n}, \frac{\theta_2^2}{\lambda + n}\right)$
- $p\left(\theta_2^2|\theta_1, \delta, \tau \leq n, \boldsymbol{x}\right) \propto \left(\frac{1}{\theta_2^2}\right)^{a_p + \frac{3}{2}} exp\left\{-\frac{2b + S_{1:n}^2 + \lambda\left(\theta_1 - \mu_0\right)^2}{2\theta_2^2} - \frac{\left(X_{\tau:n} - n_\tau\theta_1\right)\delta}{\theta_2}\right\}$

where $n_t = n - t + 1$ and $S_{t_1:t_2}^2 = \sum_{i=t_1}^{t_2}(x_i - \theta_1)^2$

Posterior distributions

- $\delta_i | \left(\theta_1, \theta_2^2, \tau \leq n, \boldsymbol{x_n}\right) \sim N\left(\mu_{\delta pi}, \sigma_{\delta pi}^2\right)$

- $\gamma | \left(\theta_1, \theta_2^2, \delta_i, \tau \leq n, \boldsymbol{x_n}\right) \sim Ber\left(\dfrac{\pi}{\pi + (1-\pi) \cdot exp\left\{\dfrac{\mu_{\delta p2}^2}{2\sigma_{\delta p2}^2} - \dfrac{\mu_{\delta p1}^2}{2\sigma_{\delta p1}^2}\right\} \dfrac{\sigma_{\delta p2}}{\sigma_{\delta p1}}}\right)$

- $p\left(\tau = k | \theta_1, \theta_2^2, \delta, \boldsymbol{x_n}\right) = \dfrac{exp\left\{\dfrac{\delta\left(X_{k:n} - n_k\theta_1\right)}{\theta_2} - \dfrac{n_k\delta^2}{2}\right\}\left((1-p)^{(k-1)^\beta} - (1-p)^{k^\beta}\right)}{\sum_{j=1}^{n} exp\left\{\dfrac{\delta\left(X_{j:n} - n_j\theta_1\right)}{\theta_2} - \dfrac{n_j\delta^2}{2}\right\}\left((1-p)^{(j-1)^\beta} - (1-p)^{j^\beta}\right)}$

where $\mu_{\delta pi} = \dfrac{\mu_{\delta i} + \sigma_{\delta i}^2\left(X_{\tau:n} - n_\tau\theta_1\right)/\theta_2}{1 + n_\tau\sigma_{\delta i}^2}$ and $\sigma_{\delta p1}^2 = \dfrac{\sigma_{\delta i}^2}{1 + n_\tau\sigma_{\delta i}^2}$

Data description:

Data description:

- We use the dataset analyzed by Hawkins (1987). Due to confidentiality issues, the standardized dataset was presented in the paper, subtracting the mean and dividing by the standard deviation.

Data description:

- We use the dataset analyzed by Hawkins (1987). Due to confidentiality issues, the standardized dataset was presented in the paper, subtracting the mean and dividing by the standard deviation.

- It refers to $n = 55$ chemical laboratory that carries out routine indirect (instrumental) assays for precious metals of batches of a feedstock. As a control measure, a sample of a standard reference material is assayed along with each batch of unknowns.

Data description:

- We use the dataset analyzed by Hawkins (1987). Due to confidentiality issues, the standardized dataset was presented in the paper, subtracting the mean and dividing by the standard deviation.

- It refers to $n = 55$ chemical laboratory that carries out routine indirect (instrumental) assays for precious metals of batches of a feedstock. As a control measure, a sample of a standard reference material is assayed along with each batch of unknowns.

- The observations arrive sequentially, assuming:

$$\boldsymbol{X}_i|\boldsymbol{\theta} \overset{iid}{\sim} N\left(\theta_1, \theta_2^2\right)$$

Non-informative prior setting:

Non-informative prior setting:

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ (reference prior, Bernardo, 1979)
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/55, 1) \equiv G(1/55)$

Non-informative prior setting:

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ (reference prior, Bernardo, 1979)
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/55, 1) \equiv G(1/55)$

Decision limit elicitation:

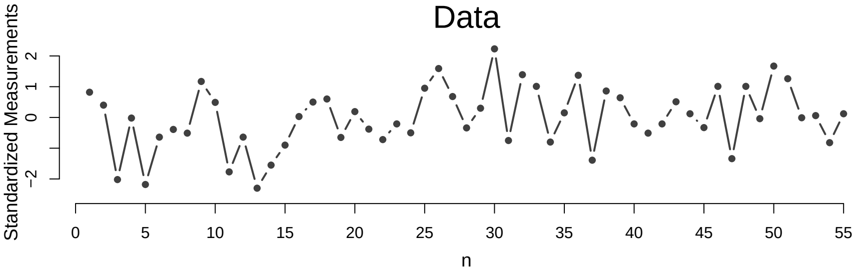
Non-informative prior setting:

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ (reference prior, Bernardo, 1979)
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/55, 1) \equiv G(1/55)$

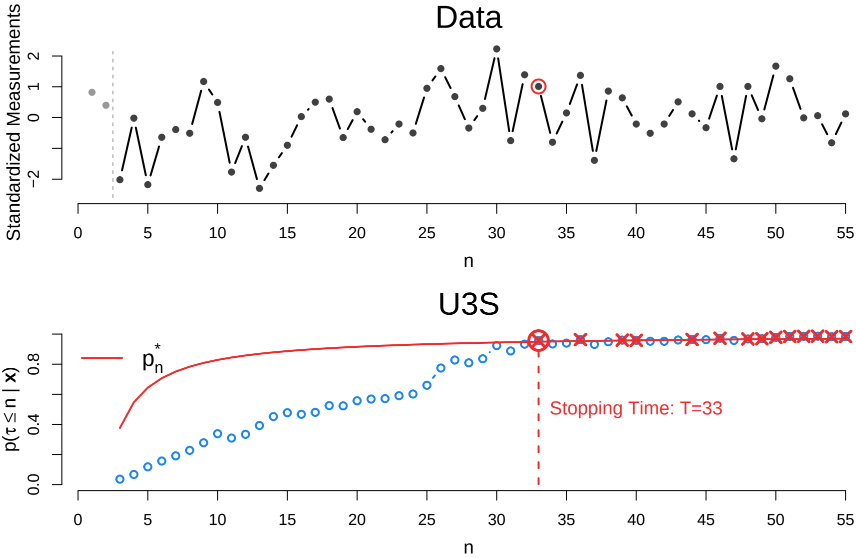
Decision limit elicitation:

- We set ${p_n}^*$ to control $PFA = 20\%$ for $n = 55$ data points.

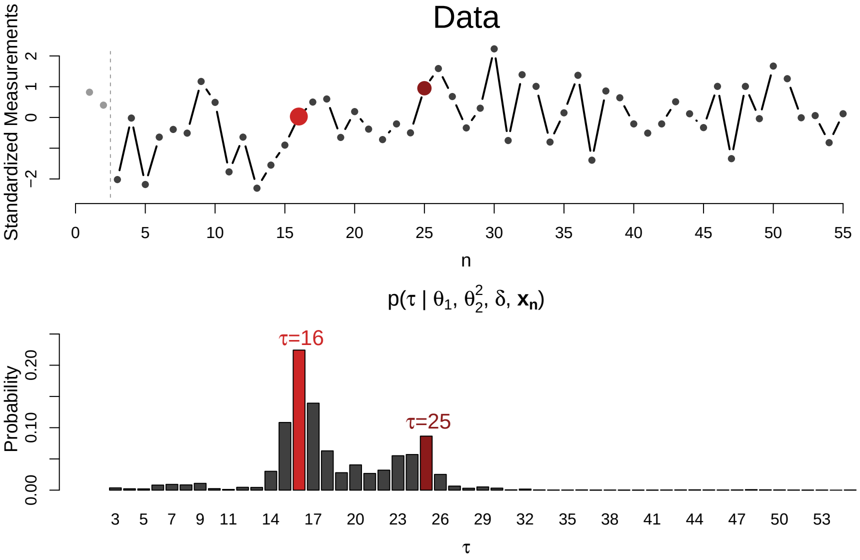
Application to real data

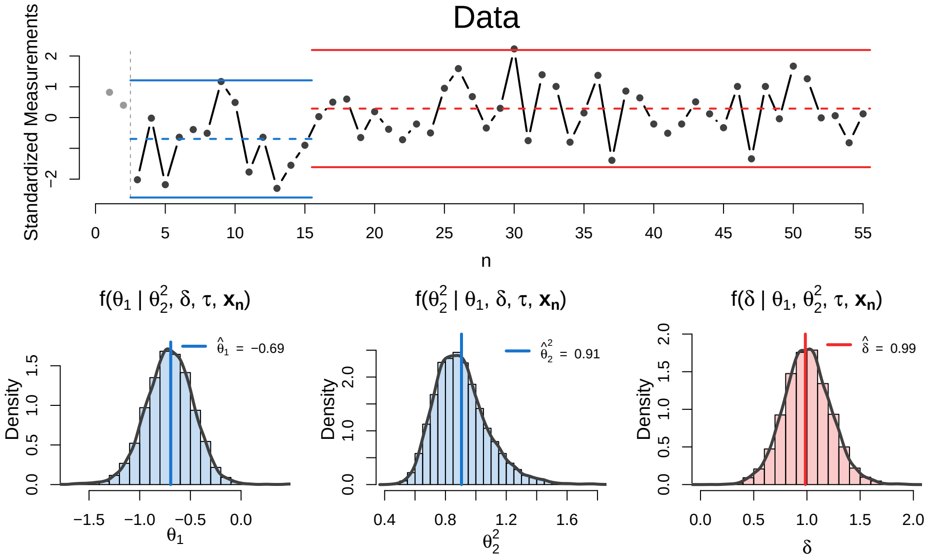


Application to real data



Application to real data





Data
Standardized Measurements
−2
0
1
2
0
5
10
15
20
25
30
35
40
45
50
55
n
$f(\theta_1 \mid \theta_2^2, \delta, \tau, \mathbf{x_n})$
$\hat{\theta}_1 = -0.69$
Density
0.0
0.5
1.0
1.5
−1.5
−1.0
−0.5
0.0
θ_1
$f(\theta_2^2 \mid \theta_1, \delta, \tau, \mathbf{x_n})$
$\hat{\theta}_2^2 = 0.91$
Density
0.0
1.0
2.0
0.4
0.8
1.2
1.6
θ_2^2
$f(\delta \mid \theta_1, \theta_2^2, \tau, \mathbf{x_n})$
$\hat{\delta} = 0.99$
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
2.0
δ

Competing methods:

Competing methods:

- Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998),

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Competing methods:

- Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998),
- Recursive Segmentation and Permutation (RS/P, Capizzi and Masarotto, 2013),

Competing methods:

- Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998),
- Recursive Segmentation and Permutation (RS/P, Capizzi and Masarotto, 2013),
- Univariate Self-starting Shiryaev (U3S).

Competing methods:

- Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998),
- Recursive Segmentation and Permutation (RS/P, Capizzi and Masarotto, 2013),
- Univariate Self-starting Shiryaev (U3S).

IC data:

Competing methods:

- Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998),
- Recursive Segmentation and Permutation (RS/P, Capizzi and Masarotto, 2013),
- Univariate Self-starting Shiryaev (U3S).

IC data:

- For $N = 50$, we assume $X_i | \left(\theta_1, \theta_2^2\right) \overset{i.i.d.}{\sim} N\left(\theta_1, \theta_2^2\right)$, where $\theta_1 = 0$ and $\theta_2^2 = 1$. We simulate $10,000$ iterations of each random sample.

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Competing methods:

- Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998),
- Recursive Segmentation and Permutation (RS/P, Capizzi and Masarotto, 2013),
- Univariate Self-starting Shiryaev (U3S).

IC data:

- For $N = 50$, we assume $X_i | \left(\theta_1, \theta_2^2\right) \overset{i.i.d.}{\sim} N\left(\theta_1, \theta_2^2\right)$, where $\theta_1 = 0$ and $\theta_2^2 = 1$. We simulate $10,000$ iterations of each random sample.

OOC scenarios:

Competing methods:

- Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998),
- Recursive Segmentation and Permutation (RS/P, Capizzi and Masarotto, 2013),
- Univariate Self-starting Shiryaev (U3S).

IC data:

- For $N = 50$, we assume $X_i | \left(\theta_1, \theta_2^2\right) \overset{i.i.d.}{\sim} N\left(\theta_1, \theta_2^2\right)$, where $\theta_1 = 0$ and $\theta_2^2 = 1$. We simulate $10,000$ iterations of each random sample.

OOC scenarios:

- Step changes for the mean from a $N(1, 1)$ and initiating at location 11, or 26, or 41.

U3S prior setting (reference & constant hazard function (r,c)):

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/50, 1)$

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

U3S prior setting (reference & constant hazard function (r,c)):

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/50, 1)$

SSC tuning parameter:

- We set $k = 0.5$

U3S prior setting (reference & constant hazard function (r,c)):

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/50, 1)$

SSC tuning parameter:

- We set $k = 0.5$

RS/P parameter for the maximum number of change points:

- We set $K = 1$

Performance metrics:

Performance metrics:

- We select the appropriate decision limits for each method, so that all of them will have identical Probability of False Alarm (PFA):

$$PFA = P\left(T \leq N | \tau > N\right) = 0.05$$

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Performance metrics:

- We select the appropriate decision limits for each method, so that all of them will have identical Probability of False Alarm (PFA):

$$PFA = P\left(T \leq N \middle| \tau > N\right) = 0.05$$

- We estimate the Probability of Successful Detection (PSD) for each method:

$$PSD\left(\tau\right) = P\left(\tau \leq T \leq N\right)$$

Performance metrics:

- We select the appropriate decision limits for each method, so that all of them will have identical Probability of False Alarm (PFA):

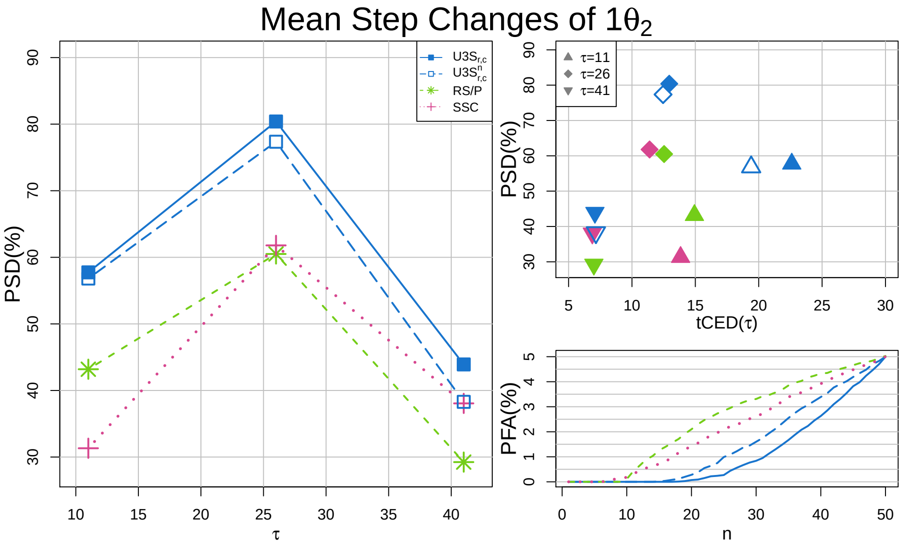
$$PFA = P\left(T \leq N \middle| \tau > N\right) = 0.05$$

- We estimate the Probability of Successful Detection (PSD) for each method:

$$PSD\left(\tau\right) = P\left(\tau \leq T \leq N\right)$$

- We estimate the truncated Conditional Expected Delay (tCED)

$$tCED(\tau) = E_{\tau}(T - \tau + 1|\tau \leq T \leq n)$$


Mean Step Changes of 1θ2
U3S r,c
U3S n r,c
RS/P
SSC
PSD(%)
τ
τ=11
τ=26
τ=41
tCED(τ)
PFA(%)
n

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) ϕ to be ***unknown***

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) ϕ to be ***unknown***
- offering a more ***flexible*** prior for the change point τ

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) ϕ to be ***unknown***
- offering a more ***flexible*** prior for the change point τ
- providing ***posterior inference*** for all the parameters of interest regarding the IC or the OOC scenario.

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) ϕ to be ***unknown***
- offering a more ***flexible*** prior for the change point τ
- providing ***posterior inference*** for all the parameters of interest regarding the IC or the OOC scenario.

Compared to the Frequentist based and Nonparametric alternatives, U3S:

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) ϕ to be ***unknown***
- offering a more ***flexible*** prior for the change point τ
- providing ***posterior inference*** for all the parameters of interest regarding the IC or the OOC scenario.

Compared to the Frequentist based and Nonparametric alternatives, U3S:

- achieves greater detection percentages

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) ϕ to be ***unknown***
- offering a more ***flexible*** prior for the change point τ
- providing ***posterior inference*** for all the parameters of interest regarding the IC or the OOC scenario.

Compared to the Frequentist based and Nonparametric alternatives, U3S:

- achieves greater detection percentages
- has similar or smaller detection delay

Conclusions

U3S process is a generalization of Shiryaev process, enriching the existed methodology in ***three*** ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) ϕ to be ***unknown***
- offering a more ***flexible*** prior for the change point τ
- providing ***posterior inference*** for all the parameters of interest regarding the IC or the OOC scenario.

Compared to the Frequentist based and Nonparametric alternatives, U3S:

- achieves greater detection percentages
- has similar or smaller detection delay
- is more resistant in absorbing an OOC scenario.

References

- Bernardo, J. M. (1979), "Reference Posterior Distributions for Bayesian Inference", *Journal of the Royal Statistical Society* Series B (Methodological), 41, pp. 113-147.
- Capizzi, G. and Masarotto, G. (2013), "Phase I distribution-free analysis of univariate data", *Journal of Quality Technology*, 45, pp. 273-284.
- Hawkins, D. M. (1987), "Selfstarting CUSUM charts for location and scale", *Journal of the Royal Statistical Society: Series D (The Statistician)*, 36, 4, 299-316.
- Hawkins, D. M., and Olwell, D. H. (1998), *Statistics for engineering and physical science-cumulative sum charts and charting for quality improvement*, Springer-Verlag, New York.
- Ibrahim J. & Chen M. (2000). "Power Prior Distributions for Regression Models" *Statistical Science*, Vol. 15, pp. 46-60.
- Shiryaev A. (1963). "On optimum methods in quickest detection problems", *Theory of Probability & Its Applications*, Vol. 8, No. 1, pp. 22-46.

Thank you!