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

Konstantinos Bourazas[1], Panagiotis Tsiamyrtzis[2]

[1]Department of Statistics, Università Cattolica del Sacro Cuore, Milano
[2]Department of Mechanical Engineering, Politecnico di Milano

konstantinos.bourazas@unicatt.it,
panagiotis.tsiamyrtzis@polimi.it

- Statistical Process Control/Monitoring (SPC/M) is an effective area of Statistics that includes all methods that deal with the **quick and valid detection of any disorder** in an ongoing process. Its main aim is to detect when a process deteriorates from its **In Control (IC)** state to the **Out Of Control (OOC)** state

  IC state: only natural causes of variation are observed,

  OOC state: exogenous to the process variation is present

- Statistical Process Control/Monitoring (SPC/M) is an effective area of Statistics that includes all methods that deal with the **quick and valid detection of any disorder** in an ongoing process. Its main aim is to detect when a process deteriorates from its **In Control (IC)** state to the **Out Of Control (OOC)** state

  IC state: only natural causes of variation are observed,

  OOC state: exogenous to the process variation is present

- SPC/M is applied in a **plethora of disciplines**, like:.

  industrial processes,

  medical laboratories,

  economics,

  geophysics etc.

- The **type of shifts**, i.e. the OOC states, that are most often considered in practice, are:

    Transient shifts: an isolated unusual value, i.e. an outlier. It is typically of large size.

    Persistent shifts: systematic changes to at least one parameter of a procedure , e.g. step changes, scale shifts, linear trends, rotations etc.. It is typically of small/medium size.

- The **type of shifts**, i.e. the OOC states, that are most often considered in practice, are:

  Transient shifts: an isolated unusual value, i.e. an outlier. It is typically of large size.

  Persistent shifts: systematic changes to at least one parameter of a procedure , e.g. step changes, scale shifts, linear trends, rotations etc.. It is typically of small/medium size.

- The majority of the proposed SPC/M methods, which are designed to efficiently detect them, typically requires two phases (I/II).

  Phase I is the training and typically offline phase, where independent IC data are gathered and the goal is to perform calibration of the monitoring scheme.

  Phase II follows and it is the testing and typically online phase, where new observations are collected and compared against the IC standards that established in phase I.

Issues in the standard setup

- **Violations** of the assumptions during phase I (like autocorrealtion, parameter shifts etc.) **jeopardize** performance in both phase I and II.

Issues in the standard setup

- **Violations** of the assumptions during phase I (like autocorrealtion, parameter shifts etc.) **jeopardize** performance in both phase I and II.

- Furthermore, the phase I/II separation has certain **restrictions**.

  In phase I analysis a **large amount of independent IC samples** is needed to provide (offline) reliable estimates of the unknown parameter(s).

  The **estimation error** for the parameter(s) of interest is typically not taken into account.

  The **IC information**, which is available from phase II data, **is wasted** using one-off plugged in phase I estimates.

Self-starting methods

- **Self-starting methods** in SPC/M have been proposed to mitigate the problems arising from phase I/II separation. The developed methodology provides testing from the early start of a process without any preliminary calibration.

Self-starting methods

- **Self-starting methods** in SPC/M have been proposed to mitigate the problems arising from phase I/II separation. The developed methodology provides testing from the early start of a process without any preliminary calibration.
- Nevertheless, there does not seem to be a concrete definition in the literature of what can be called "self-starting" and what not.

Self-starting methods

- **Self-starting methods** in SPC/M have been proposed to mitigate the problems arising from phase I/II separation. The developed methodology provides testing from the early start of a process without any preliminary calibration.
- Nevertheless, there does not seem to be a concrete definition in the literature of what can be called "self-starting" and what not.

Definition

A control chart will be called as self-starting if:

- it can provide testing, without the need of a preliminary training phase,
- it allows monitoring and inference after each incoming data point becomes available (online) and not retrospectively (offline),
- the IC and the OOC states contain at least one unknown parameter.

From now on, we will characterize a method as self-starting based on the above definition.

Research goal

In this work the focus is placed on:

- **individual univariate** short horizon data,
- the online detection of **persistent disorders** and the reliable **inference** for the unknown process parameter(s),
- adopting the **Bayesian perspective**, without the requirement of any calibration phase (self-starting).

Research goal

In this work the focus is placed on:

- **individual univariate** short horizon data,
- the online detection of **persistent disorders** and the reliable **inference** for the unknown process parameter(s),
- adopting the **Bayesian perspective**, without the requirement of any calibration phase (self-starting).

Our proposal:

- relaxes the strict assumption of known parameters,
- utilizes prior information (if available),
- focuses on detecting change points,
- provides posterior inference for all parameters of interest.

Self-starting Shiryaev (3S)

- We will propose a family of innovative Bayesian online change point models under the At Most One Change (AMOC) scenario, named Self-Starting Shiryaev (3S).

Self-starting Shiryaev (3S)

- We will propose a family of innovative Bayesian online change point models under the At Most One Change (AMOC) scenario, named Self-Starting Shiryaev (3S).
- 3S is a generalization of the Shiryaev's process (Shiryaev, 1963) and it is based on the posterior marginal probability of a change point occurrence.

Self-starting Shiryaev (3S)

- We will propose a family of innovative Bayesian online change point models under the At Most One Change (AMOC) scenario, named Self-Starting Shiryaev (3S).
- 3S is a generalization of the Shiryaev's process (Shiryaev, 1963) and it is based on the posterior marginal probability of a change point occurrence.
- We will provide all the assumptions and the methodological framework to handle univariate (U3S) data with changes in the mean or the variance.

Self-starting Shiryaev (3S)

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

Self-starting Shiryaev (3S)

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

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)

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)

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 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)
- ϕ is the vector of the Out Of Control (OOC) unknown parameter(s)
- $g(\boldsymbol{\theta}, \phi)$ is a known 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 function that represents the OOC scenario
- τ is the unknown change point

The likelihood will be:

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

Self-starting Shiryaev (3S)

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

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

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

Self-starting Shiryaev (3S)

- The marginal distributions involved in the computation are:

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

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

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 $\boldsymbol{x_{1:s}}$ necessary to make the posterior $p(\boldsymbol{\theta}|\boldsymbol{x_{1:s}})$ proper and use it instead of $\pi(\boldsymbol{\theta})$.

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.

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.

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

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

Self-starting Shiryaev (3S)

IC scenario ($\tau > n$)

OOC scenario ($\tau \leq n$)

Self-starting Shiryaev (3S)

IC scenario ($\tau > n$)

↓

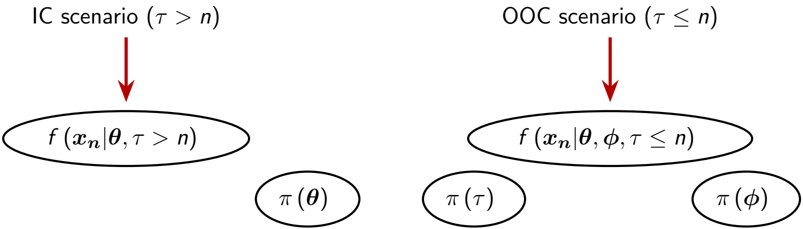
$f\left(\boldsymbol{x_n} | \boldsymbol{\theta}, \tau > n\right)$

OOC scenario ($\tau \leq n$)

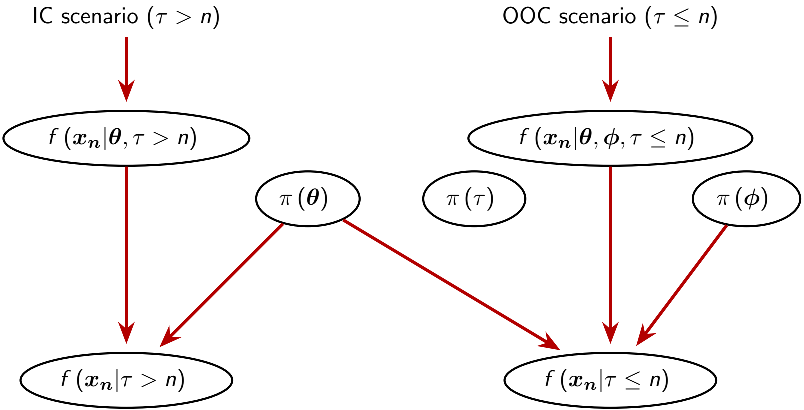
↓

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

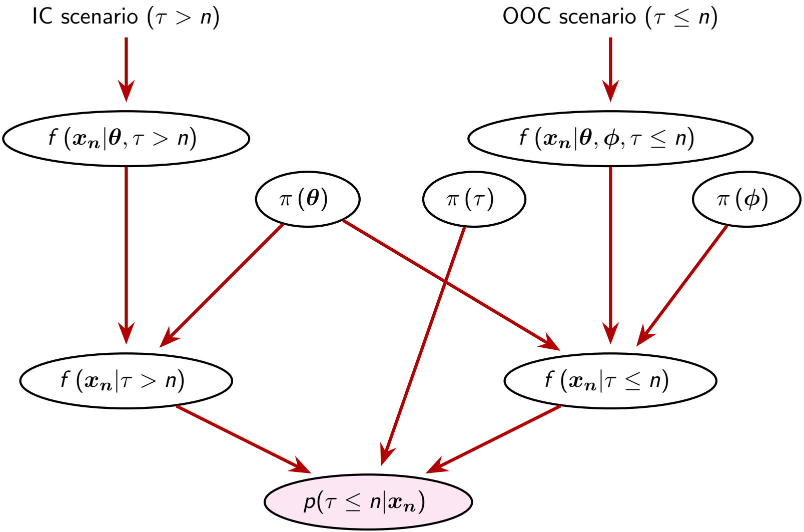
Self-starting Shiryaev (3S)



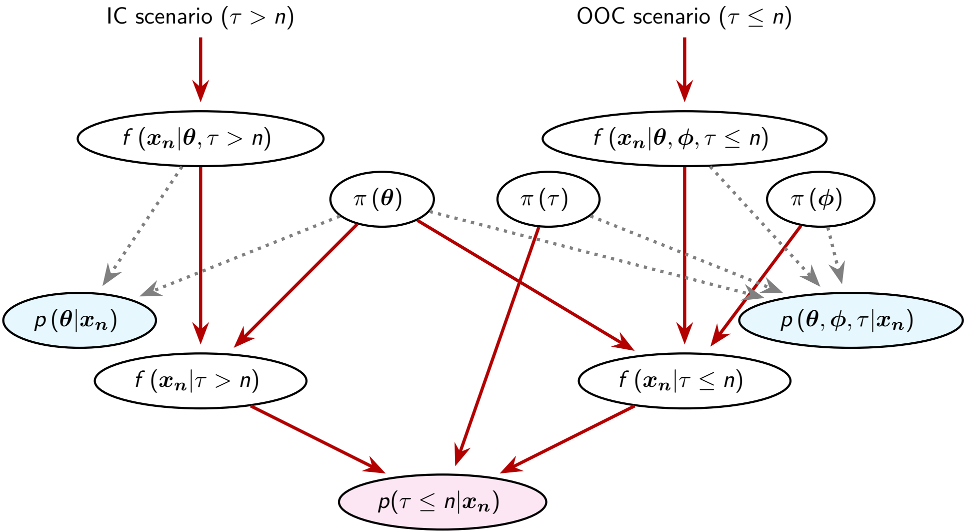
Self-starting Shiryaev (3S)



Self-starting Shiryaev (3S)



Self-starting Shiryaev (3S)



U3S for the Normal mean

- **Model parameters**

- **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:
  $\boldsymbol{Y} = (y_1, ..., y_{n_0})$ 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:
  $\boldsymbol{Y} = (y_1, ..., y_{n_0})$ 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*)

Application to real data

Data description:

Data description:

- We will use data that come from the daily Internal Quality Control (IQC) routine of a medical laboratory and specifically from the area of clinical hemostasis. We are interested in the variable “Factor V”, measured in % regarding the international standards and it is related to blood clotting.

Data description:

- We will use data that come from the daily Internal Quality Control (IQC) routine of a medical laboratory and specifically from the area of clinical hemostasis. We are interested in the variable "Factor V", measured in % regarding the international standards and it is related to blood clotting.

- We sequentially gathered $n = 21$ normally distributed IQC observations (x_i) from a medical lab.

Data description:

- We will use data that come from the daily Internal Quality Control (IQC) routine of a medical laboratory and specifically from the area of clinical hemostasis. We are interested in the variable "Factor V", measured in % regarding the international standards and it is related to blood clotting.

- We sequentially gathered $n = 21$ normally distributed IQC observations (x_i) from a medical lab.

- The observations arrive sequentially, assuming:

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

Weakly informative prior setting:

Weakly informative prior setting:

- $\pi_0\left(\theta_1, {\theta_2}^2 | \boldsymbol{\tau}\right) \sim NIG\left(31.8, 1/2, 2, 4.41\right)$

  Also, we have $n_0 = 37$ IC historical data with $\bar{\mathbf{y}} = 31.73$ and $var(\mathbf{y}) = 3.31$ $(\alpha_0 = 1/n_0)$. Combining these two sources of information, we obtain:

  $\pi\left(\theta_1, {\theta_2}^2 | \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right) \sim NIG\left(31.75, 3/2, 5/2, 6.02\right)$

- $\delta | \gamma \sim \gamma \cdot N(1, 0.25^2) + (1 - \gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/21, 1) \equiv G(1/21)$

Weakly informative prior setting:

- $\pi_0\left(\theta_1, {\theta_2}^2 | \boldsymbol{\tau}\right) \sim NIG\left(31.8, 1/2, 2, 4.41\right)$

  Also, we have $n_0 = 37$ IC historical data with $\bar{\mathbf{y}} = 31.73$ and $var(\mathbf{y}) = 3.31$ $(\alpha_0 = 1/n_0)$. Combining these two sources of information, we obtain:

  $\pi\left(\theta_1, {\theta_2}^2 | \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right) \sim NIG\left(31.75, 3/2, 5/2, 6.02\right)$
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/21, 1) \equiv G(1/21)$

Decision limit elicitation:

Weakly informative prior setting:

- $\pi_0\left(\theta_1, {\theta_2}^2 | \boldsymbol{\tau}\right) \sim NIG\left(31.8, 1/2, 2, 4.41\right)$

  Also, we have $n_0 = 37$ IC historical data with $\bar{\mathbf{y}} = 31.73$ and $var(\mathbf{y}) = 3.31$ $(\alpha_0 = 1/n_0)$. Combining these two sources of information, we obtain:

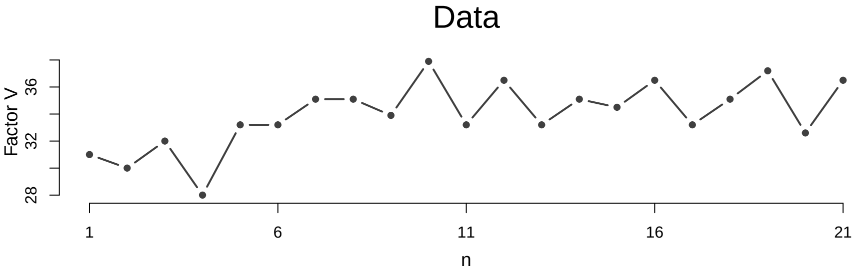
  $\pi\left(\theta_1, {\theta_2}^2 | \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right) \sim NIG\left(31.75, 3/2, 5/2, 6.02\right)$

- $\delta | \gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/21, 1) \equiv G(1/21)$

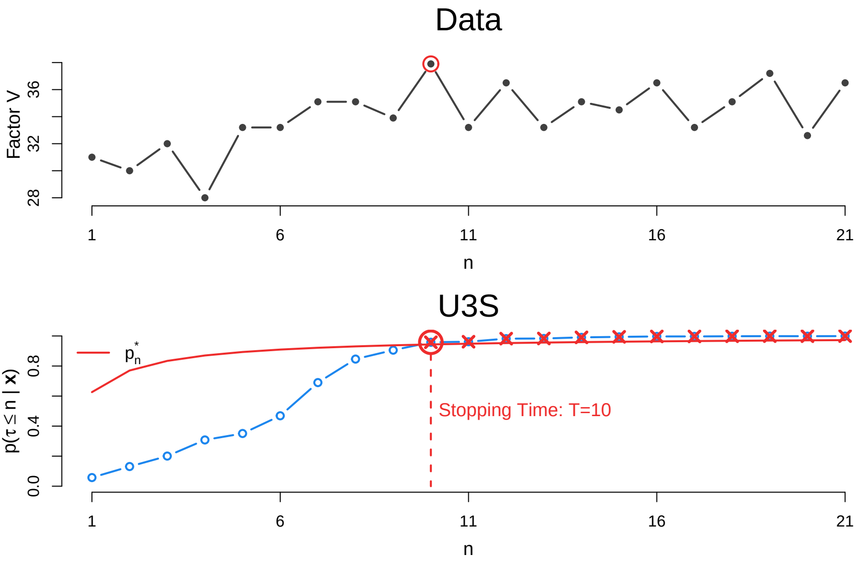
Decision limit elicitation:

- We set ${p_n}^*$ to control $PFA = 5\%$ for $n = 21$ data points.

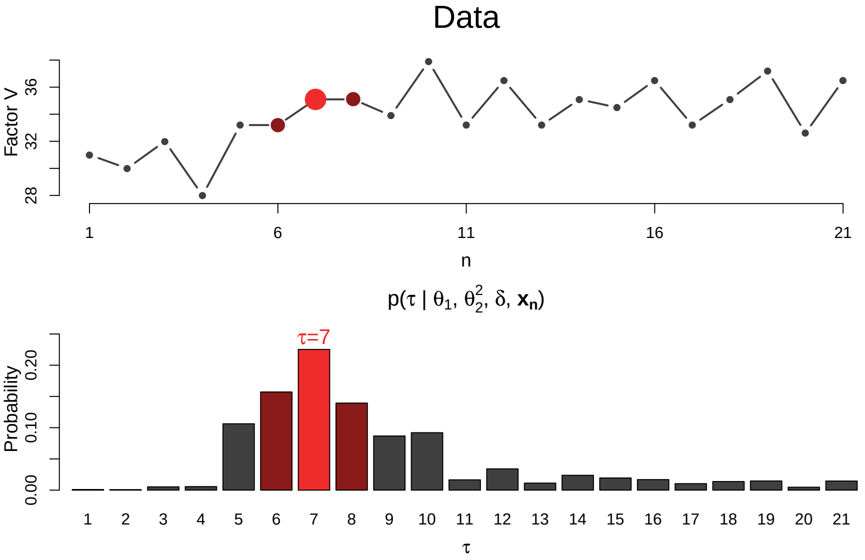
Model illustration



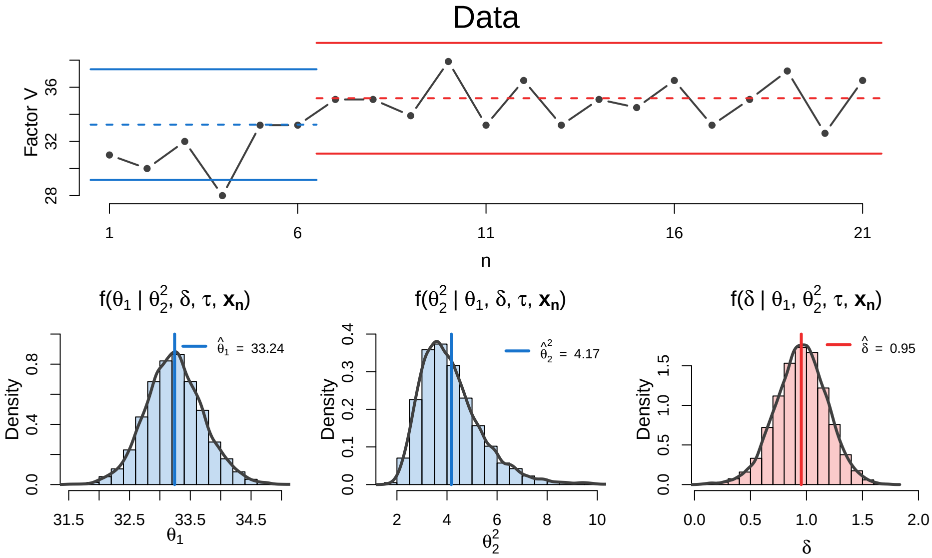
Model illustration



Model illustration



Model illustration



U3S for the Normal variance

- **Model parameters**

- **Model parameters**

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

  $\phi = \kappa$: the magnitude of a shift

  $g(\boldsymbol{\theta}, \phi) = \kappa \cdot \theta_2^2$

U3S for the Normal variance

- **Model parameters**

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

  $\phi = \kappa$: the magnitude of a shift

  $g(\boldsymbol{\theta}, \phi) = \kappa \cdot \theta_2^2$

- **Model states**

U3S for the Normal variance

- **Model parameters**

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

  $\phi = \kappa$: the magnitude of a shift

  $g(\boldsymbol{\theta}, \phi) = \kappa \cdot \theta_2^2$

- **Model states**

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

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

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:
  $\boldsymbol{Y} = (y_1, ..., y_{n_0})$ 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.
- $\kappa = \gamma \cdot \kappa_1 + (1 - \gamma) \cdot \kappa_2$ (mixture of shifts), where:
  $\kappa_i \sim IG(c_i, d_i)$,
  $\gamma \sim Ber(\pi)$,
  π is the prior probability of the shift κ_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:
  $\boldsymbol{Y} = (y_1, ..., y_{n_0})$ 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.

- $\kappa = \gamma \cdot \kappa_1 + (1-\gamma) \cdot \kappa_2$ (mixture of shifts), where:
  $\kappa_i \sim IG(c_i, d_i)$,
  $\gamma \sim Ber(\pi)$,
  π is the prior probability of the shift κ_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 $\tau \sim Geom(p)$ (*Geometric*)

Data description:

U3S application to real data

Data description:

- The dataset comes from Villanueva-Guerra et al. (2017) and it refers to $n = 60$ monthly increments in the S&P 500, which is an American stock market index.

Data description:

- The dataset comes from Villanueva-Guerra et al. (2017) and it refers to $n = 60$ monthly increments in the S&P 500, which is an American stock market index.

- We are interested in a two-sided U3S, either for an inflation or a shrinkage of the variance.

U3S application to real data

Data description:

- The dataset comes from Villanueva-Guerra et al. (2017) and it refers to $n = 60$ monthly increments in the S&P 500, which is an American stock market index.
- We are interested in a two-sided U3S, either for an inflation or a shrinkage of the variance.
- 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:

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2 \equiv \mathit{NIG}(0, 0, -1/2, 0)$ (reference prior, Bernardo, 1979)
- $\kappa|\gamma \sim \gamma \cdot \mathit{IG}(50, 200) + (1-\gamma) \cdot \mathit{IG}(50, 12.5)$
- $\gamma \sim \mathit{Ber}(1/2)$
- $\tau \sim \mathit{DW}(1/60, 1) \equiv \mathit{G}(1/60)$

Non-informative prior setting:

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ (reference prior, Bernardo, 1979)
- $\kappa|\gamma \sim \gamma \cdot IG(50, 200) + (1-\gamma) \cdot IG(50, 12.5)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/60, 1) \equiv G(1/60)$

Decision limit elicitation:

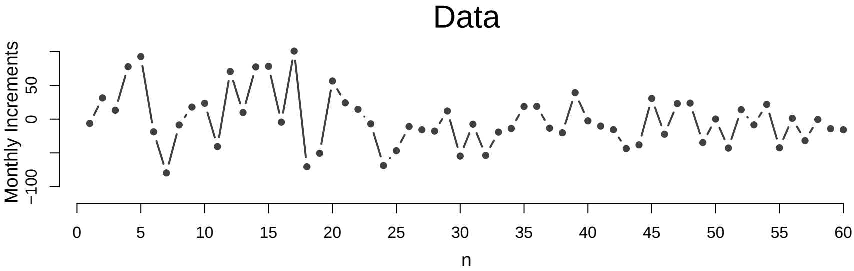
Non-informative prior setting:

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ (reference prior, Bernardo, 1979)
- $\kappa|\gamma \sim \gamma \cdot IG(50, 200) + (1-\gamma) \cdot IG(50, 12.5)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/60, 1) \equiv G(1/60)$

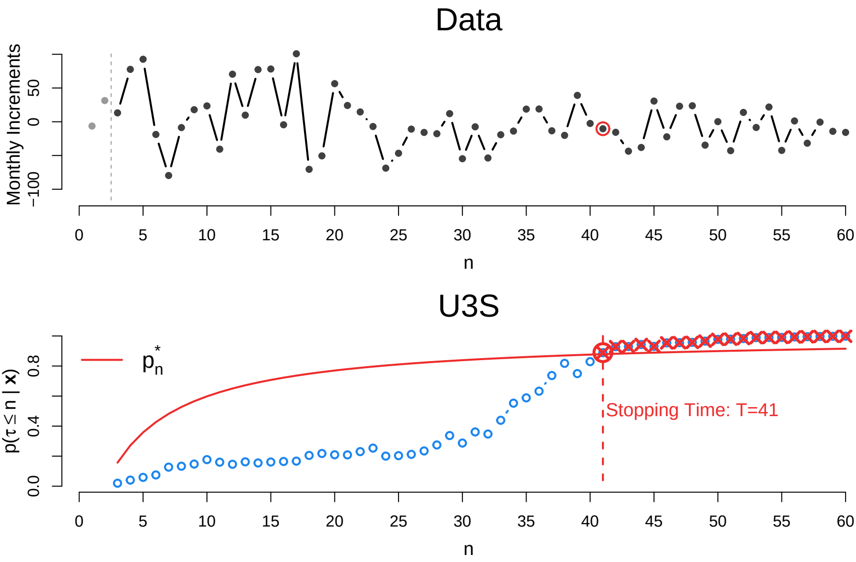
Decision limit elicitation:

- We set ${p_n}^*$ to control $PFA = 10\%$ for $n = 60$ data points.

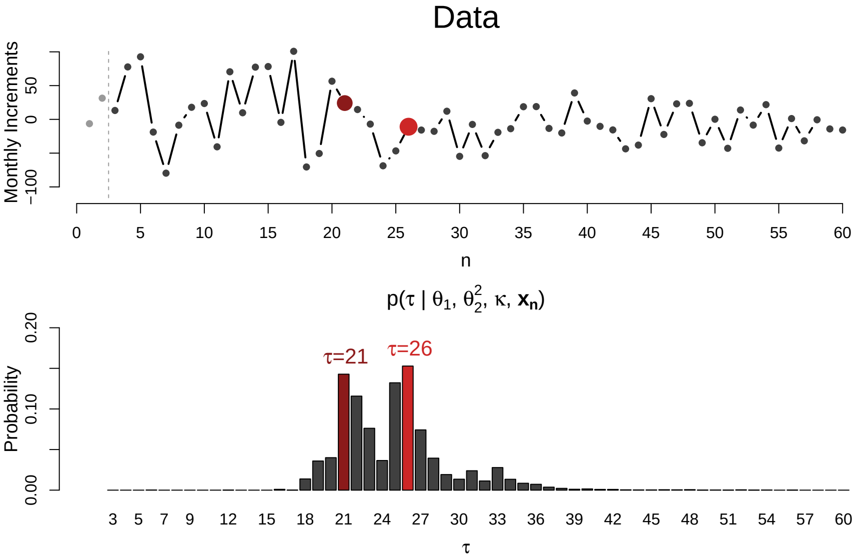
Model illustration

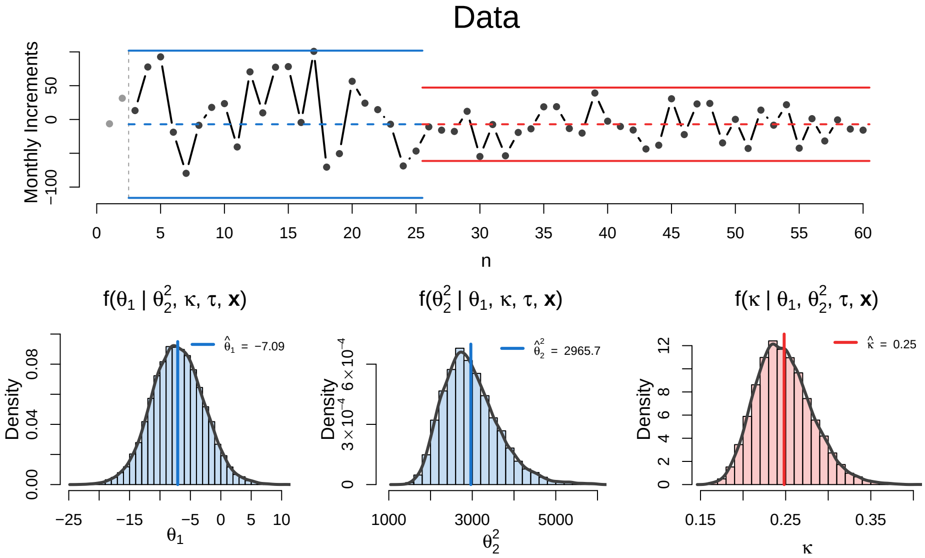


Model illustration



Model illustration




Data
Monthly Increments
−100
0
50
0
5
10
15
20
25
30
35
40
45
50
55
60
n
f(θ1 | θ2², κ, τ, x)
Density
0.00
0.04
0.08
θ̂1 = −7.09
−25
−15
−5
0
5
10
θ1
f(θ2² | θ1, κ, τ, x)
Density
0
3×10⁻⁴
6×10⁻⁴
θ̂2² = 2965.7
1000
3000
5000
θ2²
f(κ | θ1, θ2², τ, x)
Density
0
2
4
6
8
12
κ̂ = 0.25
0.15
0.25
0.35
κ

Competing methods and sensitivity analysis

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 and sensitivity analysis

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:

- **Mean**: 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.
- **Variance**: 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.

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:

- **Mean**: 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.
- **Variance**: 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:

- **Mean**: Step changes for the mean from a $N(1, 1)$ and initiating at location 11, or 26, or 41.
- **Variance**: 50% sd inflation shift, i.e. the OOC is $N(0, 1.5)$, initiating at location 11, or 26, or 41.

U3S prior settings:

U3S prior settings:

Non Informative

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

U3S prior settings:

Non Informative

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

Informative

- $\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/50, 1)$

U3S prior settings:

Non Informative

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

Informative

- $\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/50, 1)$

SSC tuning parameter:

- We set $k = 0.5$

U3S prior settings:

Non Informative

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

Informative

- $\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$
- $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$
- $\gamma \sim Ber(1/2)$
- $\tau \sim DW(1/50, 1)$

SSC tuning parameter:

- We set $k = 0.5$

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

- We set $K = 1$

U3S prior settings:

U3S prior settings:

Non Informative

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ (reference prior)
- $\kappa \sim IG(50, 112.5)$
- $\tau \sim DW(1/50, 1)$

U3S prior settings:

Non Informative

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ (reference prior)
- $\kappa \sim IG(50, 112.5)$
- $\tau \sim DW(1/50, 1)$

Informative

- $\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$
- $\kappa \sim IG(50, 112.5)$
- $\tau \sim DW(1/50, 1)$

U3S prior settings:

Non Informative

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ (reference prior)
- $\kappa \sim IG(50, 112.5)$
- $\tau \sim DW(1/50, 1)$

Informative

- $\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$
- $\kappa \sim IG(50, 112.5)$
- $\tau \sim DW(1/50, 1)$

SSC tuning parameter:

- We set $k \approx 1.46$

U3S prior settings:

Non Informative

- $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ (reference prior)
- $\kappa \sim IG(50, 112.5)$
- $\tau \sim DW(1/50, 1)$

Informative

- $\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$
- $\kappa \sim IG(50, 112.5)$
- $\tau \sim DW(1/50, 1)$

SSC tuning parameter:

- We set $k \approx 1.46$

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 Family Wise Error Rate (FWER):

$$FWER = P\left(T \leq N | \omega > N\right) = 0.05$$

Performance metrics:

- We select the appropriate decision limits for each method, so that all of them will have identical Family Wise Error Rate (FWER):

$$FWER = P\left(T \leq N | \, \omega > N\right) = 0.05$$

- We estimate the Probability of Successful Detection (PSD, Frisen, 1992) for each method:

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

Performance metrics:

- We select the appropriate decision limits for each method, so that all of them will have identical Family Wise Error Rate (FWER):

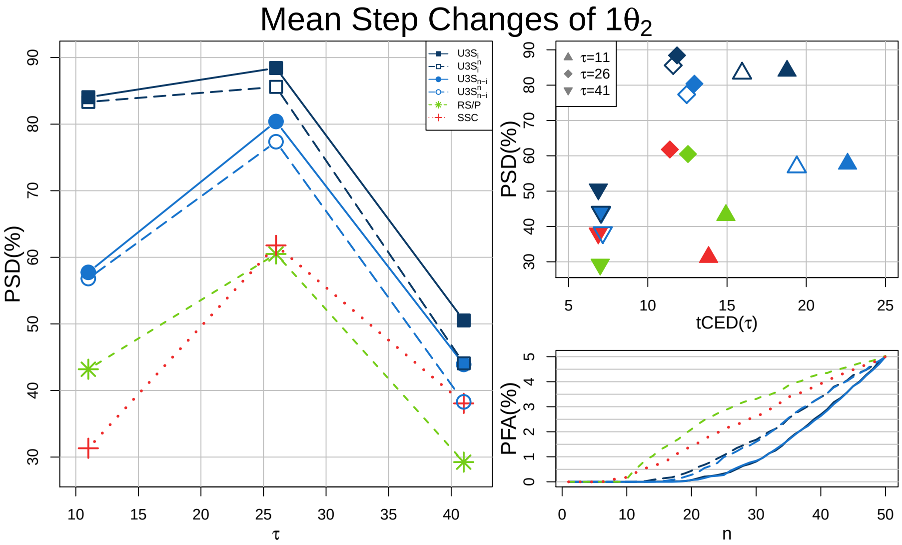
$$FWER = P\left(T \leq N \middle| \omega > N\right) = 0.05$$

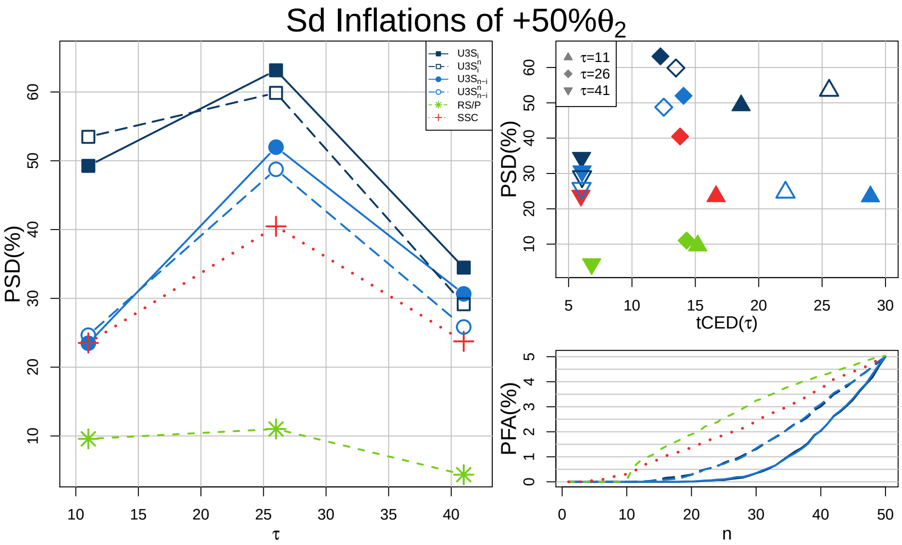
- We estimate the Probability of Successful Detection (PSD, Frisen, 1992) for each method:

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

- We estimate the truncated Conditional Expected Delay (tCED)

$$tCED(\omega) = E_{\omega}(T - \omega + 1 | \omega \leq T \leq N)$$


Mean Step Changes of 1θ2
U3S_i
U3S_i^n
U3S_{n−i}
U3S_{n−i}^n
RS/P
SSC
PSD(%)
τ
τ=11
τ=26
τ=41
tCED(τ)
PFA(%)
n


Sd Inflations of +50%θ2
U3S_i
U3S_i^n
U3S_n−i
U3S_n−i^n
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.

Comparing 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.

Comparing 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.

Comparing to the Frequentist based and Nonparametric alternatives, U3S:

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

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.

Comparing 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.

Frisen M. (1992). "Evaluations of methods for statistical surveillance" *Statistics in Medicine*, Vol. 11, pp. 1489-1502.

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!

Questions?