Latin American Journal of Mathematics (ISSN 2965-0798)
https://doi.org/10.14244/lajm.v3i1.28

Change point detection in $AR(1)$ series by optimal stopping technique*

Reza Habibi[1]

[1]Department of Banking-Iran Banking Institute Central Bank of Iran,
Tehran, Iran, Postal code: 1664647111

r_habibi@ibi.ac.ir, habibi1356@yahoo.com

Abstract. *The change point problem arises in various practical fields such as economics, engineering, medicine, quality control, statistical process control, and financial time series. The aim of this study is to detect location and time of change point at which the behavior of underlying statistical models changes in mean, variance or some other influential parameters. Change point detection methods are divided into two main branches: online methods, that aim to detect changes as soon as they occur in a real-time setting and offline methods that retrospectively detect changes when all samples are received. In practice, there are many parametric (including maximum likelihood and information criterion) and non-parametric methods. Bayesian change point detection introduces a modular Bayesian framework for online estimation of changes in the generative parameters of sequential data. Time series analysis has become increasingly important in diverse fields including medicine, aerospace, finance, business, meteorology, and entertainment. Time series data are sequences of measurements over time describing the behavior of systems. These behaviors can change over time due to external events and/or internal systematic changes in dynamics/distribution. In the current paper, change point analysis in AR(1) is studied using the optimal stopping technique. The logit of probability of having a change at a specific time is studied using the Bayesian and non-Bayesian methods. Snell envelopment method is applied to locate the possible change. Finally, concluding remarks are proposed.*

Keywords – *AR(1), Bayesian, Change point, Logit function, Snell envelopment*
MSC2020 – *62M10, 62F15*

*Received 12 February 2024; received in revised form 13 October 2024; accepted 19 November 2024. Available online 16 December 2024

1. Introduction

The first order autoregressive process AR(1), has been widely used and implemented in time series analysis. Different estimation methods have been employed in order to estimate the autoregressive parameter. The AR(1) process is an important building block of Box-Jenkins approach of time series modeling. As a first-order Markov linear dependence between data, the AR(1) has been played critical role to model the local trends of financial series such as stocks, shares and exchange rates. Different type of estimation methods for the AR(1) process have been developed and proposed in literature. Frequent estimation methods, including method of moments estimation (MME), conditional least squares estimation (CLS), exact maximum likelihood estimation (MLE), and conditional maximum likelihood estimation (CMLE) are commonly used, see [1]. To improve the current estimation procedures, specifically for small samples, a Bayesian method of estimation is considered for the AR(1) model. In general, reasons involving Bayesian approach in time series analysis are; firstly, this approach can successfully provide logical interpretation for statistical inferences in time series analysis, and secondly, results can always be updated based on assimilation of new information. Autoregressive models are statistical models used for time series analysis, where current values are predicted based on a linear combination of past values, see [2]. These models assume that past behavior influences future outcomes, making them useful for forecasting trends and patterns in data over time. A statistical model is autoregressive if it predicts future values based on past values. For example, an AR(1) model might seek to predict a stock's future prices based on its past performance.

Because of frequent changes in economic and financial environments, parameters of AR(1) may have time varying structures. For example, in practice, it is necessary to check the stability of AR(1) model over the time. For example, the slope parameter of AR(1) series is suspected to change throughout the time passes. Let

$$X_i = \beta_i X_{i-1} + \epsilon_i,\ i \geq 2,$$

be the dynamic AR(1) process at which $|\beta_i| < 1$ (stationary condition of process) ϵ_t is a sequence of independent and identically distributed (iid) random variables with zero mean and standard deviation $\sigma_\epsilon < \infty$. Observing the first n data X_i, $i = 1, \ldots, n$, it is interested to test the null hypothesis of no change point

$$H_{0n} : \beta_i = \beta,\ i = 1, \ldots, n$$

vs. the alternative AMOC (at most one change point) hypothesis

$$H_{1n} : \beta_i = \beta,\ i = 1, \ldots, \kappa \text{ and } \beta_i = \beta + \delta,\ i = \kappa + 1, \ldots, n.$$

Here, δ is the unknown magnitude of change and κ is the random change point with prior probabilities $\pi_i = P(\kappa = i)$, $i \geq 1$ The change point analysis for time series, specially AR(1) series, has received considerable attention applying various statistical methods, see [3] and references therein for a comprehensive review. To detect the change point in AR(1) process under the Bayesian setting, we use the optimal stopping technique, which is a special type of optimal control methods. Optimal control refers to solving problems where decisions are made continuously over time, considering changing information, by using the concept of 'cost-to-go' to minimize integrated costs from a current state to a target state. The theory of optimal stopping is concerned with the problem of choosing a time to take a given action based on sequentially observed random variables in order to maximize an expected payoff or to minimize an expected cost. Problems of this type are found in the area of statistics, where the action taken may be to test a hypothesis or to estimate a parameter, and in the area of operations research, where the action may be to replace a machine, hire a secretary, or reorder stock, etc.

In the current paper, the change point problem in AR(1) process is studied using the optimal stopping time theory. In the next section, change point analysis, using Bayesian (see [4]) and non-Bayesian approach are studied and concluding remarks are given in section 3.

2. Two approaches

Here, the change point detection using the optimal stopping method, based on two non-Bayesian and Bayesian approaches is presented.

2.1. Preliminaries

Following Veeravalli and Banerjee [5], let $X_1^n = (X_1, \dots, X_n)$ observations up to time n and the posterior probability

$$p_n = P(\kappa \leq n | X_1^n)$$

denote that a change has been occurred before the n-th observation. Throughout the current paper, conditional probabilities and expectation with respect to X_1^n is equivalent to these conditional quantities given the information set (σ-field) generated by X_1^n and are used exchangeable. Then, the Bayesian theorem implies that

$$p_{n+1} = P(\kappa \leq n+1 | X_1^n, X_{n+1})$$

is proportional to

$$f_1(X_{n+1} | X_1^n, \kappa \leq n+1)\tilde{p}_n,$$

at which

$$\tilde{p}_n = P(\kappa \leq n+1 | X_1^n)$$

and f_1 is the conditional density of observations after the change point. Clearly, given $\kappa \leq n+1$, the density f_1 reduces to $f_1(X_{n+1}|X_n)$ (since change has occurred at $n+1$ time or before that). Also,

$$\tilde{p}_n = p_n + P(\kappa = n+1 | X_1^n).$$

To compute

$$P(\kappa = n+1 | X_1^n),$$

notice that X_1^n says that if $\kappa > n$ or not? In the case of $\kappa > n$, probability of $\kappa = n+1$ is measured via prior distributions. Thus,

$$\tilde{p}_n = p_n + P(\kappa = n+1|\kappa > n)P(\kappa > n|X_1^n) = p_n + \gamma_{n+1}(1 - p_n),$$

where

$$\gamma_{n+1} = P(\kappa = n+1|\kappa > n) = \frac{P(\kappa = n+1)}{P(\kappa > n)} = \frac{\pi_{n+1}}{\sum_{j \geq n+1} \pi_j} \in (0,1).$$

As soon as,

$$\pi_i = \rho(1-\rho)^{i-1}$$

be the geometric prior, then $\gamma_n = \rho$. One can see that the posterior ratio Y_{n+1} is

$$Y_{n+1} = \frac{p_{n+1}}{1 - p_{n+1}} = \frac{\tilde{p}_n}{1 - \tilde{p}_n}\Delta_{n+1},$$

where Δ_{n+1} is the likelihood ratio

$$\frac{f_1(X_{n+1}|X_n)}{f_0(X_{n+1}|X_n)}.$$

From a practical point of view, when p_n is close to 1, then, Y_n converges to infinity. To overcome this difficulty, a refinement as

$$Y_n = \frac{p_n}{1 - p_n + 0.05}$$

is considered. Then, the maximum value of Y_n when $p_n \approx 1$, is 20. Here f_0 is the conditional density of observations before the change point. However, one can see that

$$\frac{\tilde{p}_n}{1 - \tilde{p}_n} = \frac{1}{1 - \gamma_{n+1}} \frac{p_n}{1 - p_n} + \frac{\gamma_{n+1}}{1 - \gamma_{n+1}}.$$

Thus,

$$Y_{n+1} = \left(\frac{1}{1 - \gamma_{n+1}} Y_n + \frac{\gamma_{n+1}}{1 - \gamma_{n+1}} \right) \Delta_{n+1}.$$

Since Y_n is increasing function of p_n, thus, the large values of Y_n (equivalently, p_n) indicates existence of a change at time n. Indeed, a stopping time τ is found at which $E(Y_\tau)$ attains its maximum. A stopping time (referred also as Markov time) is a random time at which a stochastic process is stopped (controlled). Generally, stopping time techniques are used to derive the statistical properties of hitting time, exit time, up-crossing, and first passages; to prove the maximal inequalities and variants of optional sampling theorems; to propose distribution of maximum or minimum of a specific process, and to study the behavior of stopped process. Usually, Dynkin theorems and local martingales are closely related to stopping time techniques. Financial applications of stopping times are American option pricing and decision about its early exercise, financial trading and detection of momentum time, analyzing Dynkin game and Israeli option pricing. Stopping times and stopped process have increasingly applications in control engineering and signal processing fields, see [6].

Here, $p_0 = 0$, hence, $Y_0 = 0$; thus,

$$Y_1 = \frac{\gamma_1}{1-\gamma_1}\Delta_1.$$

Under the null hypothesis, since all X_n's are distributed as f_0, then,

$$E(\Delta_n|F_{n-1}) = \int \frac{f_1(x_n)}{f_0(x_n)} f_0(x_n)dx_n = 1, \quad \text{for each } n,$$

however, when X_n comes from f_1 (say, observations after the change, under the alternative hypothesis), then

$$E(\Delta_n|F_{n-1}) = \int \frac{f_1(x_n)}{f_0(x_n)} f_1(x_n)dx_n$$

may be larger or smaller than 1.

2.2. Non-Bayesian approach

Suppose that γ_n for each n is close to zero, then, $Y_1 = \Delta_1$ and

$$Y_{n+1} = \Delta_{n+1}Y_n.$$

Here, Y_n reduces to the likelihood ratio statistic. In this case, under the null hypothesis,

$$E(Y_{n+1} \mid F_n) = Y_n,$$

where F_n is the information set generated by Y_1^n. Thus, Y_n is a martingale process. Using the optional sampling theorem, for each finite stopping time τ, we have:

$$E(Y_\tau) = E(Y_1) = E(\Delta_1) = 1.$$

Thus, there is no stopping time such that $E(Y_\tau)$ attains its maximum.

Remark 2.1. Notice that

$$\log(Y_i) = \log(Y_{i-1}) + \log(\Delta_i).$$

Under the null hypothesis, as soon as

$$E(\log(\Delta_i)) = 0,$$

then $\log(Y_i)$ is a random walk, and renewal theory in optimal stopping of Veeravalli and Banerjee [5] is applicable, as follows. Before that, to make sure that $E(\log(\Delta_i))$ is small, notice that, as f_1 is close to f_0, then, clearly, it gets small. To see this fact in detail, suppose that $f_0(x_i|x_{i-1})$ is the density of the normal

$$N(\beta x_{i-1}, \sigma^2)$$

distribution and $f_1(x_i|x_{i-1})$ is the density of

$$N((\beta+\delta)x_{i-1},\sigma^2)$$

distribution. Then,

$$\log(\Delta_i)=\frac{\delta}{\sigma^2}(x_i-\beta x_{i-1}-0.5\delta x_{i-1}).$$

Thus, as soon as δ converges to zero or $\sigma^2\to\infty$, then $\log(\Delta_i)$ tends to zero. Let

$$y_i=\log(Y_i),\ \tau=\inf\{i\mid y_i\geq b\},\ \text{and}\ \tau_+=\inf\{i\mid y_i\geq 0\}.$$

Then

$$\lim_{b\to\infty}P(y_\tau-b>u)=(E(y_{\tau_+}))^{-1}\int_u^\infty P(y_{\tau_+}>x)\,dx,$$

see Theorem 2.5 of Veeravalli and Banerjee [5].

2.2.1. Snell envelopment

Under the alternative hypothesis, the Snell envelopment procedure may be applied to find the possible location of the change point. To this end, notice that

$$g_i=\max(Y_i,E(g_{i+1}|F_i));\ g_n=Y_n.$$

One can see that

$$E(g_n\mid F_{n-1})=E(\Delta_n|F_{n-1})Y_{n-1}.$$

Because of the Markov property of Δ_n, then $E(\Delta_n|F_{n-1})$ is a function say $\theta_{n-1}(\cdot)$ of X_{n-1}. Thus,

$$E(g_n|F_{n-1})=\theta_{n-1}(X_{n-1})Y_{n-1}.$$

It is also seen that

$$g_{n-1}=\max\left(1,\theta_{n-1}(X_{n-1})\right)Y_{n-1}.$$

Let $\aleph_{n-1}(X_{n-1})=\max(1,\theta_{n-1}(X_{n-1}))$. Therefore,

$$g_{n-1}=\aleph_{n-1}(X_{n-1})Y_{n-1}.$$

As follows, using the Snell recursive structure, we derive the functional forms of the functions $\theta_i(X_i)$ and $\aleph_i(X_i)$. To this end, by taking the expectation, it is seen that

$$\begin{aligned}E(g_{n-1}|F_{n-2})&=E(\aleph_{n-1}(X_{n-1})\Delta_{n-1}Y_{n-2}|F_{n-2})\\&=E(\aleph_{n-1}(X_{n-1})\Delta_{n-1}|X_{n-2})Y_{n-2}\\&=\theta_{n-2}(X_{n-2})Y_{n-2}.\end{aligned}$$

Hence, the next backward g_{n-2} is given by

$$g_{n-2} = \max(1, \theta_{n-2}(X_{n-2}))Y_{n-2} = \aleph_{n-2}(X_{n-2})Y_{n-2}.$$

Generally, using induction, the auxiliary process of the Snell procedure is identified by two functions $\theta_i(X_i)$ and $\aleph_i(X_i)$; i.e.,

$$g_i = \aleph_i(X_i)Y_i.$$

One can see that

$$\begin{aligned}\aleph_i(X_i) &= \max(1, \theta_i(X_i)),\\ \theta_i(X_i) &= E(\aleph_{i+1}(X_{i+1})\Delta_{i+1}|X_i),\end{aligned}$$

where

$$\aleph_n(X_n) = 1,\ \theta_{n-1}(X_{n-1}) = E(\Delta_n|X_{n-1}),\ \text{and}\ \aleph_{n-1}(X_{n-1}) = \max(1, \theta_{n-1}(X_{n-1})).$$

By finding $\theta_i(X_i)$ and $\aleph_i(X_i)$, the rule of change point detection by the Snell procedure is found. It is seen that, before the actual change, all $\aleph_i(X_i)$ and $\theta_i(X_i)$ are one. However, after the change, a shift occurs in g_i's and the first (stopping) time at which $\mathbb{E}(Y_\tau)$ attains its maximum is κ itself.

2.2.2. Second criterion

Following Veeravalli and Banerjee [5], another version of the test statistic Y_n is given. First, assume that f_0 and f_1 are completely known. Then,

$$Y_n = \frac{\sum_{i=1}^n \pi_i \prod_{j=1}^i f_0(x_j|x_{j-1}) \prod_{j=i+1}^n f_1(x_j|x_{j-1})}{\left(\sum_{i=n+1}^\infty \pi_i\right) \prod_{j=1}^n f_0(x_j|x_{j-1})} = \frac{\sum_{i=1}^n \pi_i \prod_{j=i+1}^n \Delta_j}{\sum_{i=n+1}^\infty \pi_i}.$$

It is easy to see that

$$Y_{n+1} = \left(\frac{\pi_n}{\Pi_{n+1}} + \frac{\Pi_n}{\Pi_{n+1}} Y_n\right)\Delta_{n+1},$$

where

$$\Pi_n = \sum_{i=n+1}^\infty \pi_i.$$

As

$$\frac{\Pi_n}{\Pi_{n+1}} \to 1 \text{ and } \frac{\pi_n}{\Pi_{n+1}} \to 0,$$

the Bayesian test statistic reduces to the likelihood ratio test statistic. Next, in the case of small δ (local change), then

$$\log(\Delta_{n+1}) = \log(f_{\beta+\delta}(x_{n+1}|x_n)) - \log(f_\beta(x_{n+1}|x_n)) \approx \frac{\partial}{\partial\beta}\log(f_\beta(x_{n+1}|x_n)),$$

where β is the unknown parameter before the change point which is changed to $\beta+\delta$ after the change point.

For example, if f_0 is normal, then

$$\log(\Delta_{n+1}) \approx \frac{\delta}{\sigma}\left(\frac{x_{n+1}-\beta x_n}{\sigma}\right) = \lambda z_{n+1},$$

where

$$z_{n+1} = \frac{x_{n+1}-\beta x_n}{\sigma}$$

and assume that $\lambda = \frac{\delta}{\sigma}$ is known. Therefore, $\text{logit}(Y_{n+1})$ is a random walk as follows:

$$\text{logit}(Y_{n+1}) = \text{logit}(Y_n) + \lambda z_{n+1}.$$

Therefore, the posterior probability of having a change at some time n depends on its value in previous time and λ. Let

$$\hat{\beta}_n = \frac{\sum_{j=2}^{n} X_j X_{j-1}}{\sum_{j=2}^{n} X_{j-1}^2}$$

be the least squares estimate of β, under H_{0n}. Under the null hypothesis, we used all samples from 1 to n to derive $\hat{\beta}_n$. In this way, z_{n+1} is estimated by

$$\frac{x_{n+1}-\hat{\beta}_n x_n}{\hat{\sigma}_n}.$$

Here, another version of Y_n based on the estimated residual process e_i is given by

$$e_i = X_i - \hat{\beta}_n X_{i-1} = \epsilon_i - (\hat{\beta}_n - \beta)X_{i-1},\ i \geq 2.$$

To this end, consider that:

i) Under the alternative hypothesis, assuming the change point is i, the parameters before and after the changes are estimated by

$$\hat{\beta}_i = \frac{\sum_{j=2}^{i} X_j X_{j-1}}{\sum_{j=2}^{i} X_{j-1}^2},\quad \hat{\beta}_i^* = \frac{\sum_{j=i+1}^{n} X_j X_{j-1}}{\sum_{j=i+1}^{n} X_{j-1}^2}.$$

ii) Let $f_\vartheta(x_j|x_{j-1})$ be the density of the Normal distribution $N(\vartheta x_{j-1}, \sigma^2)$ computed at x_j. Therefore, notice that

$$Y_n = \frac{\sum_{i=1}^{n} \pi_i \prod_{j=1}^{i} f_{\hat{\beta}_i}(x_j|x_{j-1}) \prod_{j=i+1}^{n} f_{\hat{\beta}_i^*}(x_j|x_{j-1})}{\left(\sum_{i=n+1}^{\infty} \pi_i\right) \prod_{j=1}^{n} f_{\hat{\beta}_n}(x_j|x_{j-1})}.$$

First, assume that σ^2 is known. By some algebraic manipulation, it is seen that:

$$Y_n = \frac{\sum_{i=1}^{n} \pi_i \exp\left(\frac{1}{2\sigma^2} L_i\right)}{\left(\sum_{i=n+1}^{\infty} \pi_i\right)},$$

where

$$L_i = \sum_{j=1}^{n}(x_j - \hat{\beta}_n x_{j-1})^2 - \sum_{j=1}^{i}(x_j - \hat{\beta}_i x_{j-1})^2 - \sum_{j=i+1}^{n}(x_j - \hat{\beta}_i^* x_{j-1})^2.$$

Following Pan et al. [7], notice that:

$$\sum_{j=1}^{n}(x_j - \hat{\beta}_n x_{j-1})^2 = \sum_{j=1}^{i}(x_j - \hat{\beta}_i x_{j-1} + (\hat{\beta}_i - \hat{\beta}_n)x_{j-1})^2 + \sum_{j=i+1}^{n}(x_j - \hat{\beta}_i^* x_{j-1} + (\hat{\beta}_i^* - \hat{\beta}_n)x_{j-1})^2.$$

Let

$$s_i = \sum_{j=1}^{i} x_{j-1}^2, \; w_i = \frac{s_i}{s_n}.$$

Therefore,

$$L_i = s_i(\hat{\beta}_i - \hat{\beta}_n)^2 + (s_n - s_i)(\hat{\beta}_i^* - \hat{\beta}_n)^2.$$

Notice that

$$\hat{\beta}_i^* = \frac{\hat{\beta}_n - w_i \hat{\beta}_i}{1 - w_i}.$$

So,

$$L_i = s_n \frac{w_i}{1 - w_i}(\hat{\beta}_i - \hat{\beta}_n)^2.$$

Also,

$$\hat{\beta}_i - \hat{\beta}_n = \frac{\sum_{j=1}^{i} x_{j-1} e_j}{s_n w_i}.$$

Hence

$$L_i = \frac{\left(\sum_{j=1}^{i} x_{j-1} e_j\right)^2}{s_n w_i (1 - w_i)}.$$

By replacing $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=2}^{n} e_j^2$, it is seen that

$$Y_n = \frac{\sum_{i=1}^{n} \pi_i \exp\left(\frac{t_i^2}{2s_n}\right)}{\sum_{i=n+1}^{\infty} \pi_i},$$

where

$$t_i = \frac{\sum_{j=1}^{i} x_{j-1} e_j}{\hat{\sigma}\sqrt{w_i(1 - w_i)}}.$$

Lemma 2.2. *Under H_{0n}, given $\hat{\beta}_n$, the residual process e_k is an ARMA time series, presented by*

$$e_i = \beta e_{i-1} + \epsilon_i + \hat{\beta}_n \epsilon_{i-1}.$$

The moments of e_t are as follow

$$E(e_i) = 0, \; var(e_i) = \frac{1 + 2\beta E(\hat{\beta}_n) + E(\hat{\beta}_n^2)}{1 - \beta^2},$$

and

$$cov(e_i, e_{i-h}) = \begin{cases} \frac{E((\hat{\beta}_n+\beta)(\hat{\beta}_n+\beta+1))}{1-\beta^2}\sigma_\epsilon^2, & |h| = 1, \\ \beta E(e_i e_{i-h+1}|\hat{\beta}_n)\sigma_\epsilon^2, & |h| > 1. \end{cases}$$

Proof. It is easy to see that:

$$e_i - \beta e_{i-1} = X_i - \beta X_{i-1} - \hat{\beta}_n(X_{i-1} - \beta X_{i-2}).$$

Since ϵ_i is independent of $\hat{\beta}_n$, e_i is an ARMA process. Thus,

$$E(e_i|\hat{\beta}_n) = 0, \ \text{var}(e_i|\hat{\beta}_n) = \frac{1 + 2\beta\hat{\beta}_n + \hat{\beta}_n^2}{1-\beta^2}.$$

Hence,

$$E(e_i) = E(E(e_i|\hat{\beta}_n)) = 0,$$

and

$$\text{var}(e_i) = E(\text{var}(e_i|\hat{\beta}_n)) + \text{var}(E(e_i|\hat{\beta}_n)) = \frac{1 + 2\beta E(\hat{\beta}_n) + E(\hat{\beta}_n^2)}{1-\beta^2}.$$

Regarding the covariance terms, we know that

$$E(e_i e_{i-h}|\hat{\beta}_n) = \frac{(\hat{\beta}_n + \beta)(\hat{\beta}_n + \beta + 1)}{1-\beta^2}\sigma_\epsilon^2 \ \text{if } |h| = 1,$$

and

$$E(e_i e_{i-h}|\hat{\beta}_n) = \beta E(e_i e_{i-h+1}|\hat{\beta}_n)\sigma_\epsilon^2 \quad \text{if } |h| > 1.$$

Thus, taking the expectations on both sides proves the lemma. □

Remark 2.3. The partial sum process of e_i is given by

$$Q_i = \sum_{j=2}^{i} X_{j-1}e_j.$$

Empirically, it is seen that, under H_{0n}, the plot of Q_i against the number of observations i oscillates around zero. It remains between two specified boundaries with high probability. However, when there is a change in the slope parameter, the plot of Q_i creates a peak out of the boundary.

2.3. Bayesian Approach

Notice that

$$Y_i = (a_i Y_{i-1} + b_i)\Delta_i,$$

where

$$a_i = \frac{1}{1-\gamma_i} \text{ and } b_i = \frac{\gamma_i}{1-\gamma_i}.$$

Under the null hypothesis, it is seen that

$$E(Y_i|F_{i-1}) = a_i Y_{i-1} + b_i.$$

Thus, Y_i is not a martingale. Suppose that

$$Z_i = c_i Y_i + d_i$$

is a martingale. Therefore,

$$E(Z_i|F_{i-1}) = c_i a_i Y_{i-1} + c_i b_i + d_i = c_{i-1} Y_{i-1} + d_{i-1}.$$

To make sure the above equation holds, it suffices to assume that

$$c_i a_i = c_{i-1},\ c_i b_i + d_i = d_{i-1}.$$

It is seen that

$$c_i = c_0 \frac{1}{\prod_{j=1}^{i} a_j} = c_0 \prod_{j=1}^{i} (1-\gamma_j),\ d_i = d_0 - \sum_{j=1}^{i} c_j b_j.$$

One can see that Z_i is a decreasing function of Y_k. Thus,

$$U_i = -Z_i$$

is an increasing function and martingale. Hence, there is no stopping time that $E(U_\tau)$ gets its maximum value. However, under the alternative hypothesis, assuming $E(\Delta_n|X_{n-1}) > 1$, the first point at which $E(U_\tau)$ attains its maximum is κ itself.

2.3.1. Asymptotic envelopment

Here, following Lustri et al. [8], the asymptotic expression for the expectation of Snell envelopment sequences h_i is derived. In what follows, the optimal stopping time is searched to maximize $E(Y_\tau)$. Before going ahead, the following lemma is proposed, which is necessary for asymptotic envelopment.

Lemma 2.4. *For every continuous random variable X and positive real number a, then*

$$E(\max(X-a)) = \int_a^\infty P(X > x)\, dx.$$

Notice that

$$h_i = \max(Y_i, E(h_{i+1}|F_i));\ h_n = Y_n.$$

Let

$$W_i = E(h_{i+1}|F_i) \text{ and } E(h_i) = \mu_i.$$

Suppose that the variance of h_i is small. Since

$$\text{var}(W_i) \leq \text{var}(h_i),$$

it follows that var(W_i) is small. Therefore, h_i is close to its expectation, μ_i. Notice that

$$E(W_i) = \mu_{i+1}.$$

Write h_i as

$$h_i = W_i + \max(Y_i - W_i, 0).$$

Therefore, by taking the expectation on both sides, see Shah [2], it is seen that

$$\mu_i = \mu_{i+1} + E(\max(Y_i - \mu_{i+1} - (W_i - \mu_{i+1}), 0)).$$

Notice that since the variance of $W_i - \mu_{i+1}$ is small and its expectation is zero, then this term is negligible, see Fathan and Delage [9]. Therefore,

$$\mu_i = \mu_{i+1} + E(\max(Y_i - \mu_{i+1}, 0)) = \mu_{i+1} + \int_{\mu_{i+1}}^{\infty} \zeta_i(z) dz,$$

where

$$\zeta_i(z) = P(Y_i > z).$$

Following Lustri et al. [8], then,

$$\mu_i' \sim -\int_{\mu_i}^{\infty} \zeta_i(z)\, dz \text{ and } \zeta_i(\mu_i) \sim \frac{\mu_i''}{\mu_i'}.$$

Also, let

$$\varphi_i = E(Y_i | F_{i-1}).$$

Then,

$$\varphi_i' \sim \int_{\mu_i}^{\infty} P(Y_i > z | F_{i-1}) dz.$$

The notation μ_i' and μ_i'' represent the first and second derivatives of μ_i with respect to i. Additionally, the notation $u_i \sim w_i$ indicates that the asymptotic behaviors of u_i and w_i are the same.

Remark 2.5. One can notice that

$$\int_{\mu_i}^{\infty} \zeta_i(z)\, dz = E(Y_i - \mu_i)_+,$$

where $x_+ = \max(x, 0)$. Then,

$$\mu_i' \sim -E(Y_i - \mu_i)_+.$$

Also, a numerical approximation of the equation

$$\zeta_i(\mu_i) \sim \frac{\mu_i''}{\mu_i'}$$

is given by

$$\frac{\mu_i - 2\mu_{i-1} + \mu_{i-2}}{\mu_i - \mu_{i-1}} = \zeta_i(\mu_{i-2}).$$

To derive the asymptotic Snell envelopment, it is assumed that the variance of h_i is small. To survey this assumption more precisely, notice that

$$h_i = \max(Y_i, E(h_{i+1}|F_i)).$$

Therefore,

$$E(h_i) \geq E(h_{i+1}).$$

Given F_i, the covariance of $h_{i+1} + h_i$ and $h_{i+1} - h_i$ is positive. Thus,

$$E((h_{i+1} + h_i)(h_{i+1} - h_i)|F_i) \geq E((h_{i+1} + h_i)|F_i)E((h_{i+1} - h_i)|F_i).$$

Therefore, it is seen that

$$E(h_{i+1}^2 - h_i^2|F_i) \geq (h_i + E(h_{i+1}|F_i))(E(h_{i+1}|F_i) - h_i).$$

Before the $\arg\max$ of h_i, say τ, we know that

$$E(h_{i+1}|F_i) = h_i,$$

and hence,

$$E(h_{i+1}^2|F_i) \geq E(h_i^2|F_i).$$

By taking the expectation on both sides, it is concluded that

$$E(h_{i+1}^2) \geq E(h_i^2).$$

So,

$$\text{var}(h_{i+1}) \geq \text{var}(h_i).$$

Indeed, before τ, the variance of $h_\tau = Y_\tau$ is the maximum of var(h_i). After τ (i.e., $i \geq \tau$), if $h_i = Y_i$, then their variances are equal, and if

$$h_i = E(h_{i+1}|F_i),$$

then

$$\text{var}(h_{i+1}) \geq \text{var}(h_i).$$

Two cases happen: if for all $i \geq \tau$,

$$\text{var}(h_{i+1}) \geq \text{var}(h_i),$$

then the variance of all h_i is less than the variance of $h_n = Y_n$, and if for some i, $h_i = Y_i$, then their variances are equal. So to ensure that the variance of h_i is small, it is enough to ensure that the variance of Y_i is small.

Remark 2.6. A necessary condition for the variance of Y_i to converge to zero is

$$Y_n \prod_{j=i+1}^{n} (1-\gamma_j)\Delta_j^{-1} \to^p 0,$$

as $n \to \infty$, for each $i > 0$. Notation $\to^p$ stands for convergence in probability. To see that, notice that

$$Y_{i+1} = (a_{i+1}Y_i + b_{i+1})\Delta_{i+1} \geq a_{i+1}Y_i\Delta_{i+1}.$$

Therefore,

$$\frac{Y_i}{Y_{i+1}} \leq (1-\gamma_{i+1})\Delta_{i+1}^{-1}.$$

By solving this recursive, it is seen that

$$\frac{Y_i}{Y_n} \leq \prod_{j=i+1}^{n-1} (1-\gamma_{j+1})\Delta_{j+1}^{-1}.$$

Notice that, assuming f_1 is close to f_0, before the change, Δ_{j+1}^{-1} is close to 1 and after the change it is close to zero. Under H_{1n}, for large n, Y_n is bounded by 20. Thus, Y_i^2 is bounded by an upper bound that converges to zero in probability. Therefore,

$$\text{var}(Y_i) \leq E(Y_i^2)$$

is small.

3. Concluding remarks

In this paper, the change point detection in AR(1) process is studied. The Bayesian and non-Bayesian settings are proposed. The change point is considered the first point at which a process is optimized. Therefore, the optimal stopping techniques are applicable. The underlying process for change point analysis is AR(1), because the likelihood ratio process and posterior process is better approximated and the Snell envelopment procedure is better done. However, methods applied here can be extended to AR(p) process, straightforward. Some advantages of this paper are described as follows:

1. The probability of n-th point being the change point is represented as a recursive relation which shows the effects magnitude of change, change point locations, and variances of sequence in the performances of proposed method.

2. In the non-Bayesian setting, the logit of probability existing a change at a specified time point, is decomposed additively to its lag and logarithm of likelihood ratio, which shows the random walk structure of logit function.
3. In the non-Bayesian setting, the Snell auxiliary process is solved and represented by recursive functions with a known structure.
4. Also, the Lustri method which gives an alternative method for derivation of Snell process and then finding the location of change point becomes simple.
5. For all methods, the accuracies of proposed methods and corresponding limit theorems are given. There are some similarities with other works in this filed:
6. Following Veeravalli and Banerjee [5], probabilities of changes are proposed, recursively, which may be served as an early warning system to alarm the existence of possible future changes. Although, their logits are decomposed additively and form a random walk process with suitable properties to find the location of changes points.
7. The method of Lustri et al. [8] in optimal stopping field is developed to change point detection.
8. Snell method is proposed in optimal stopping problems is proposed for change point detection.

References

[1] Li X, Ma J. Non-central Student-t mixture of Student-t processes for robust regression and prediction. In: Intelligent Computing Theories and Application: 17th International Conference, ICIC 2021, Shenzhen, China, August 12–15, 2021, Proceedings, Part I. Springer; 2021. p. 499-511.

[2] Shah V. Optimal stopping problems: autonomous trading over an infinite time horizon. Imperial College London, Department of Mathematics; 2020.

[3] Ma L, Grant AJ, Sofronov G. Multiple change point detection and validation in autoregressive time series data. Statistical Papers. 2020;61:1507-28.

[4] Cappello L, Padilla OHM, Palacios JA. Scalable Bayesian change point detection with spike and slab priors. arXiv preprint. 2021;arXiv:2106.10383.

[5] Veeravalli VV, Banerjee T. Quickest change detection. In: Academic Press Library in Signal Processing. vol. 3. Elsevier; 2014. p. 209-55.

[6] Øksendal B. Stochastic Differential Equations. 6th ed. Springer; 2003.

[7] Pan J, Liu Y, Shu J. Gradient-based parameter estimation for a nonlinear exponential autoregressive time-series model by using the multi-innovation. International Journal of Control, Automation and Systems. 2023;21(1):140-50.

[8] Entwistle HN, Lustri CJ, Sofronov GY. On asymptotics of optimal stopping times. Mathematics. 2022;10(2):194.

[9] Fathan A, Delage E. Deep reinforcement learning for optimal stopping with application in financial engineering. arXiv preprint. 2021;arXiv:2105.08877.