

# Study Note of BSPT

Haixi Li

April 23, 2020

## 1 Some Notations

- $\epsilon$ : OLS residuals
- $\zeta$ : OLS residuals, a sequence of random variable that is subject to structural breaks:  $\zeta^t = \{\zeta_0, \dots, \zeta_t\}$
- $Z$ : random variable of the random process,  $Z^t = \{Z_0, Z_1, \dots, Z_t\}$  and  $Z = \{Z_0, \dots, Z_t, \dots\}$
- $\mathcal{Z}$  denotes the space that contains the random variable
- $\kappa$ : structural breaking time
- $\tau$ : optimal stopping time
- $T$ : the sequential stopping procedure that consists of a sequence of stopping times:  $T = \{\tau_0, \tau_1(\zeta^1), \tau_2(\zeta^2), \tau_3(\zeta^3), \dots, \tau_t(\zeta^t) \dots\}$ .

## 2 The Problem

Consider a dynamic regression model in the following form:

$$y_t = x_t' \beta_t + \epsilon_t, \quad (1)$$

where  $x_t$  is a  $k$  dimensional vector of regressors and  $\beta_t$  is a  $k \times 1$  vector of regression coefficients, and the error terms  $\epsilon_t$  is i.i.d white noise. Assume in the history  $t \in \{1, \dots, n\}$ , the coefficients in Equation () are constant and equal to  $\beta_0$ . We want to monitor new data from  $n + 1$  onward to test if there is any structural break happening since. The null hypothesis is

$$\beta_t = \beta_0 \forall t \quad (2)$$

against the alternative that from some unknown time  $\kappa > n$ , the coefficients changes to  $\beta_1 \neq \beta_0$ . To be more precise, we assume the following statistical process:

$$\begin{aligned} H_0 : y_t &= x_t' \beta_0 + \epsilon_t \text{ for } t = 0, 1, 2, \dots, \kappa - 1, \\ H_1 : y_t &= x_t' \beta_1 + \epsilon_t \text{ for } t = \kappa, \kappa + 1, \dots \end{aligned}$$

In practice, practitioners estimate the model with historical data, and start monitoring the out of sample performance of the estimated model with newly arrived data. Let  $\hat{\beta}^n$  denote the OLS estimates with historical data up to  $t = n$ . Under the null, the residuals can be approximated as

$$\begin{aligned} \zeta_t &= \epsilon_t - x_t' (\hat{\beta}^n - \beta_0) \\ &= \epsilon_t - x_t' \left( \left( \sum_{i=0}^n x_i' x_i \right)^{-1} \sum_{i=0}^n x_i' \epsilon_i \right) \\ &\approx \epsilon_t. \end{aligned}$$

The approximation is reasonable when  $\hat{\beta}^n$  is estimated consistently. Under the alternative, the residuals can be approximated as

$$\begin{aligned}\zeta_t &= \epsilon_t - x_t'(\hat{\beta}^n - \beta_1) \\ &= \epsilon_t - x_t'(\beta_0 - \beta_1 + (\sum_{i=0}^n x_i' x_i)^{-1} \sum_{i=0}^n x_i' \epsilon_i) \\ &\approx \epsilon_t - x_t'(\beta_0 - \beta_1) \\ &\approx \epsilon_t - \mu.\end{aligned}$$

### 3 Bayesian Sequential Probability Test

#### 3.1 The Stochastic Process

There are two different distribution  $f_0, f_1(\theta)$ , we have a piece wise i.i.d. random sequence  $\{\zeta_0, \zeta_1, \dots\}$ , and there is a random variable, structural break time  $\kappa$ , before  $\kappa$  we have  $\zeta_t \sim f_0$  and after  $\kappa$  we have  $\zeta_t \sim f_1(\theta)$ . The structural break time  $\kappa$  follows geometric distribution. Let  $\zeta^t = \{\zeta_0, \zeta_1, \dots, \zeta_t\}$

$$\zeta_t \sim \begin{cases} f_0 = \mathcal{N}(0, \sigma^2) & \text{for } t = 1, 2, \dots, \kappa - 1 \\ f_1 = \mathcal{N}(\mu, \sigma^2) & \text{for } t = \kappa, \kappa + 1, \dots \end{cases} \quad (3)$$

The set of parameters  $\theta = \{\mu, \sigma\}$ , which is included in parameter space  $\theta \in \Theta$ .

The structural break time  $\kappa$  follow geometrical distribution:

$$P(\kappa = k) = \begin{cases} \pi & \text{if } k = 0 \\ (1 - \pi)\rho(1 - \rho)^{k-1} & \text{if } k = 1, 2, \dots \end{cases} \quad (4)$$

#### 3.2 The Probability Measure of the Process

Let  $\Pi_0(\theta)$  denote the prior distribution over  $\Theta$ , and  $\Pi_t(\theta)$  denote the posterior distribution of the parameters given information  $\mathcal{I}_t$ . So the prior includes two components: the geometric prior distribution for  $\kappa$  and prior distribution  $\Pi_0(\theta)$  over the parameter space  $\Theta$ .

Denote  $\zeta^t = \{\zeta_1, \dots, \zeta_t\}$  as the history of random variable  $\zeta_t$  and for each set  $A = \{\omega: \zeta_1 \leq x_1, \dots, \zeta_t \leq x_t\}$ , given  $\theta \in \Theta$

$$\begin{aligned}P_\theta(A) &= \pi P_\theta^1(A) + (1 - \pi) \sum_{i=0}^{t-1} \rho(1 - \rho)^i P^0\{\zeta_1 \leq x_1, \dots, \zeta_i \leq x_i\} \\ &\quad \times P_\theta^1\{\zeta_{i+1} \leq x_{i+1}, \dots, \zeta_t \leq x_t\} + (1 - \pi)(1 - \rho)^t P^0(A)\end{aligned}$$

The  $P^0$  and  $P_\theta^1$  are the probability measures underlying the null distribution and the alternative distribution given  $\theta \in \Theta$  respectively. More generally, the probability measure given distribution  $\Pi$  over parameter space  $\Theta$  is defined as:

$$P^\Pi(A) = \int_{\Theta} P_\theta^\pi(A) d\Pi(\theta) \quad (5)$$

#### 3.3 The Decision Problem

##### 3.3.1 The Objective Function and its Recursive Form

The loss function given parameter  $\theta$  is given by the following equation:

$$L(\theta, \tau) = E_\theta\{1(\tau < \kappa) + c(\tau - \kappa)^+\}. \quad (6)$$

Shiryayev (1963) solved this problem given the parameter  $\theta$ . When  $\theta$  is unknown, the Bayes risk function for any stopping time  $\tau$  became:

$$r^{\Pi_0}(\pi, \tau) = E^{\Pi_0} L(\theta, \tau), \quad (7)$$

where  $\Pi_0$  is the prior distribution over the parameter space  $\Theta$ . The Bayes risk function will be derived by taking the expectation over the parameter space with the probability distribution defined over the parameter space. The expectation  $E^{\Pi_0}$  is defined by probability measure defined by Equation (5).

Given the expectation, the loss function can be converted to

$$r^{\Pi_0}(\pi, \tau) = E^{\Pi_0} (1 - \pi_\tau + c \sum_{k=0}^{\tau-1} \pi_k) \quad (8)$$

At time  $t = 0$ , the expectation is taken with the prior  $\Pi_0$ . But the distribution  $\Pi_t$  over the parameter space is an evolving object, this problem does not show up in with known alternative or simple alternative which is solved by Shiryayev (1963).

We will need a sequential procedure which is a sequence of stopping times  $T = \{\tau_0, \tau_1, \dots\}$ , the stopping time is index by time to reflect the information available at the time. The stopping is simply the time to reject the null in favor of the alternative and stop sampling, it is a random function of information available at the time, it can be formally stated as

$$N(Z) = \min\{t \geq 0 : \tau_t(Z^t) = 1\} \quad (9)$$

To find the recursive presentation of the optimization problem, define the value function of continuation from any time  $t$  on with prior distribution  $\Pi_t$  as

$$v^{\Pi_t}(\pi, \tau) = \inf_{\tau \in \mathcal{T}} E^{\Pi_t} (1 - \pi_\tau + c \sum_{k=t}^{\tau-1} \pi_k) \quad (10)$$

To show this recursively, we need to went through the following notation:

$$v^{\Pi_0}(\pi, \tau) = \inf_{\tau \in \mathcal{T}} E^{\Pi_0} (1 - \pi_\tau + c \sum_{k=0}^{\tau-1} \pi_k) \quad (11)$$

At  $t = 0$ , if  $\tau = 0$ , then there is no delay and only false alarm:  $1 - \pi_0$ . if we decide to continue, then that means  $\tau \geq 1$ , so the expected delay will be  $c\pi_0$  and plus the expected value of continuation. At  $t = 1$ , we have:

$$v^{\Pi_1}(\pi, \tau) = \inf_{\tau \in \mathcal{T}} E^{\Pi_1} (1 - \pi_\tau + c \sum_{k=1}^{\tau-1} \pi_k) \quad (12)$$

If we stop  $\tau = 1$ , that will yield loss  $1 - \pi_1$ , if we decide to continue, the loss will become

In summary, at time zero, the decision maker can stop or continue, which is reflected in following equation:

$$v^{\Pi_0}(\pi) = \min\{1 - \pi, c\pi + E^{\Pi_0} v^{\Pi_1}(\pi' | \pi)\} \quad (13)$$

Define the continuation value function as

$$v^{\Pi_t}(\pi) = \inf_{\tau \in \mathcal{T}} E^{\Pi_t} (1 - \pi_\tau + c \sum_{k=t}^{\tau-1} \pi_k) \quad (14)$$

In recursive format:

$$v^{\Pi_t}(\pi) = \min\{1 - \pi, c\pi + E^{\Pi_t} v^{\Pi_{t+1}}(\pi' | \pi)\} \quad (15)$$

The intuition is at any time  $t$ , the decision maker will face a new problem from  $t$  on. And he will do so every time he receive a new data to update his prior to posterior. In more general form, for any generic prior distribution  $\Pi$ :

$$v^{\Pi}(\pi) = \min\{1 - \pi, c\pi + v^{\Pi}(\pi' | \pi)\} \quad (16)$$

### 3.3.2 The Convergence of the Recursive Form

One can define a finite time problem that  $t \in \{0, 1, 2, \dots, n\}$

$$v_{n-1,n}^{\Pi}(\pi) = \min\{1 - \pi, c\pi + v_n^{\Pi}(\pi' | \pi)\}, \quad (17)$$

Futher more:

$$v_{n-2,n}^{\Pi}(\pi) = \min\{1 - \pi, c\pi + v_{n-1,n}^{\Pi}(\pi' | \pi)\}, \quad (18)$$

We can define more general any time  $t < n$ :

$$v_{t,n}^{\Pi}(\pi) = \min\{1 - \pi, c\pi + v_{t+1,n}^{\Pi}(\pi' | \pi)\}, \quad (19)$$

It is evident that  $v_{t,n}^{\Pi} \leq v_{t+1,n}^{\Pi}$ , since the option of stopping has more value when there are more time left. In the same manner, we can also define a sequence in the same way:  $v_{t,n}^{\Pi}, v_{t,n+1}^{\Pi}, \dots$ , in the same spirit, we have  $v_{t,n}^{\Pi} \geq v_{t,n+1}^{\Pi}$ , in this way we have construct a montone non-increasing sequence, and the limit of which is well defined

$$v_t^{\Pi} = \lim_{n \rightarrow \infty} v_{t,n}^{\Pi}. \quad (20)$$

It is also trivial to show that the objective function is bounded from below, since as long as the exercise started, that  $\tau \geq 1$ , there will be at least one period of expected delay. And also the convergence is point wise for the state variable  $\pi$ .

## References

Shiryaev, A. N. (1963), 'On Optimum Methods in Quickest Detection Problems', *Theory of Probability & Its Applications* **8**(1), 22–46.