Journal of Algorithms and Computation

journal homepage: http://jac.ut.ac.ir

A Note on Change Point Analysis Using Filtering

Reza Habibi*[1]

[1]Iran Banking Institute, Central Bank of Iran, Tehran, Iran.

ABSTRACT

The Kalman-Bucy filter is studied under different scenarios for observation and state equations, however, an important question is, how this filter may be applied to detect the change points. In this paper, using the Bayesian approach, a modified version of this filter is studied which has good and justifiable properties and is applied in change point analysis.

Keyword: Bayesian theorem, Change point, Kalman-Bucy filter

ARTICLE INFO

Article history:
Note
Received 15, December 2022
Received in revised form 18, April 2023
Accepted 02 May 2023
Available online 01, June 2023

AMS subject Classification: 60G35

*Corresponding author: R. Habibi. Email: r_habibi@ibi.ac.ir

1 Introduction

The Kalman-Bucy filter is one of the widely used techniques applied in the filtering scope. It has too applications in control engineering, fault detection, change point analysis, tracking problems proposed in many fields of research such as engineering, statistics and economics. To derive the traditional form of this filter, the observation equation is considered as a regression model without intercept term with time varying slop β_t while the state equation is supposed to be a first order auto-regressive AR(**??**) process. In the current note, following Guo and Meyn (1986), the observation equation is also a discrete time AR(**??**) process for single variable y_t as

$$y_t = \beta_t y_{t-1} + \varepsilon_t; t \geq 2,$$

at which ε_t's are independent random variables with zero mean and finite variance $\sigma^2 < \infty$, denoted by $\varepsilon_t \sim (0, \sigma^2)$. Three scenarios for specifying the state process β_t are considered. The main aim of the current paper is studying the change point analysis in β_t throughout the passing time. Similar problem is studied by Habibi *et al.* (2017) using the least square method and construction of adaptive filters. To this end, three types of β_t are considered including
(i) β_t's are independent and identically distributed with zero mean and variance $v^2 < \infty$, i.e., $\beta_t \sim (0, v^2)$,
(ii) the AR(**??**) process $\beta_t = \alpha\beta_{t-1} + \zeta_t$, where $\zeta_t \sim (0, v^2)$ and $\alpha \neq 0$ and $|\alpha| < 1$, and finally,
(iii) the random walk process $\beta_t = \beta_{t-1} + \zeta_t$, with $\zeta_t \sim (0, v^2)$.
Notice that cases (i) and (iii) are special cases of (ii) with $\alpha = 0$ and $\alpha = 1$, respectively. In the current paper, for simplicity arguments, a simplified version of Kalman-Bucy filter is introduced and used. To find the filter, first, it is assumed that ε_t's and ζ_t's are normally distributed and using the Bayesian theorem, the filter is derived. However, this filter is also achievable using the minimum variance method without the assumption of normality. For extracting the Kalman-Bucy filter, often, the conditional expectation $E(\beta_t|F_t^y)$ is computed and since it is function of β_{t-1}, maximization is taken over β_{t-1}, to remove its effect. Here, it is assumed that β_{t-1} is known, and the filter is derived. Thus, the maximization step is deleted. As soon as, the filter is computed, under the some mild conditions, the assumption of β_{t-1} being known is relaxed.
The corresponding filter is $\widehat{\beta_t} = E(\beta_t|F_t^y, \beta_{t-1})$, where F_t^y is the information set made by $\{y_s,\ s = 1, 2, \ldots, t\}$. Under case ($ii$), it is seen that $\widehat{\beta_t}$ is a weighted linear combination of moment estimator $\alpha\beta_{t-1}$ and unbiased estimator $\frac{y_t}{y_{t-1}}$, where weights of each estimator depends on its accuracy (its variance). Similar results are proposed in credibility theory of insurance field. Similar filters are obtainable in cases (i) and (iii). Using error and scenario analyses and under some mild conditions, the assumption of known β_{t-1} is relaxed and recursive relation for $\widehat{\beta_t}$ based on $\widehat{\beta_{t-1}}$ is proposed. Some advantages of this filter are its simple usage and derivation, its mixture structure with mixing proportion with related to accuracy of each estimator directly, the use of Bayesian method

which itself have good advantages such as updating procedures as soon as new observations come, and useful interpretations. As stated, this filter is obtainable using the minimum variance method without assumption of normality. This filter is achievable for change point analysis using state equation (iv) $\beta_t = (1-J_t)\beta_{t-1} + J_t Z_t$ where $J_t = 1$ if a change has occurred in β_{t-1} and $\beta_t = \beta_{t-1} + Z_t$, while if $J_t = 0$, then no change has happened and $\beta_t = \beta_{t-1}$. Here, Z_t is a sequence of independent variable with $Z_t \sim (\mu_z, \sigma_z^2)$. This model is studied by Yao (1985) for the change point analysis in β_t.
The main idea of this filter comes from the Bayesian approach where the likelihood is derived from the observation equation and the prior is obtained from the state equation. It is seen that this filter has interesting properties and interpretations. This paper is organized as follows. In the next section, the filter proposed and error and scenario analyses are given for cases (i) - (iii). In section 3, the filter is derived for case of (iv) and it is designed to study the change point analysis in β_t. Finally, at the end of this section, conclusions are proposed.

2 Filter features

As stated in introduction, the filter is given by $\widehat{\beta_t} = E(\beta_t | F_t^y, \beta_{t-1})$. Under (ii), it is seen that, given β_{t-1}, the β_t has prior normal distribution with mean $\alpha\beta_{t-1}$ and variance v^2. For the likelihood part of Bayesian formula, notice that given y_{t-1}, the observation y_t has normal distribution with mean $\beta_t y_{t-1}$ and variance σ^2. **2.1. Filter derivation**. The posterior distribution of β_t, given β_{t-1}, y_t and y_{t-1} is proportional to (notations f and $\propto$ serve as density function and algebraic proportion)

$$f(\beta_t|\beta_{t-1}, y_t, y_{t-1}) \propto f\left(\beta_t \mid \beta_{t-1}\right) f(y_t|\beta_t, y_{t-1})$$

$$\propto \exp\{\frac{-1}{2\sigma^2}(y_t - \beta_t y_{t-1})^2\}\exp\{\frac{-1}{2v^2}(\beta_t - \alpha\beta_{t-1})^2\}$$

$$\propto \exp\{\frac{-1}{2\gamma_t^2}\left(\beta_t - \mu_t\right)^2\},$$

2.1 Filter derivation

The posterior distribution of β_t, given β_{t-1}, y_t and y_{t-1} is proportional to (notations f and $\propto$ serve as density function and algebraic proportion)

$$f(\beta_t|\beta_{t-1}, y_t, y_{t-1}) \propto f\left(\beta_t \mid \beta_{t-1}\right) f(y_t|\beta_t, y_{t-1})$$

$$\propto \exp\{\frac{-1}{2\sigma^2}(y_t - \beta_t y_{t-1})^2\}\exp\{\frac{-1}{2v^2}(\beta_t - \alpha\beta_{t-1})^2\}$$

$$\propto \exp\{\frac{-1}{2\gamma_t^2}\left(\beta_t - \mu_t\right)^2\},$$

where $\gamma_t^{-2} = \frac{y_{t-1}^2}{\sigma^2} + \frac{1}{v^2}$ and $\mu_t = \gamma_t^2\left(\frac{y_t y_{t-1}}{\sigma^2} + \frac{\alpha\beta_{t-1}}{v^2}\right)$. Thus, the posterior distribution is normal with mean μ_t and variance γ_t^2. Hence,

$$\widehat{\beta_t} = \mu_t = \frac{\frac{y_t y_{t-1}}{\sigma^2} + \frac{\alpha\beta_{t-1}}{v^2}}{\frac{y_{t-1}^2}{\sigma^2} + \frac{1}{v^2}} =$$

$$w\left(\frac{y_t}{y_{t-1}}\right) + (1-w)(\alpha\beta_{t-1}),$$

where the weight is given by $w = \frac{y_{t-1}^2}{y_{t-1}^2+\theta^2} \in (0,1)$ at which $\theta^2 = \frac{\sigma^2}{v^2} > 0$ is the variance ratio of observation y_t over the variance of β_t (in the sense of prior). In the other words, $\frac{w}{1-w} = \frac{y_{t-1}^2}{\theta^2}$. It is seen that $\widehat{\beta_t}$ is a linear combination of unbiased estimator $\frac{y_t}{y_{t-1}}$ (notice that $\frac{y_t}{y_{t-1}} = \beta_t + \frac{\varepsilon_t}{y_{t-1}}$) and the moment estimator $\alpha\beta_{t-1}$. Notice that if β_{t-1} is known,, in practice, motivated by the state equation, the estimator $\alpha\beta_{t-1}$ is suggested for estimating β_t by practitioners. In this way, the estimation from the observation equation is ignored, while the Bayesian method advises a linear combination of both estimators. Here, two points are given. The following proposition gives a brief summary of above discussion.

Proposition 1. Under the model (ii), the filter is given by

$$w\left(\frac{y_t}{y_{t-1}}\right) + (1-w)(\alpha\beta_{t-1}),$$

Here, $w = \frac{y_{t-1}^2}{y_{t-1}^2+\theta^2}$ at which $\theta^2 = \frac{\sigma^2}{v^2}$.

2.2 Two points

Here, properties of filter are studied.
(a) *Errors analysis*. Here, some points about the error analysis are given.
(a_1) Here, it is seen that the filter can be derived using the minimum variance criterion and by canceling the normality assumption. To this end, suppose that ε_t's and ζ_t's are not normal, however, consider a linear combination $\widehat{\beta_t} = w\left(\frac{y_t}{y_{t-1}}\right) + (1-w)(\alpha\beta_{t-1})$. Notice that $\frac{y_t}{y_{t-1}} = \beta_t + \frac{\varepsilon_t}{y_{t-1}}$ and $\alpha\beta_{t-1} = \beta_t - \zeta_t$. Thus, $\widehat{\beta_t} = w\left(\beta_t + \frac{\varepsilon_t}{y_{t-1}}\right) + (1-w)(\beta_t - \zeta_t) = \beta_t + e_t$, where $e_t = w\frac{\varepsilon_t}{y_{t-1}} - (1-w)\zeta_t$. Thus, $\widehat{\beta_t} - \beta_t = e_t$ and $E(e_t) = 0$ and $E\left(e_t^2\right) = var(e_t) = \frac{w^2\sigma^2}{y_{t-1}^2} + (1-w)^2 v^2$. By minimizing the mean squared error $E(\widehat{\beta_t} - \beta_t)^2 = E(e_t^2)$ with respect to w again the weight of Bayesian method is derived. This fact makes us to relax the normality assumption of errors terms ε_t's and ζ_t's.
(a_2) Here, the assumption of known β_{t-1} is relaxed. Again, notice that $E(\widehat{\beta_t} - \beta_t)^2 \leq v^2(1 + \frac{\theta^2}{y_{t-1}^2})$. Therefore, as $\theta \to 0$ (for example, $\sigma \to 0$), then, $E(\widehat{\beta_t} - \beta_t)^2 \leq v^2$. As soon

as, $v \to 0$, then $E(\widehat{\beta_t} - \beta_t)^2$ tends to zero. In this cases, the $\widehat{\beta_t}$ is a consistent estimator for β_t. Also,

$$\widehat{\beta_t} = w\left(\frac{y_t}{y_{t-1}}\right) + (1-w)\,\alpha\left(\widehat{\beta}_{t-1} - e_{t-1}\right) =$$

$$w\left(\frac{y_t}{y_{t-1}}\right) + (1-w)\,\alpha\widehat{\beta}_{t-1} - (1-w)\,\alpha e_{t-1}.$$

For the cases of $\sigma, v \to 0$, then $E(e_{t-1}^2) \to 0$, thus, $e_{t-1} \approx 0$, with probability one. Therefore,

$$\widehat{\beta_t} = w\left(\frac{y_t}{y_{t-1}}\right) + (1-w)\,\alpha\widehat{\beta}_{t-1}.$$

(a_3) Here, the analysis variance procedure is proposed. Notice that $y_t - \beta_t y_{t-1} = y_t - \widehat{\beta_t} y_{t-1} + (\widehat{\beta_t} - \beta_t)y_{t-1}$ and this fact that $E(y_t - \beta_t y_{t-1})^2 = \sigma^2$. About the first term, by replacing $\widehat{\beta_t}$ in this term, one can see that

$$y_t - \widehat{\beta_t} y_{t-1} = y_t - y_{t-1} w\left(\frac{y_t}{y_{t-1}}\right) - y_{t-1}\,(1-w)\,(\alpha\beta_{t-1}) =$$

$$= (1-w)\,\{y_t - \alpha\beta_{t-1} y_{t-1}\} =$$

$$= (1-w)\left\{y_t - (\beta_t - \zeta_t)y_{t-1}\right\} =$$

$$(1-w)\left\{\varepsilon_t + \zeta_t y_{t-1}\right\}.$$

Thus, $E(y_t - \widehat{\beta_t} y_{t-1})^2 = (1-w)^2\left(\sigma^2 + v^2 y_{t-1}^2\right) = v^2(1-w)^2(\theta^2 + y_{t-1}^2)$. By replacing $w = \frac{y_{t-1}^2}{y_{t-1}^2 + \theta^2}$, it is seen that

$$E(y_t - \widehat{\beta_t} y_{t-1})^2 = \sigma^2\,(1-w).$$

About the second term, it is seen that $\left(\widehat{\beta_t} - \beta_t\right)y_{t-1} = y_{t-1}e_t$. Thus, $E(y_{t-1}e_t)^2 = y_{t-1}^2 var(e_t)$. Again, it easy to see that $E(\left(\widehat{\beta_t} - \beta_t\right)y_{t-1})^2 = \sigma^2 w$.

From, the analysis of variance perspective, it is seen that, as w approaches to one (zero), then the variance of second (first) term has effective role in $E(y_t - \beta_t y_{t-1})^2$.

(b) *Scenario analysis*. Here, some points about the scenario analysis are given.

(b_1) Here, the effect of extreme values of hyper-parameter θ^2 is studied. Notice that as θ^2 goes to the infinity; i.e., as the variance (accuracy) of observation y_t gets large (small) with respect to the variance (accuracy) state β_t, then the weight assigned to unbiased estimator is larger than the moment estimator. Conversely, as θ^2 tends to zero, then w goes to zero.

(b_2) Also, notice that, all above results can be extended to the cases of (i) and (iii), by assuming $\alpha = 0$ and $\alpha = 1$, respectively.

(b_3) To study the effect of outlier data y_{t-1} in w, consider the function $w(z) = \frac{z}{z+\theta^2}$, $z > 0$. This function is increasing in z. Thus, as z approaches to infinity, then $w(z)$ converges to 1. So, if there is an outlier at time $t-1$, then $w(z)$ goes to the 1 and this fact may cause the results become spurious.

3 Filter under Changing β_t

Under the model (iv), then, $\beta_t = (1 - J_t)\beta_{t-1} + J_t Z_t$. Let $\pi_t = P\left(J_t = 1 \mid y_t\right)$ be the posterior of probability of having change in β_t at time t, where using the Bayesian theorem, we have,

$$\pi_t = \frac{pf\left(y_t \mid J_t = 1\right)}{pf\left(y_t \mid J_t = 1\right) + (1-p)f\left(y_t \mid J_t = 0\right)}.$$

Notice that, the logit function of π_t, *i.e.*, $\operatorname{logit}(\pi_t) := \log\left(\frac{\pi_t}{1-\pi_t}\right)$ is given by

$$\operatorname{logit}(\pi_t) = \operatorname{logit}(p) + \log(\Lambda),$$

where is the likelihood ratio of $\Lambda = \frac{f(y_t|J_t=1)}{f(y_t|J_t=0)}$. It is easy to see that, under the normality assumption, given β_{t-1}, y_{t-1} and $J_t = 0$, then, y_t has normal distribution with mean $\beta_{t-1}y_{t-1}$ and variance σ^2, and under the assumption of $J_t = 1$, then it is normally distributed with mean $\mu_z y_{t-1}$ and variance $\sigma^2 + \sigma_z^2 y_{t-1}^2$. To derive the filter, notice that

$$\beta_t = \beta_t^* = \begin{cases} \beta_{t-1} & \text{if } J_t = 0, \\ \beta_{t-1} + Z_t & \text{if } J_t = 1. \end{cases}$$

Let $\mu^* = E(\beta_t^*)$ and $\sigma^{2*} = \operatorname{var}(\beta_t^*)$. Here, the filter is given by

$$\widehat{\beta_t^*} = w^*\left(\frac{y_t}{y_{t-1}}\right) + (1 - w^*)\mu^*,$$

where $w^* = \frac{y_{t-1}^2}{y_{t-1}^2 + \theta^{2*}}$ at which $\theta^{2*} = \frac{\sigma^2}{\sigma^{2*}}$. The following lemma, gives the values of μ^* and σ^{2*}.

Lemma 1. $\mu^* = \beta_{t-1} + (1-\pi_t)\mu_z$ and $\sigma^{2*} = \pi_t(1-\pi_t)\mu_z^2 + \pi_t\sigma_z^2$.

Proof. One can see that

$$E\left(\beta_t^* \mid J_t\right) = \begin{cases} \beta_{t-1} & \text{if } J_t = 0, \\ \beta_{t-1} + \mu_z & \text{if } J_t = 1. \end{cases}$$

Then, $\mu^* = E\left(E\left(\beta_t^* \mid J_t\right)\right) = \pi_t\beta_{t-1} + (1-\pi_t)(\beta_{t-1} + \mu_z) = \beta_{t-1} + (1-\pi_t)\mu_z$. Again,

$$\operatorname{var}\left(\beta_t^* \mid J_t\right) = \begin{cases} 0 & \text{if } J_t = 0, \\ \sigma_z^2 & \text{if } J_t = 1. \end{cases}$$

Using the formula $\sigma^{2*} = var\left(E\left(\beta_t^* \mid J_t\right)\right) + E(\operatorname{var}\left(\beta_t^* \mid J_t\right))$, it is seen that

$$\sigma^{2*} = \pi_t(1-\pi_t)\mu_z^2 + \pi_t\sigma_z^2.$$

Proposition 2. Under the model (iv), the filter is given by

$$\widehat{\beta_t^*} = w^* \left(\frac{y_t}{y_{t-1}} \right) + (1 - w^*)\, \mu^*,$$

where $\mu^* = \beta_{t-1} + (1 - \pi_t)\mu_z$ and $\sigma^{2*} = \pi_t (1 - \pi_t)\, \mu_z^2 + \pi_t \sigma_z^2$. Here, $w^* = \frac{y_{t-1}^2}{y_{t-1}^2 + \theta^{2*}}$ at which $\theta^{2*} = \frac{\sigma^2}{\sigma^{2*}}$.

4 Conclusions

The changing behavior of parameters of models and studying its effects on a modified version of Kalman-Bucy filter is studied. It is seen that, this modified filter have good properties and it is applicable in change point analysis.

References

[1] Guo, L., and Meyn, S. P. Adaptive control for time-varying systems: a combination of martingale and Markov chain techniques. International Journal of Adaptive Control and Signal Processing 3 (1989), 1-14.

[2] Habibi, H., Howard, I., and Habibi, R. Bayesian Sensor Fault Detection in a Markov Jump System. Asian Journal of Control 19(2017) 1465–1481.

[3] Yao, Y. C. Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. Annals of Statistics 12 (1984), 1434-1447.