Journal of Probability and Statistical Science
24(1), 1-15 Feb. 2026

Applications of Hamilton Filter in Regime Switching Models

Reza Habibi
Iran Banking Institute
Central Bank of Iran, Iran

ABSTRACT

Regime switching models play important roles in the fields of economics and finance. Hamilton filter is applied to diagnosis the regime switching models. It is re-write based on logit models and the properties of the new model are surveyed. Time varying case of Hamilton filter is studied and its sensitivity analysis is studied. Finally, a concluding remarks section is given.

Keywords: Hamilton filter, Logit model, Sensitivity analysis, Time varying

1. Introduction

The process of fault detection plays critical role in the early diagnosis of faults to remove any abnormal progression in a controlled system. This process can detect the possibility of occurrences of faults by an online monitoring scheme. It is an important task in the field of control engineering. Equivalent monitoring processes in the statistical quality control, economics and finance and medicine are the change point analysis, structural breaks detection and surveillance. In practice, when a fault occurs in some components of a system, parameters of a statistical model of a system may change. There are many different mathematical models for studying shift (change) in parameters including change point models, fault detections, threshold models and regime switching models. For a fast review on these approaches, see Isermann (2006).

A system which has time varying parameters from a regime to another (i.e., has the regime shifting property) can be modeled by a regime switching model. The latent state process which governs on this time-variation phenomenon is a short memory Markov process. That is, the value of state process depends only on its previous values and independent of former historical information of state process. This feature induces a parameter shift property to the available statistical model. Hence, the value of parameter in a specific regime is effected by its value of previous regime which is fully determined a transition probability matrix of mentioned Markov chain of state process. For a review on Markov chain see Norris (1997). In practice, there is a close connection between change point and regime switching models. There is a vast literature about this issue. For example, Chang *et al*. (2017), using the threshold models, shows a link between change point models and regime switching model. Also, similar to their work, by search in the literatures, it is seen that Yao (1984), used a sequence of independent binary dummy variable indicating the existence of a change to study the possible future change points in a step noisy function setting. As an extension of Yao's (1984) work, his problem may be re-studied by assuming the binary variable id a Markov chain process.

The regime switching models belong to the family of non-linear time series models. Nonlinear time series models have influential applications in economics, data analysis and control engineering. From the late of 1970s up to now, different types of non-linear time series analysis are presented. One of main member of family of non-linear models is the regime switching model. It also has vast applications in finance, econometrics, insurance sciences, statistics and electrical engineering. For comprehensive review on regime-switching models

- Received September 7, 2025, in final form February 2026.
- Reza Habibi (corresponding author) is affiliated with Iran Banking Institute, Central Bank of Iran, Tehran, Iran
habibi1356@yahoo.com

and their applications in the field of engineering, say, control of queuing system, repair-replace models, radar tracking problems, analyzing networked systems, see Isermann (2006). Similar to the multiple change points setting, the main idea in the heart of regime switching models is that its parameters can change (shift) over time from a specific regime to the next one. However, in spite of multiple change point models, this shifting in the parameters occurs according to a state Markov process.

There are two important problems throughout studying of a regime switching problem, namely, estimations of transition probability matrix and parameters within different regimes. There are many different accurate methods for estimating transition matrix and parameters during analyzing a regime switching problem, however, in this field, the filtering theory, based on probability approach, plays important role. Hamilton (2005) proposed a filter to estimate the transition probabilities of hidden states as well as the unknown parameters of statistical model within a regime. In the current paper, first this filter is reviewed and some new results, based on odds ratios, are presented. Next, some extensions, applications, simulations and discussion are proposed.

Let $x_k = \mu(s_k) + \sigma z_k$ be a regime switching process where z_k's are independent and identically distributed with zero mean and standard deviation 1, i.e., (0,1) random variables, $0 < \sigma < \infty$, and s_k be a hidden Markov chain with states $s_k \in \{0,1\}$ with transition probabilities

$$p_{ij} = \mathrm{P}(\mathrm{s_{k+1}} = \mathrm{j}|\mathrm{s_k} = \mathrm{i}); \mathrm{i}, j = 0,1.$$

The initial mean of observations is $\mu(0) = \mu_0$ and after the first change has occurred, the mean of observations is $\mu(1) = \mu_1 \neq \mu_0$. Hamilton (2005) filter proposes a useful setting for detecting and predicting future possible regimes, using the conditional probability p_k, given by

$$p_k = P(s_k = 1|\mathfrak{F}_{k-1}).$$

Here, notation $\mathfrak{F}_{k-1}$ stands for the information set up to time $k-1$. It is easy to see that

$$p_k = \frac{p_{11}p_{k-1}f_1(x_k) + p_{01}(1-p_{k-1})f_0(x_k)}{l_k},$$

at which $l_k = p_{k-1}f_1(x_k) + (1-p_{k-1})f_0(x_k)$. In terms of odds ratio of $\theta_k = \frac{p_k}{q_k}; k \geq 1$, this filter is summarized as

$$\theta_k = \frac{p_{11}\theta_{k-1}\Delta_k + p_{01}}{p_{10}\theta_{k-1}\Delta_k + p_{00}}, p_k = \frac{\theta_k}{1+\theta_k},$$

where $\Delta_k = \frac{f_1(x_k)}{f_0(x_k)}$ is the likelihood ratio of two densities f_0, f_1 and

$$q_k = 1 - p_k = \frac{p_{10}p_{k-1}f_1(x_k) + p_{00}q_{k-1}f_0(x_k)}{l_k}.$$

Interested readers may follow Menard (2002), for the application of odds-ratio in the log-linear models. Notice that the likelihood function computed at k-th observation x_k is given by

$$l_k = p_{11}p_{k-1}f_1(x_k) + p_{01}(1-p_{k-1})f_0(x_k) + p_{10}p_{k-1}f_1(x_k) + p_{00}q_{k-1}f_0(x_k).$$

The density functions f_0, f_1 are distributions densities of x_k's within consecutive regimes. Unknown parameters of these densities such as μ_0, μ_1 and σ are estimated by maximizing the log-likelihood function of unknown parameters based on observations $x_k; k = 1,2,\ldots,n$, given by $\sum_{k=1}^{n} \log(l_k)$.

The rest of paper is organized as follows. In the next section, some new results including Hamilton's filter approximation, sensitivity analysis and its parameter estimation are discussed. Also, the case is considered that the probabilities of transition matrix depend on the duration that the state process remains at it. Finally, at the end of this section, a state space format of regime switching model is developed and model is extended to the case of state space models and features of Hamilton's filter are studied, in this case. This part may also be considered as state process estimation, in the presence of regime switching patterns, as extension of Yao's (1984) work. Throughout the section 3, the change point analysis is reviewed as the regime switching model and the Hamilton's filter is applied in this case. Finally, conclusions are given in section 4.

2. Regime identification

In this section, Hamilton's filter approximation, its parameter estimation, diagnosing the regime limits and sensitivity analysis are studied. To the best knowledge of authors, these results are not presented, before.

(a) Filter approximation

Hereafter, filter approximation is proposed about this filter. To this end, consider the odds ratio $\theta_k = \frac{p_{11}\theta_{k-1}\Delta_k + p_{01}}{p_{10}\theta_{k-1}\Delta_k + p_{00}}$. Notice that as $\Delta_k \to \infty$, then $\theta_k \to \frac{p_{11}}{p_{10}}$ and $p_k \to p_{11}$. If $\Delta_k \to 0$, then $\theta_k \to \frac{p_{01}}{p_{00}}$ and $p_k \to p_{01}$. Hereafter some other representations for θ_k is given. It is easy to see that, based on odds ratios $\beta = \frac{p_{11}}{p_{10}}, \alpha = \frac{p_{01}}{p_{00}}$, then

$$\theta_k = \frac{\beta\theta_{k-1}\Delta_k + \alpha\vartheta}{\theta_{k-1}\Delta_k + \vartheta}$$

where $\vartheta = \frac{p_{00}}{p_{10}}$. One can see that, as the transition matrix P is close to "near the identity", following Yin and Krishnamurthy (2005), then, $\theta_k \approx \theta_{k-1}\Delta_k + 1$. Let $\gamma_k = \frac{\theta_k}{\alpha}$, be the scaled filter. Then,

$$\gamma_k = \frac{\beta\Delta_k\gamma_{k-1} + \vartheta}{\alpha\Delta_k\gamma_{k-1} + \vartheta}$$

If $\Delta_k \to 0$, then and if $\Delta_k \to \infty$, then $\gamma_k \to \frac{\beta}{\alpha}$. If $\alpha = \beta$, i.e., (determinant of P is zero), then, detecting the change point is too difficult.

Above, it was shown that as $\Delta_k \to 0$ or $\Delta_k \to \infty$, then θ_k is close to α or β, respectively. It is interesting to study the rates of convergence. To this end, using the Taylor expansion, it is seen that as $\Delta_k \to 0$, then, $\theta_k = \alpha + \frac{(\beta-\alpha)\theta_{k-1}\Delta_k}{\vartheta}$, and as $\Delta_k \to \infty$, then, $\theta_k = \beta + \frac{(\alpha-\beta)\vartheta}{\theta_{k-1}\Delta_k}$.

The all parameters are kept in both of formula to show that how the rate of convergence and converging scheme depends to these parameters.

As follows, an interesting approximation of $\zeta_k = \log(\theta_k)$ and its application, in diagnosing regime limits, is presented. The ζ_k is the logit of p_k (Focardi *et al.*, 2002). Before going ahead, first, the following lemma is given.

Lemma 1. Let $f(x) = \log(ax + b)$, where a, b, x are positive numbers. Then,

$$f(x) \approx \log(a+b) + \frac{a}{a+b}\log(x).$$

Proof. Notice that $f(x) = g(y) = \log(ae^y + b), y = \log(x)$. Notice that $g(y=0) = \log(a+b)$ and $g'(y=0) = \frac{a}{a+b}$. Using the Taylor expansion about $y = 0$, see McMullen (2018), it is seen that

$$f(x) = g(y) \approx \log(a+b) + \frac{a}{a+b}y.$$

Using the above lemma, it is seen that

$$\zeta_k = \log(p_{11}\theta_{k-1}\Delta_k + p_{01}) - \log(p_{10}\theta_{k-1}\Delta_k + p_{00}) \approx$$
$$\log(p_{11}\Delta_k + p_{01}) + \frac{p_{11}\Delta_k}{p_{11}\Delta_k + p_{01}}\zeta_{k-1} -$$

$$\log(p_{10}\Delta_k + p_{00}) - \frac{p_{10}\Delta_k}{p_{10}\Delta_k + p_{00}}\zeta_{k-1}.$$

Thus, $\zeta_k = A_k\zeta_{k-1} + B_k$, where

$$\begin{cases} A_k = \dfrac{p_{11}\Delta_k}{p_{11}\Delta_k + p_{01}} - \dfrac{p_{10}\Delta_k}{p_{10}\Delta_k + p_{00}}, \\ B_k = \log\left(\dfrac{p_{11}\Delta_k + p_{01}}{p_{10}\Delta_k + p_{00}}\right). \end{cases}$$

Notice that as Δ_k goes to zero (∞), then A_k tends to zero (0) and B_k tends to $\log(\alpha)$ ($\log(\beta)$). Here, $\log(\alpha)$ and $\log(\beta)$ may be interpreted as the logit of p_{01} and p_{11}, respectively. Throughout a specific regime, Δ_k is large (converges to infinity) or small (converges to zero), thus, ζ_k is close to the $\log(\alpha)$ or $\log(\beta)$. However, when a specific regime shifts to another regime, then small (large) Δ_k changes to the large (small) ones. Thus, as soon as regime shifts, then, ζ_k jumps from $\log(\alpha)$ ($\log(\beta)$) to the $\log(\beta)$ ($\log(\alpha)$). This method may be applied to diagnosis the regime limits.

Here, following Yin and Krishnamurthy (2005), a differential equation is derived for θ_k. One can see that

$$\theta_k - \theta_{k-1} = \left(\beta - 1 - \frac{(\alpha-\beta)}{\vartheta}\Delta_k\right)\theta_{k-1} + \alpha - \beta.$$

That is, θ_k is a discrete realization of continuous process θ_t, $t > 0$, derived by differential equation

$$\frac{d\theta_t}{dt} = \left(\beta - 1 - \frac{(\alpha-\beta)}{\vartheta}\Delta_t\right)\theta_t + \alpha - \beta.$$

If $\alpha = \beta$, then $\frac{d\theta_t}{dt} = (\beta - 1)\theta_t$.

The above differential equation depends on stochastic process Δ_k. Here, its sampling properties are studied. Its moments are proposed as follows:

$$m_0 = E_{f_0}(\Delta_k) = \int_{-\infty}^{\infty} f_0(x)\frac{f_1(x)}{f_0(x)}dx = 1$$

$$m_1 = E_{f_1}(\Delta_k) = \int_{-\infty}^{\infty} f_1(x)\frac{f_1(x)}{f_0(x)}dx = \int_{-\infty}^{\infty}(\frac{f_1(x)}{f_0(x)})^2 f_0(x)dx = E_{f_0}(\Delta_k^2).$$

$$E_{f_1}(\Delta_k - 1) = E_{f_0}(\Delta_k^2) - E_{f_0}^2(\Delta_k) = \sigma_0^2 = var_{f_0}(\Delta_k) \geq 0,$$

Thus, $m_1 \geq 1$. Also, $\sigma_1^2 = var_{f_1}(\Delta_k) = E_{f_1}(\Delta_k^2) - E_{f_1}^2(\Delta_k)$ and

$$E_{f_1}(\Delta_k^2) = \int_{-\infty}^{\infty}(\frac{f_1(x)}{f_0(x)})^2 f_1(x)dx = \int_{-\infty}^{\infty}(\frac{f_1(x)}{f_0(x)})^3 f_0(x)dx = E_{f_0}(\Delta_k^3)$$

Let $\varepsilon_k^j = \frac{\Delta_k - m_j}{\sigma_j}$, $j = 0,1$ and $k \geq 1$. For large R, it is seen that, as soon as all standardized observations ε_k^j belong to a specific regime, i.e., (j is fixed for all index of observations), then $R^{-0.5}\sum_{k=1}^{[Ru]}\varepsilon_k^j$ converges to the some stochastic differential equation, see Hamilton (2005).

(b) Sensitivity analysis

Hereafter, the sensitivity analysis of θ_k with respect to the parameters θ_{k-1}, Δ_k, p_{00} and p_{11} are studied. To this end, by some algebraic manipulations, it is seen that

Table 1: Various derivatives of θ_k

$\frac{d\theta_k}{d\theta_{k-1}} = \frac{\det(P)\,\Delta_k}{(p_{10}\theta_{k-1}\Delta_k + p_{00})^2}$,
$\frac{d\theta_k}{d\Delta_k} = \frac{\det(P)\,\theta_{k-1}}{(p_{10}\theta_{k-1}\Delta_k + p_{00})^2}$,
$\frac{d\theta_k}{dp_{11}} = \frac{\theta_{k-1}\Delta_k(1+\theta_{k-1}\Delta_k)}{(p_{10}\theta_{k-1}\Delta_k + p_{00})^2} > 0$,
$\frac{d\theta_k}{dp_{00}} = \frac{-(1+\theta_{k-1}\Delta_k)}{(p_{10}\theta_{k-1}\Delta_k + p_{00})^2} < 0$.

Here, the determinant of matrix P is given by

$$\det(P) = p_{11}p_{00} - p_{10}p_{01} = \\ p_{11}p_{00} - (1-p_{11})(1-p_{00}) = \\ p_{11} + p_{00} - 1 = tr(P) - 1,$$

is the determinant of transition matrix $P = (p_{ij})_{i,j=0,1}$ and $tr(P)$ is the trace of P. The sign of $\frac{d\theta_k}{dp_{11}}$ and $\frac{d\theta_k}{dp_{00}}$ are positive and negative, respectively. It means that, as soon as p_{11}, and p_{00} gets large, then θ_k gets large and small, respectively. However, the sign of $\frac{d\theta_k}{d\theta_{k-1}}$ and $\frac{d\theta_k}{d\Delta_k}$ depends to the sign of $\det(P) = p_{11} + p_{00} - 1$. If $p_{11} + p_{00} > 1$, then as Δ_k (θ_{k-1}) gets large, then, $\frac{d\theta_k}{d\theta_{k-1}}$ ($\frac{d\theta_k}{d\Delta_k}$) gets large. If $p_{11} + p_{00} < 1$, converse results happens. If $p_{11} + p_{00} = 1$, then θ_k does not change if θ_{k-1} or Δ_k varies. Also, as Δ_k (θ_{k-1}) $\to \infty$ or 0 then $\frac{d\theta_k}{d\theta_{k-1}}$ ($\frac{d\theta_k}{d\Delta_k}$) converges to zero. Finally, notice that as Δ_k or θ_{k-1} converges to zero, then both of $\frac{d\theta_k}{dp_{11}}$ and $\frac{d\theta_k}{dp_{00}}$ converges to zero. Then, as Δ_k or θ_{k-1} converges to ∞, then $\frac{d\theta_k}{dp_{11}}$ and $\frac{d\theta_k}{dp_{00}}$ tends to the $\frac{1}{p_{10}^2}$ and $-\frac{1}{p_{00}^2}$, respectively. The following proposition summarizes the above discussion.

Proposition 1. Let $p_k = P(s_k = 1|\mathfrak{F}_{k-1})$. Consider the odds ratio $\theta_k = \frac{p_k}{1-p_k}$ and $\zeta_k = \log(p_k)$. Then, (1)-(2) are correct.

1) $\theta_k = \frac{p_{11}\theta_{k-1}\Delta_k + p_{01}}{p_{10}\theta_{k-1}\Delta_k + p_{00}}, p_k = \frac{\theta_k}{1+\theta_k}$ and $\zeta_k = A_k\zeta_{k-1} + B_k$, where $A_k = \frac{p_{11}\Delta_k}{p_{11}\Delta_k + p_{01}} - \frac{p_{10}\Delta_k}{p_{10}\Delta_k + p_{00}}$ and $B_k = \log\left(\frac{p_{11}\Delta_k + p_{01}}{p_{10}\Delta_k + p_{00}}\right)$. Within a regime, ζ_k is close to the $\log(\alpha)$ or $\log(\beta)$. While regime shifts, it jumps from $\log(\alpha)$ to $\log(\beta)$, or conversely.

2) If $p_{11} + p_{00} > 1$ (<1, =1), $\frac{d\theta_k}{d\theta_{k-1}}$ ($\frac{d\theta_k}{d\Delta_k}$) gets large (small, remains unchanged) as Δ_k (θ_{k-1}) gets large. As p_{00} (p_{11}) gets large, then θ_k gets small (large).

(c) Parameter estimations

As follows, the estimation of unknown parameters are discussed. First, suppose that f_0, f_1 are known functions. Thus, Δ_k is known. It is understood that θ_k is constant over a specified regime. That is, $\theta_k \approx \theta_{k-1}$. Then, by substituting this equation in the relation between θ_k, θ_{k-1}, it is seen that

$$h(\theta_k) = p_{10}\Delta_k\theta_k^2 + p_{00}\theta_k - p_{11}\Delta_k\theta_k - p_{01} \approx 0.$$

Then, the derivative of $h(\theta_k)$ with respect to the θ_k is zero. Thus,

$$h'(\theta_k) = 2p_{10}\Delta_k\theta_k + p_{00} - p_{11}\Delta_k = 0.$$

In this case, then,

$$\Delta_k = \frac{p_{00}}{-2p_{10}\theta_k + p_{11}}.$$

Therefore,

$\log(\Delta_k) = \log(p_{00}) - \log(-2p_{10}\theta_k + p_{11})$.

Notice that function $g(x) = \log(a - 2bx)$ is well approximated by

$g(x) \approx \log(a) - \frac{2b}{a}x$,

using the Taylor expansion. Thus,

$$\log(\Delta_k) \approx \log(p_{00}) - \log(p_{10}) + \frac{2p_{10}}{p_{11}}\theta_k.$$

That is, within a specific regime, the logarithm of Δ_k is constant. Indeed, the sign of $\log(\Delta_k)$ is different from the current regime to the next one. In this way, the limits of different regimes are obtained. Then, the elements of transition matrix P are easily estimated, empirically. For example, for calculating p_{10} and p_{01}, it is enough to count the numbers of shifts from regime 0 to 1 and 1 to 0 over the number of observations n. Also, the empirical estimate of p_k is the number of state 1 of underlying regime switched process among observations x_1 to x_k throughout n observations divided by k. As soon as, regimes and p_k are derived, then using the likelihood function $\sum_{k=1}^{n} \log(l_k)$, all unknown parameters of f_0, f_1 are estimated.

For unknown Δ_k's, suppose that f_0, f_1 are indexed by parameters μ_0 and μ_1. The parameter μ_0 is known and $\mu_1 = \mu_0 + \delta$. Supposing f_0, f_1 belong to the class of normal distributions, then,

$$\log(\Delta_k) = \frac{-1}{2\sigma^2}\{(x_k - \mu_1)^2 - (x_k - \mu_0)^2\} = \frac{-1}{2\sigma^2}\{(x_k - \mu_0 - \delta)^2 - (x_k - \mu_0)^2\} = \frac{2\delta(x_k - \mu_0) - \delta^2}{2\sigma^2}.$$

As σ is small, and x_k belongs to the regime with $E(x_k) = \mu_0$, then, $(x_k - \mu_0)$ is negligible and $\log(\Delta_k)$ is negative. Also, notice that

$$E(\log(\Delta_k)) = \frac{-\delta^2}{2\sigma^2} < 0.$$

However, if $E(x_k) = \mu_1$, then

$$E(\log(\Delta_k)) = \frac{\delta^2}{2\sigma^2} > 0,$$

or $(x_k - \mu_0)$ is large and $\log(\Delta_k)$ is positive. If μ_0 is unknown, it may be replaced by the average of third or fourth observations. To estimate δ, it is enough to plot the sequential mean or rolling mean of observations. As soon as a jump is appeared, the difference of means of consecutive regimes estimates δ. For non-normal cases, notice that, as $\delta \to 0$, then, $\log(\Delta_k)$ is well approximated by

$$\delta \frac{\partial}{\partial \mu_0} \log\left(f_{\mu_0}(x_k)\right) + \frac{\delta^2}{2} \frac{\partial^2}{\partial \mu_0^2} \log\left(f_{\mu_0}(x_k)\right).$$

Hence, to diagnosis the jumps (regimes), it is enough to visualize the sign of above function, using estimated parameters μ_0 and δ.

Remark 1. One can see that, up to now, proposed results can be extended straightforwardly, to the general case that x_k's are independent and they are distributed according to the density function f_{φ_k}, where $\varphi_k = \mu_0$ or μ_1, depending x_k belongs to which regime. Yet, these results may be extended to the case of linear models, such as regression models similar to the case of Yin and Krishnamurthy (2005).

Remark 2. In practice, to avoid large variability of $\log(\Delta_k)$, in practice, it is replaced by its rolling mean defined by mowing average

$$\frac{\sum_{j=k+1}^{k+l}\log(\Delta_j)}{l},$$

for some length of rolling window l. This rolling mean is positive or negative for all observations during a specific regime. As soon as, a regime shifts, the sign of this rolling mean is changed. In our problem, since the change has happened in the mean of observations, rolling means of observations can help to detect the different regimes.

Remark 3. In practice, following Durland and McCurdy (1994) the probabilities of transition matrix may be time varying and they depend on the duration state process remains in a specific regime. In this case, the important concept is the length of a regime which state process lives on it which is

$av_i = \frac{1}{1-p_{ii}}, i = 0,1.$

Also, notice that, it is assumed that

$logit(p_{ii}) = \log\left(\frac{p_{ii}}{1-p_{ii}}\right) = a_i d_i + b_i, i = 0,1.$

Here, d_i is the length of i-th regime. Focardi *et al*. (2019) used the same argument for modeling local trends of long-lived stock return series. Usually, the maximum likelihood method is applied to estimate parameters $a_i, b_i, i = 0,1$.

The following practical instruction summarizes the above discussion.

Practical instruction:

To detect regimes the below steps are followed:

(1) Based on observations $x_1, \dots, x_k$, compute the likelihood ratio Δ_k (or use its approximation, using estimated parameters). Detect different regimes from the sign shifting of Δ_k.

(2) Compute the elements of transition matrix P, and probability of p_k. Use the likelihood function to estimate the unknown parameters.

(3) Using new observations $x_{k+i}, i \geq 1$, compute the future odds ratios $\theta_{k+i}, i \geq 1$ and detect future possible regimes using probabilities $p_{k+i} = \frac{\theta_{k+i}}{1+\theta_{k+i}}$.

(4) State space model. Because of control engineering problems type, the Hamilton's filter is extended to the state space models in the presence of regime switched state process. To review these types of models, interested readers are referred to Zeng and Wu (2010). This section may be considered as the extension of Yao's (1984) work. The mean function $\mu(s_k)$, in the previous section, is given by

$$\mu(s_k) = \mu_0 s_k + \mu_1(1 - s_k).$$

This formulation, although, allows different regimes, however, there are only two values for the state process, i.e., 0,1. Hence, there are only two possible values for mean functions throughout regimes which are μ_0 and μ_1. However, in practice, the mean function may take more than two values during regime shifts. One way, to solve this difficulty is to consider a state process with more than two states values (number of realizations of state process). In this setting, the dimension of state process can be high, and, then, working with it is too hard. Notice that,

$$\mu(s_k) = \mu_0 + s_k\delta, k \geq 1,$$

where $\delta = \mu_1 - \mu_0$ is the magnitude of change. To overcome the abovementioned problem (state process being high dimensional), it is enough to assume that, at time k, the magnitude of change depends on k, that is, $\delta = \delta_k$ and μ_0 is replaced with μ_{k-1}, the value of mean function at $k - 1$. More accurately, let

$$\mu(s_k) = \mu_k \text{ and } \mu_k = \mu_{k-1} + s_k\delta_k.$$

In this way, throughout a specific regime no change occur and $s_k = 0$ and $\mu_k = \mu_{k-1}$. As soon as a regime switch occurs, then, $s_k = 1$, and a new value μ_k for mean function is made. However, because of variation of δ_k, it can take more than a fixed new value μ_1 which does not correspond with practice, in applied engineering problems. In this way, the dimension of state process is two, yet. As stated in remark 2, in these cases, the time varying transition probabilities are needed. Next, using the above descriptions and remarks, the problem is proposed, more rigorously, as follows. The observation process obeys the formulation

$$x_k = \mu_k + \sigma z_k,$$

where σ and z_k are defined as section 1. Here,

$$\mu_k = \mu_{k-1} + s_k\delta_k,$$

where δ_k is the magnitude of change process and s_k is $0-1$ Markov process at which $s_k = 1$ shows the existence of regime-shift. The transition probability matrix is P.

Remark 4. Yao (1984) applied a same formulation to model a noisy step function, however, there, the binary indicator variable which shows the existence of a change is not Markov process, like our problem. Here, the occurrence of shift may have effect on another, a stylized fact which happens, in practice, frequently. Another difference between the current work and Yao's work id, that Yao assumes that the magnitude of change is random and uses the Bayesian setting, which is a complex and time-consuming computational approach and needs true and accurate non-sampling prior information, which in the current work, these requirements are not needed.

Remark 5. Modeling the magnitude of change, instead of working with the new value for mean function, for example μ_1, has many advantages. For example, it permits to study the local regime shifts (when the magnitude of change is small) and densities functions approximations, as it is stated before remark 1 which is too useful when the new values are unknown. It may be formulated as function sample size of observations. In this section, it is assumed, that, the magnitude of change depends the time and a new formulation is born.

The following proposition extends the result of proposition 1, in the current case.

Proposition 2. By changing μ_0 with μ_{k-1} and μ_1 with $\mu_{k-1} + \delta_k$, in the proposition 1, results of proposition 1 and abovementioned practical instruction are kept by replacing Δ_k with

$$\Delta_k^* = \frac{f_{\mu_{k-1}+\delta_k}(x_k)}{f_{\mu_{k-1}}(x_k)}.$$

Proof. Apply the arguments of proposition 1, straightforwardly.

Remark 6. An important parameter in the above proposition is the magnitude of change δ_k. There are many different formulations for modeling δ_k. For example, Yao (1989) considered a Bayesian setting and considered the magnitude of change as a realization of independent normal random variables. The local change point model assumes that the magnitude of change converges to zero. The gradual change model assumes that the magnitude of change is a linear function of time and the sharp change supposes that the magnitude of change is a deterministic unknown parameter which is estimated throughout the inferential process. As soon as, the type of change point model is declared, then the practical instruction of part *(c)* is applicable, here.

Following Yin and Krishnamorthy (2017), a differential equation is derived for $e_k = \hat{\mu}_k - \mu_k$. Consider the following equations

$$\begin{cases} x_k = \mu_k + \sigma z_k \\ \mu_k = \mu_{k-1} + s_k\delta_k \\ \hat{\mu}_k = \hat{\mu}_{k-1} + \lambda_k(x_k - \hat{\mu}_{k-1}) \end{cases}$$

Combining the above equations together, it is seen that

$$e_k - e_{k-1} = -\lambda_k e_{k-1} - (1-\lambda_k)s_k\delta_k + \sigma z_k.$$

One can see that the above equation is

$$de_t = -(\lambda_t e_t + (1-\lambda_t)s_t\delta_t)dt + \sigma dw_t$$

3. Change point as regime switching

Chang *et al*. (2017) showed that, if there is a latent factor generated by a first order autoregressive AR(1) state process which plays the role of threshold variable, then, the change point analysis of parameters of a linear time series model is a type of regime switching model. In applied cases, there is no general method to find the mentioned (latent/even observed) state process, however, in some special cases, this goal is achievable. Some of special cases are given as follows

a) Change test statistic. Chen and Gupta (2005), page 20, showed that the test statistic for detecting a change in the mean of observations, in a sequential manner, is an AR(1) process. Thus, this process may be considered as threshold variable. Indeed, to find the AR state process playing the role of threshold variable, adaptive estimates of parameter of interest (to study its change point analysis) produces the factor AR(1) state process.

b) Rolling estimates. The other way to find the threshold variable is using the rolling estimates. The rolling estimates of interesting parameters, often, are used to diagnosis the stability of statistical parameter (model) over the time, see Zivot and Wang (2005), page 313. These types of estimates are derived by computing, for example, the maximum likelihood estimations of unknown parameters throughout a rolling window with a given size over the windowed sample of original sample. According to the Brockwell and Davis (2005), page 11, the rolling estimate constitutes a moving average process. Also, there is connectivity between moving average and autoregressive process. Thus, in this way, the autoregressive state process is defined and then the abovementioned state process and subsequent threshold process may be defined.

c) Adaptive estimates. Usually, techniques such as stochastic approximation and exponential moving average (EWMA) are useful to obtain the adaptive estimates, consequently, the adaptive filtering scheme, of interested parameters. Here, results of Yin and Krishnamurthy (2005) are reviewed, in our case.

Here, motivated by Chang *et al*. (2017), the at most one change point (AMOC) model is analyzed using the regime switching approach. To this end, suppose that $\{x_i\}_{i=1}^n$ are independent random variables such that $\{x_i\}_{i=1}^{k_0}$ have common density function f_{μ_0} $(k_0 \le n)$ and the remaining observations $\{x_i\}_{i=k_0+1}^n$ come from another distribution with density f_{μ_1} where $\mu_1 \ne \mu_0$. Here, the aim is to estimate the change point k_0 and parameters before and after the change μ_1, μ_0 and some scale parameters such as $\sigma > 0$. To coordinate the current problem with the model of proposed in section 2, it is assumed that both $f_{\mu_j}, j = 0,1$ have a location-scale family of distribution formats; *i.e.*,

$$f_{\mu_j}(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu_j}{\sigma}\right), j = 1,2,$$

for some location parameters $\mu_j, j = 1,2,$ and scale parameters σ. Assume that Φ is the distribution function of density ϕ. Although, theoretical results proposed here are devoted for location-scale family of distributions, however, all results are extendable for a general class of distributions, straightforwardly.

Usually, the adaptive estimates of location parameters are given by running a stochastic approximation method as

$$\hat{\mu}_k = (1-\lambda_k)\hat{\mu}_{k-1} + \lambda_k x_k, k \ge 1,$$

for some forgetting factor $\lambda_k \in (0,1)$. For example, when the location parameter plays the role of population mean of observations, then $\lambda_k = k^{-1}$ and $\hat{\mu}_k = \frac{\sum_{i=1}^k x_i}{k}, k = 1,2,\dots,n$. Again, let $\mathcal{F}_k$ be the information set of $\{x_1,\dots,x_k\}$. Also, similar notations of section 1, assume that $s_k =$

1 denotes a change has occurred at time k and $s_k = 0$ stands no change has happened at k-th time point. Here, it is assumed that $\hat{\mu}_k$ plays the role of latent autoregressive process of Change *et al*. (2017), although, it is observed. Indeed, it is assumed that $s_k = 0$ if $\hat{\mu}_k = \hat{\mu}_{k-1}$. Notice that

$$\hat{\mu}_k = \hat{\mu}_{k-1} + \lambda_k(x_k - \hat{\mu}_{k-1}).$$

Thus, $\hat{\mu}_k = \hat{\mu}_{k-1}$, if and only if $\lambda_k|x_k - \hat{\mu}_{k-1}| < \varepsilon$, for some pre-determined threshold $\varepsilon > 0$, where ε is a positive small number.

Notice that $p_k = P(s_k = 1|\mathrm{F}_{k-1})$ is proportional to $p_{11}p_{k-1}f_1(x_k) + p_{01}(1 - p_{k-1})f_0(x_k)$. Also, notice that

$$p_k = P(s_k = 1|\mathrm{F}_{k-1}) = P(s_k = 1|\hat{\mu}_{k-1}) =$$
$$P_{s_k=1}(\lambda_k|x_k - \hat{\mu}_{k-1}|< \varepsilon|\hat{\mu}_{k-1}) =$$
$$P_{s_k=1}\left(x_k \le \hat{\mu}_{k-1} + \frac{\varepsilon}{\lambda_k}\right) - P_{s_k=1}\left(x_k \le \hat{\mu}_{k-1} - \frac{\varepsilon}{\lambda_k}\right) =$$
$$\Phi\left(\frac{\hat{\mu}_{k-1} - \mu_1}{\sigma} + \frac{\varepsilon}{\sigma\lambda_k}\right) - \Phi\left(\frac{\hat{\mu}_{k-1} - \mu_1}{\sigma} - \frac{\varepsilon}{\sigma\lambda_k}\right).$$

As soon as, $\frac{\varepsilon}{\sigma\lambda_k} \to 0$, thn, using the Taylor expansion, it is seen that

$$p_k \approx \frac{2\varepsilon}{\sigma\lambda_k}\phi\left(\frac{\hat{\mu}_{k-1} - \mu_1}{\sigma}\right).$$

Computing the odds ration of, $\theta_k = \frac{p_k}{q_k}$, given by

$$\theta_k = \frac{2\varepsilon\sigma^{-1}\lambda_k^{-1}\phi\left(\frac{\hat{\mu}_{k-1} - \mu_1}{\sigma}\right)}{1 - 2\varepsilon\sigma^{-1}\lambda_k^{-1}\phi\left(\frac{\hat{\mu}_{k-1} - \mu_1}{\sigma}\right)}.$$

Again, the practical instruction of section 2, is applicable, here.

4. Concluding remarks.

Hamilton filter is applied to diagnosis the regime switching models. It is re-write based on logit models and the properties of the new model are surveyed. Time varying case of Hamilton filter is studied and its sensitivity analysis is studied. Finally, a concluding remarks section is given.

References

[1]. Brockwell, P. J. and Davis, R. A. (2005). *Time series: theory and methods*. Springer. USA.

[2]. Chang, Y., Choi, Y. and Park, J. Y. (2017). A new approach to model regime switching. *Journal of Econometrics* 106, 127-143.

[3]. Cheng, J. and Gupta, A. K. (2012). *Parametric statistical change point analysis: with applications to genetics, medicine, and finance*. Birkhauser. USA.

[4]. Durland, J. M., and McCurdy, T. H. (1994). Duration-dependent transitions in a Markov model of U.S. gnp growth. *Journal of Business & Economics* 12, 179-288.

[5]. Focardi, S. M., Fabozzi, F. J., and Mazza, D. (2019). Modeling local trends with regime shifts models with time varying probabilities. *International Journal of Financial Analysis* 66, 25-40.

[6]. Hamilton, J. D. (2005). Regime-switching models. *Technical report*. Department of Economics, University of California. San Diego, USA.

[7]. Isermann, R. (2006). *Fault-diagnosis Systems: an introduction from fault detection to fault tolerance*. Springer. Berlin.

[8]. McMullen, C. (2018). *Essential calculus skills practice workbook with full solutions*. Zishka publishing. USA.
[9]. Menard, S. W. (2002). *Applied logistic regression*. Wiley. USA.
[10]. Norris, J. R. (1997). *Markov chains*. Cambridge University Press. UK.
[11]. Yao, Y. C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *The Annals of Statistics* 12, 1434-1447.
[12]. Yin, G.G. and Krishnamurthy, V. (2005). Least mean square algorithms with Markov regime-switching limit. *IEEE Transactions on Automatic Control* 50, 577 - 593.
[13]. Zeng, Y., and Wu, S. (2010). *State-space models: applications in economics and finance*. New York: Springer.
[14]. Zivot, E. and Wang, J. (2005). *Modeling financial time series with SPLUS*. Springer. USA.