Computation of ARL on a CUSUM Control Chart for a Long-Memory Seasonal Autoregressive Fractionally Integrated Process with Exogenous Variables

W Peerajit

Department of Applied Statistics, Faculty of Applied Science, King Mongkut’s University of Technology North Bangkok, Bangkok, Thailand

* Corresponding author: wilasinee.p@sci.kmutnb.ac.th

Abstract. The aim of this study is to derive explicit formulas for the average run length (ARL) of a long-memory seasonal autoregressive fractionally integrated model with exogenous variables process on a cumulative sum (CUSUM) control chart with exponential white noise. The efficiency of the proposed explicit formulas was compared with a numerical integral equation (NIE) method via the relative percentage change to determine the accuracies of the ARL derivations. Although both methods were in good agreement, only 1 second of CPU time was required for the explicit formulas while 32-35 minutes were required for the NIE method. Some application examples using real datasets are provided to the detection ability of both methods.

1. Introduction

The control chart, a statistical process control tool, is widely used to monitor the quality characteristics of industrial processes. They are also used in various fields, such as biology, genetics, medicine, and finance [1]. The Shewhart, cumulative sum (CUSUM), and exponential weighted moving average (EWMA) control charts are the most commonly used. The Shewhart control chart is well known for its effectiveness in detecting large shifts in the process mean. On the other hand, the CUSUM and EWMA control charts are better for small-to-moderate shifts in the process mean. The CUSUM control chart has gained significant attention for monitoring uncommon variations in manufacturing and service processes. The CUSUM location chart is also used for monitoring changes in the process mean. Many researchers have offered theories for the CUSUM charts focusing on location charts with normally distributed observations (for a detailed review in this area, see [2]-[3] and the references therein).

The average run length (ARL) is a characteristic commonly used to determine the efficacy of a control chart. It is divided into two states: ARL_0 is used to denote the process that is in control and ARL_1 is used to denote a process that is out of control. Specifically, ARL_0 is the expectation of the observations prior to a false alarm on the control chart indicating that the in-control process has gone out of control, while ARL_1 refers to the expectation of the observations being out of control and the control chart showing an alarm for the out-of-control process. ARL_0 should be sufficiently large and the ARL_1 should be minimal.

Evaluating the solution for the ARL of a control chart can be achieved via various methods. For instance, integral equations using numerical quadrature methods can be used to approximate the ARL of the CUSUM control chart for a process with normally distributed observations [4]. In addition, a finite-state Markov chain approach for finding the ARL of an EWMA control chart was implemented by Brook and Evans [5] to determine the ARL of an EWMA control chart. Meanwhile, Lucas and Saccucci [6] proposed closed-form
formulas for calculating the ARL by using the Markov chain approach to determine the optimal performance of an EWMA control chart, while the Martingale approach was used Sukparungsee and Novikov [7] to analytically derive closed-form equations for the ARL. When close-form formulas are not used, the simplest approach is using Monte Carlo simulation to simply program ARLs and to test their accuracy [8]. However, a great deal of CPU time is required owing to the great number of trajectories in this approach. Busaba et al. [9] proved explicit formulas for the ARL on an exponential CUSUM control chart by deriving them from integral equation only in terms of continuous distributions. Calculating explicit formulas for the ARL can be achieved by using the numerical Gauss Legendre approximation for both EWMA and CUSUM when the process comprises serially correlated exponential white noise observations [10]-[13]. The analysis of explicit formulas is interesting when applied to actual situations with real data. The study of time series with long-memory characteristics has drawn the attention of statisticians. Granger and Joyeux [14] first introduced the autoregressive fractionally integrated MA (ARFIMA(, , )) model with this property. Parameter is the fractional difference operator known as the long-memory parameter with the interval (0, 0.5) [15]. Long-memory processes with periodicity (seasonal ARFIMA) have also been considered [16]. Furthermore, seasonal ARFIMA models with an exogenous variable (ARFIMAX) model have been suggested. In terms of economic forecasting and other fields, cases where an exogenous variable is included in the forecasting model are usually more accurate than ones without it. Subsequently, Ramjee et al. [17] designed an HWMA forecast-based control chart particularly designed for a non-stationary ARFIMA model with autocorrelated data. Rabyk and Schmid [18] recently introduced the EWMA control chart for detecting changes in the mean of a long-memory process with the control chart’s design based on an ARFIMA(p, d, q) process. The objective of this study is to propose and derive explicit formulas for the ARL of a long-memory seasonal AR fractionally integrated with k exogenous variables (SARFIX) model process on a CUSUM control chart having exponential white noise.

2. CUSUM control chart and model for observation

The SARFIX model is a combination of seasonal AR and fractionally differentiated processes/ integration order (D order of the seasonal data) in time-series data. Suppose that , is the sequence for SARFIX(P, D, k), for a non-stationary process, then

\[ \Phi_p(B^S)(1-B^S)^D Y_t = \mu + \sum_{j=1}^{k} \omega_j X_p + \xi_t, \]

where \( \xi_t \) is an observed sequence from exponential distribution \( \xi_t \sim \text{Exp}(\beta) \), \( \mu \) is the constant process mean, \( B \) is a backward-shift operator (i.e., \( B^s Y_t = Y_{t-s} \)), \( S \) is the number of seasonal periods per year (e.g., \( S = 12 \) for monthly data), and \( \Phi_p(B^S) \) are the seasonal AR polynomials in \( B \) of order \( P; S \in \mathbb{N} \), represented by

\[ \Phi_p(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2s} - \ldots - \Phi_p B^{ps} = 1 - \sum_{i=1}^{p} \Phi_i B^{is}, \]

and \((1-B^S)^D\) is the seasonal fractional difference operator that is subjected to a binomial series expansion as follows:

\[ (1-B^S)^D = \sum_{i=0}^{D} \binom{D}{i} (-B^S)^i = 1 - DB^S - \frac{D(1-D)}{2!} B^{2s} - \ldots, \]

where \( D \) is the seasonal fractional difference operator known as the long-memory parameter in the interval (0, 0.5) [14].

By rearranging (1) and using the fractional difference operator in (2), we obtain the polynomial form of the general model in the SARFIX(P, D, k) with k exogenous variables as [19]
\[ Y_t = \mu + \sum_{j=1}^{k} \omega_j X_{j,t} + \xi_t + \left( D_{Y_{t-5}} - \frac{D(D-1)}{2!} Y_{t-25} + \ldots \right) + \left( \Phi_1 Y_{t-5} - D\Phi_1 Y_{t-25} + \frac{D(D-1)}{2!} \Phi_1 Y_{t-35} - \ldots \right) + \ldots + \left( \Phi_{r_{1-5}} Y_{t-5} - D\Phi_{r_{1-5}} Y_{t-(p+1)5} + \frac{D(D-1)}{2!} \Phi_{r_{1-5}} Y_{t-(p+2)5} - \ldots \right); t = 1,2,\ldots \] (3)

where \( X_{j,t}, j=1,2,\ldots,k \) are exogenous variables and \( \omega_j \) are coefficients corresponding to \( k \). The initial values for the SARFIX(\( P,D,k \)) process are \( Y_{t-5}, Y_{t-25}, \ldots, Y_{t-(p+1)5}, \ldots \) equal to 1, \( \xi_t = 1 \), where \( \xi_t \sim Exp(\beta) \), and the seasonal AR coefficient as \(-1 \leq \Phi_1 \leq 1; i = 1,2,\ldots,P\).

The CUSUM statistic based on a long-memory SARFIX(\( P,D,k \)) process is described by the following recursion \([2]\):

\[ Z_t = \text{max}(Z_{t-1} + Y_t - a, 0), t = 1,2,\ldots \] (4)

where \( Y_t \) is the sequence of the SARFIX(\( P,D,k \)) process with exponential white noise, \( a \) is a reference value for the chart, and \( Z_0 \) is the starting value (i.e., \( Z_0 = u \), where \( u \) is the initial value).

The stopping time of the CUSUM control chart is given by

\[ \tau_b = \inf \left\{ t > 0 : Z_t > b \right\}, b > u, \] (5)

where \( b \) is the upper control limit (UCL).

The ARL of the upper-sided CUSUM control chart for a long-memory SARFIX(\( P,D,k \)) process is given by

\[ \text{ARL} = E_0(\tau_b) \] (6)

where \( \theta \) is the change-point time and \( E_0(.) \) is the expectation under the assumption that the change point occurs at time \( \theta \). Meanwhile, \( \text{ARL}_0 = E_0(\tau_b) = L(u) \) denotes the explicit in-control ARL for the SARFIX(\( P,D,k \)) process with initial value \( u \).

### 3. The explicit formulas for the ARL of a SARFIX(\( P,D,k \)) process on a CUSUM control chart

The numerical solution to an integral equation is often used to compute the ARL. Here, some definitions from actual analyses and Banach’s fixed point theorem to assert the existence and uniqueness of the explicit ARL are mentioned.

The unique solution of the integral equation for the ARL is as follows:

\[ L(u) = 1 + \beta e^{\int_{0}^{u} e^{-\theta} d\theta + (1-e^{-\theta})L(\theta) - \beta L(\theta) e^{\int_{0}^{u} e^{-\theta} d\theta + (1-e^{-\theta})L(\theta)} d\theta} - \beta L(0) \] (7)

**Theorem 1** Banach’s fixed point theorem \([20]\)

Let \( (X, d) \) be a complete metric space and \( \mathcal{Y} : X \rightarrow X \). If \( \mathcal{Y} \) is a contraction mapping with contraction constant \( 0 \leq q < 1 \); i.e.,

\[ d(\mathcal{Y}(g), \mathcal{Y}(h)) \leq q d(g,h), \text{ for all } g, h \in X. \] (8)

Next, \( g \in X \), where \( X = C(I) \), exists and is unique such that \( \mathcal{Y}(g) = g \); \( \mathcal{Y} \) has a unique fixed point in \( C(I) \), where \( I \) is a compact interval. Now consider \( \left( C(I), \| \cdot \| \right) \) as a complete metric space where
$C(I)$ is the space of all continuous functions $L(u)$ on interval $I = [0,a]$ endowed with supremum norm $\|L\|_\infty = \sup_{u \in I} |L(u)|$, for every function

**Proof.** Let (7) also be a continuous function. This proves the existence of the solution of the integral equation

$$Y(L(u)) = 1 + \beta e \int_0^b L(z) e^{-\beta z} dz + (1 - e^{-\beta b}) L(0).$$

(9)

Therefore, (7) can be written in the form $Y(\lambda(L(u)) = \lambda(u)$, which proves the existence of the solution of the integral equation by Banach’s fixed point theorem.

**Theorem 2** The operator $Y$ is the contraction on a complete metric space \((C(I), \|\|_\infty)\) with the norm $\|L\|_\infty = \sup_{u \in I} |L(u)|$.

**Proof.** This proves the uniqueness of the solution of the integral equation.

$$\|Y(L_1) - Y(L_2)\|_\infty = \sup_{u \in I} |L_1(u) - L_2(u)|$$

$$\leq \sup_{u \in I} \left\|L_1(0) - L_2(0)\right\| (1 - e^{-\beta b}) + \|L_1 - L_2\|_\infty \beta e \int_0^b e^{-\beta z} dz$$

$$= q\|L_1 - L_2\|_\infty$$

where $q = 1 - e^{-\beta b}$ is a positive constant.

Therefore, the existence and uniqueness of the explicit ARL for SARFIX($P,D,k$) process are guaranteed Banach’s fixed point theorem 2.

Let $\nu = \int_0^b L(z) e^{-\beta z} dz$. $L(u)$ can be written as

$$L(u) = 1 + \beta e \int_0^b L(z) e^{-\beta z} dz + (1 - e^{-\beta b}) L(0).$$

(10)

If $u = 0$ then

$$L(0) = 1 + \beta e \int_0^b L(z) e^{-\beta z} dz + (1 - e^{-\beta b}) L(0).$$

Subsequently, $L(u)$ can take the form
Now, the constant $\nu$ can be found as follows:

$$\nu = \frac{\beta}{0} C(z)e^{-\beta z}dz = \frac{1}{0}(1 + \beta \nu + e^{-\beta \nu}) - e^{\beta \nu} \int_{0}^{\infty} e^{-\beta z}dz$$

$$= (1 + \beta \nu + e^{-\beta \nu}) \left( \frac{1}{\beta} (1 - e^{-\beta \nu}) - b \right)$$

$$\therefore \nu = e^{\theta_0} (1 + e^{-\theta_0})(1 + e^{-\theta_0}) - e^{\theta_0} \left( \frac{1}{\beta} (1 - e^{-\beta \nu}) - b \right)$$

By substituting constant $\nu$ into (10), we obtain

$$L(u) = e^{\theta_0}(1 + e^{-\theta_0}) - \beta \nu - e^{\theta_0}.$$
The ARL results for the explicit formulas and NIE method are reported in Table 1. The parameter values applied for the ARL on a CUSUM control chart were \( a = 3 \) and \( 3.5 \), the desired ARL0 = 370, and in-control parameter \( \beta_1 = 1 \); the magnitude of the change was given as \( \beta_0 \). In-control ARL0 = 370 was the desired level when the CUSUM control chart parameters \( a \) and \( b \) comprising the CUSUM control limit were applied. The calculation of \( b \) revealed an increase in \( a \) as \( b \) decreased.

The performance of the proposed explicit formulas was expressed as the relative percentage (\( \varepsilon \% \)) with respect to the NIE method as follows:

\[
\varepsilon\% = \frac{|L(u) - L(u)_{\text{NIE}}|}{L(u)} \times 100\%.
\]

In addition, a comparison of the CPU times needed to obtain the numerical values for ARL1 was made. When \( \beta_0 = 1 \), the results for \( \beta_0 > 1 \) were consistent with ARL1. Tables 1 report that the results obtained from the proposed explicit formulas were close to those obtained with the NIE, with \( \varepsilon \% \) values of less than 0.25%. For the ARL1 calculation efficiency, the CPU time for the explicit formulas was less than 1 second while 32–35 min was needed for NIE

**Table 1.** Comparison of the ARL for a long-memory SARFIX(\( P, D, k \))s process on a CUSUM control chart where \( \beta_0 = 1, \Phi_1 = 0.2, \alpha_1 = 0.5 \), and ARL0 = 370.

| \( \beta_1 \) | \( a = 3.0, b = 4.5721 \) | \( a = 3.5, b = 3.6612 \) | \( a = 3.0, b = 4.9647 \) | \( a = 3.5, b = 3.8436 \) |
|---|---|---|---|---|
| \( L(u) \) | \( L(u)_{\text{NIE}} \) | \( \varepsilon \% \) | \( L(u) \) | \( L(u)_{\text{NIE}} \) | \( \varepsilon \% \) | \( L(u) \) | \( L(u)_{\text{NIE}} \) | \( \varepsilon \% \) |
| 1.01 | 344.224 | 343.469 | 0.22 | 346.808 | 346.157 | 0.20 | 342.319 | 341.615 | 0.21 |
| | (0.015) | (33.90) | | (0.014) | (33.59) | | (0.014) | (34.82) | |
| 1.03 | 299.221 | 298.595 | 0.21 | 305.989 | 305.379 | 0.20 | 294.403 | 293.835 | 0.19 |
| | (0.014) | (34.30) | | (0.014) | (33.24) | | (0.015) | (34.85) | |
| 1.05 | 261.540 | 261.016 | 0.20 | 271.212 | 270.686 | 0.19 | 254.731 | 254.271 | 0.18 |
| | (0.014) | (34.10) | | (0.014) | (33.34) | | (0.014) | (34.97) | |
| 1.10 | 191.039 | 190.696 | 0.18 | 204.471 | 204.100 | 0.18 | 181.833 | 181.555 | 0.15 |
| | (0.014) | (33.49) | | (0.014) | (33.59) | | (0.014) | (34.67) | |
| 1.30 | 70.518 | 70.436 | 0.12 | 82.267 | 82.151 | 0.14 | 63.179 | 63.134 | 0.07 |
| | (0.014) | (33.61) | | (0.014) | (33.28) | | (0.014) | (34.78) | |
| 1.50 | 35.096 | 35.009 | 0.08 | 42.405 | 42.448 | 0.11 | 30.819 | 30.811 | 0.03 |
| | (0.014) | (32.70) | | (0.014) | (35.48) | | (0.014) | (34.76) | |
| 1.70 | 21.207 | 21.196 | 0.05 | 25.861 | 25.837 | 0.09 | 18.695 | 18.694 | 0.01 |
| | (0.014) | (32.84) | | (0.014) | (33.66) | | (0.014) | (34.58) | |
| 2.00 | 12.503 | 12.499 | 0.03 | 14.944 | 14.984 | 0.07 | 11.281 | 11.281 | 0.01 |
| | (0.014) | (32.64) | | (0.014) | (35.02) | | (0.014) | (34.58) | |
| 3.00 | 5.134 | 5.134 | 0.00 | 5.667 | 5.665 | 0.04 | 4.976 | 4.976 | 0.00 |
| | (0.014) | (32.79) | | (0.014) | (34.54) | | (0.014) | (34.93) | |

The CPU time is in seconds for the explicit formulas and minutes for the NIE method. The numbers in each cell show the ARL value with the CPU time used in the calculation indicated in parentheses.
Figure 1. ARL value of explicit formulas of a long-memory SARFIX(P,D,k)s process on a CUSUM control chart.

Figure 1 show comparisons of the reference values for $a = 3.0$ and 3.5, in which the magnitude of the mean ($\beta$) increased from 1.01 to 3.00, respectively. When comparing the reference values for $a = 3.0$ and 3.5, it was revealed that ARL$^1$ using the explicit formulas decreased for every level in the process mean as well as every reference value parameter ($a$) of the CUSUM control chart. Furthermore, $a = 3.0$ provided the lowest ARL$^1$ results for all magnitudes of process mean shifts with each model and was thus more effective for detecting shifts in the process mean than $a = 3.5$.

5. Application of the proposed explicit formulas with real data

The proposed explicit formulas and the NIE method on the CUSUM control chart were applied for using monthly values of export agricultural products with an exogenous variable which are reported in Table 2. The variable of monthly values obtain from the Office of Agricultural Economics and includes an exogenous variable the US dollar exchange rate with the Thai baht. The observations were collected monthly from January 2013 to April 2019. The dataset was analyzed and fitted to a long-memory SARFIX(1, 0.290731, 1) process with coefficients $\Phi_1 = 0.462373$ and $\omega_1 = 0.188104$. The white noise was tested to be significant of the exponential distribution and provided a mean ($\beta$) = $0.4769 \times 10^6$ for the in-control state. In addition, CUSUM chart parameters, reference value ($a$) = 2 and CUSUM control limit ($b$) = $1.888309 \times 10^6$ was selected to give the desired in-control ARL$^0$ = 370 as calculated using (11).

Table 2. Comparison of the ARLs generated by the explicit formulas and the NIE method for $\beta_0 = 0.4769, \Phi_1 = 0.462373, \omega_1 = 0.188104$, and ARL$^0_0 = 370$.

| Mean changes ($\beta_i \times 10^6$) | $L(a)$ | CPU time (Sec) | $L(a)_{\text{NIE}}$ | CPU time (Min) | $\epsilon\%$ |
|--------------------------------------|--------|----------------|-----------------|----------------|-----------|
| 0.4869                              | 321.867| 0.014          | 321.189         | 34.81          | 0.21      |
| 0.5069                              | 247.555| 0.014          | 247.066         | 33.48          | 0.20      |
| 0.5269                              | 194.192| 0.014          | 193.833         | 33.69          | 0.18      |
| 0.5769                              | 113.931| 0.014          | 113.75          | 34.12          | 0.16      |
| 0.7769                              | 27.321 | 0.014          | 27.296          | 34.11          | 0.09      |
| 0.9769                              | 12.211 | 0.014          | 12.205          | 33.60          | 0.05      |
| 1.1769                              | 7.389  | 0.014          | 7.387           | 34.64          | 0.03      |
| 1.4769                              | 4.627  | 0.014          | 4.626           | 33.52          | 0.02      |
| 2.4769                              | 2.362  | 0.014          | 2.362           | 33.67          | 0.00      |

The results in Table 2 follow a similar trend to those in Table 1. The numerical results from the proposed explicit formulas are very close to the NIE results for all detecting changes in the process mean. However, the CPU time of explicit ARL was actually less than that of the NIE method (less than 1 second of CPU time compared to between 33 to 35 minutes) In conclusion, the explicit formula approach is an alternative for practical applications in detecting changes process mean on a CUSUM control chart.
6. Conclusion
Explicit formulas for ARL of a long-memory SARFIMA($P, D, k$) process with exponential white noise on a CUSUM control chart were presented. The proposed explicit formulas were easy to calculate and the CPU time needed to execute them was significantly less than for the NIE. Their efficacy was also demonstrated with a real economic dataset following a SARFIMA($P, D, k$) process. From the findings, this approach could be used in a variety of financial, economic, and industrial applications involving long-memory SARFIMA($P, D, k$) processes on CUSUM control charts. In addition, the proposed approach for constructing explicit formulas for the ARL could be extended to other control charts.

7. References
[1] Montgomery D C 2000 Introduction to Statistical Quality Control (New York: Wiley) 4th ed
[2] Page ES 1954 Continuous inspection schemes (Biometrika) vol 41 pp 100-115
[3] Woodall, W H and Adams, B M 1993 The Statistical Design of CUSUM Charts (Quality Engineering) vol 5 pp 559-570
[4] S V Crowder 1987 A simple method for studying run-length distributions of exponentially weighted moving average charts (Technometrics) vol 29 pp 401-407
[5] D Brook, and D A Evans 1972 An approach to the probability distribution of the CUSUM Run Length (Biometrika) vol 59 pp 539-549
[6] Lucas JM, Saccucci MS 1990 Exponentially weighted moving average control schemes: properties and enhancements (Technometrics) vol 32 pp 1-29
[7] Sukparungsee S and Novikov AA 2006 On EWMA procedure for detection of a change in observations via martingale approach (An International Journal of Science and Applied Science) vol 6 pp 373-380
[8] S W Robert 1959 Control Chart Test Based on Geometric Moving Averages (Technometrics) vol 1 pp 239-250
[9] Busaba J, Sukparungsee S, Areepong Y, Mittielu G 2012 An analysis of average run length for CUSUM procedure with negative exponential data (Chiang Mai J. Sci.) vol 39 pp 200-208
[10] K Petcharit, S Sukparungsee, and Y Areepong 2015 Exact solution of the average run length for the cumulative sum charts for a moving average process of order $q$ (Science Asia) vol 41 pp 141-147
[11] W Peerajit, Y Areepong, and S Sukparungsee 2018 Numerical Integral Equation method for ARL of CUSUM chart for long-memory process with non-seasonal and seasonal ARFIMA models,” (Thailand Statistician) vol 6 pp 26-37
[12] Peerajit W, Areepong Y and Sukparungsee S 2019 Explicit analytical solutions for ARL of CUSUM chart for a long-memory SARFIMA model (Communications in Statistics-Simulation and Computation) vol 48 pp 1176-1190
[13] Sunthornwat R, Areepong Y 2020 Average Run Length on Cusum Control Chart for Seasonal and Non-Seasonal Moving Average Processes with Exogenous Variables (Symmetry) vol 12 pp 1-15
[14] Granger CWJ, Joyeux R 1980 An introduction to long memory time series models and fractional differencing, (Journal of Time Series Analysis) vol 1 pp 15-29
[15] Hosking JRM 1981 Fractional differencing (Biometrika) vol 68 pp 165-176
[16] A Montanari, R Rosso and M S Taqqu 2000 A Seasonal Fractional ARIMA Model Applied to the Nile River Monthly Flows at Aswan (Water Resources Research) vol 36 pp 1249-1259
[17] Ramjee R, Crato N, Ray B K 2002 A note on moving average forecasts of long memory processes with an application to quality control (International Journal of Forecasting) vol 18 pp 291–297
[18] Rabýk L, Schmid W 2016 EWMA control charts for detecting changes in the mean of a long-memory process (Metrika) vol 79 pp 267–301
[19] Degiannakis S 2008 ARFIMAX and ARFIMAX-TARCH realized volatility modeling (Journal of Applied Statistics) vol 35 pp 1169–1180
[20] Dugundji J Granas A 2003 Fixed point theory (New York: Springer)

Acknowledgments
The author gratefully acknowledges the editor and referees for their valuable comments and suggestions which greatly improve this paper. Besides, we would like to express the gratitude to Thailand Science Research and Innovation via King Mongkut’s University of Technology North Bangkok, Thailand for supporting the research fund with Contract no. KMUTNB-BasicR-64-20.