Received: 26 July 2021 | Revised: 14 October 2021 | Accepted: 29 October 2021 | Published on: 19 November 2021

DOI: 10.1002/nav.22034

RESEARCH ARTICLE

WILEY

Sequential detection of common transient signals in high dimensional data stream

Yanhong Wu[1] | **Wei Biao Wu**[2]

[1]Department of Mathematics, California State University Stanislaus, Turlock, California, USA
[2]Department of Statistics, University of Chicago, Chicago, Illinois, USA

Correspondence
Yanhong Wu, Department of Mathematics, California State University Stanislaus, Turlock, CA 95382, USA.
Email: ywu1@csustan.edu

Funding information
National Science Foundation.

Abstract

Motivated from sequential detection of transient signals in high dimensional data stream, we first study the performance of EWMA and MA charts for detecting a transient signal in a single sequence in terms of the power of detection under the constraint of false detecting probability in the stationary state. Satisfactory approximations are given for the false detection probability and the power of detection. Comparison of EWMA, MA, and CUSUM charts shows that both charts are quite competitive. A multivariate EWMA procedure is considered by using the squared sum of individual EWMA processes and a fairly accurate approximation for the false detection probability is also given. To increase the power of detection, we use the Min-δ procedure by truncating the estimated weak signals. Dow Jones 30 industrial stock prices are used for illustration.

KEYWORDS

EWMA process, false detection probability, power of detection, transient signal detection

1 | INTRODUCTION

Consider a sequence of independent normal random variables $\{X_i\}$ that follow $N(0,1)$ for $i \leq \nu$ and $N(\mu,1)$ for $i > \nu$. To detect a positive change with $\mu > 0$, the traditional Shewhart chart (Shewshart, 1931) makes an alarm when $X_i > c$. That means, the alarm is only made based on the current observation. The CUSUM (Page, 1954), EWMA (Roberts, 1959), and Shiryayev-Roberts (S-R) (Roberts (1966) and Shiryayev (1963)) charts are developed for a quick detection of change. For example, we may focus to find a stopping time τ that minimizes the conditional delay detection time $CADT(\nu) = E[\tau - \nu | \tau > \nu]$ for given $ARL_0 = T$. The optimality of CUSUM and S-R charts has been established under different criteria in terms of ν for a given reference value δ of μ.

From signal detection point of view, the CUSUM and S-R charts assume that the signal will be present forever after the change. However, in the case that the signal just appears in one observation, the Shewhart chart will possess certain optimality as shown in Pollak and Krieger (2013) and Moustakides (2014).

In this communication, we consider the detection of transient signal. That means, the signal only appears in a relatively short period of time with length L, say, from $\nu+1$ to $\nu+L$. An obvious chart is the MOSUM (moving sum) chart (Bauer & Hackel, 1978) that uses partial sum $\sum_{i=k+1}^{k+L} X_i = S_{k,L}$ to detect the signal by raising an alarm at

$$\tau_{MA} = \inf\{n > 0 : S_{n-l,l} > c\} = \inf\left\{n > l : \max_{1\leq k\leq n-l} S_{k,l} > c\right\}$$

where l is a pre-selected window size. Note that the MOSUM chart is indeed equivalent to MA (moving average) chart. General weighted moving average chart is considered in Lai (1974), and Wu (1996) also considered an optimal linear kernel smoother in order to achieve high constant efficiency for unknown strength of signal under the change point model. Several authors have proposed other procedures. Gueppie et al. (2012) proposed to use a window-limited CUSUM procedure under the constraint on the false alarm probability $P_0(\nu < \tau < \nu + L)$ for just one alarm interval. Noonan and Zhigljavsky (2020) studied the power of MA chart by using the corrected diffusion approximation under the constraint

on ARL_0. Tartakovsky et al. (2021) considered a modified CUSUM procedure that is shown to be optimal when the signal length L follows an exponential distribution under the constraint of the conditional detection false alarm probability $\max_\nu P_0[\tau \le \nu + L|\tau > \nu]$.

Here we assume that the transient signal appears far away from the beginning. In notation, we denote by τ_i for $i \ge 1$ as the intervals between consecutive false alarms where the detecting process will start new after each alarm. So $ARL_0 = E_0\tau_1$. Let

$$\kappa = \inf\left\{j > 0 : \tau_1 + \cdots + \tau_j > \nu\right\}.$$

As $\nu \to \infty$, the instantaneous returned (controlled) process at the change time will be at its stationary state. The stationary average delay detection time (SADDT) is defined as

$$SADDT = \lim_{\nu\to\infty} E\left[\tau_1 + \cdots + \tau_\kappa - \nu\right],$$

under the change point model. This criterion is considered in Srivastava and Wu (1993) and Reynolds Jr. and Stoumbos (2004). A more recent discussion by using Markov chain approach is given in Knoth (2021).

By denoting $P^*_\delta(.)$ as the probability measure when the controlled process starts at the stationary state with signal δ, we can write $SADDT = E^*_\delta[\tau]$.

Our main focus is to study the performance of the EWMA and MA charts as the detection procedures for the transient signal with length L. Following the suggestions given in Lai (1995) and Reynolds Jr. and Stoumbos (2004), we define the *power of detection* as $P^*_\mu(\tau \le L)$ for signal length L, and $P^*_0(\tau \le L)$ as the *false detection probability*. Just like in statistical hypothesis test, we evaluate the power of detection under the constraint on false detection probability. Alternatives to the false detection probability have been considered by other authors. For example, Margavio et al. (1995) considered the false alarm rate $P_0[\tau = j|\tau > j-1]$ for CUSUM and EWMA chart. Frisén (1992) also proposed to use the predictive value of an alarm.

The rest of the paper is organized as follows. In Section 2, we first review several charts in terms of SADDT under the single change point model. In Section 3, we study the performance of EWMA and MA charts under the constraint on the false detection probability $P^*_0(\tau \le L)$ in terms of the power of detection $P^*_\mu(\tau \le L)$. Satisfactory approximations are obtained for practical implementation. The comparison between EWMA, MA and CUSUM charts shows that the both procedures perform quite well. In Section 4, we consider a multivariate EWMA procedure by using the square norm of multi-dimensional EWMA process as the detecting rule. Accurate approximation for false detection probability is obtained for practical use. In Section 5, when the signal is sparse, we further propose to use the Max-K and Min-δ procedures by truncating off the weak signal estimations. Dow Jones 30 industrial stock prices are used for illustration.

2 | PERFORMANCE OF CONTROL CHARTS UNDER THE CHANGE POINT MODEL

We give the definitions of several detecting procedures and briefly review their performances in terms of the SADDT $E^*[\tau]$ at a reference value δ for μ with the same $ARL_0 = T$.

1. Shewhart chart: Define
$$\tau_{SH} = \inf\{n > 0 : X_n > c\}.$$
Obviously, $ARL_0 = \frac{1}{1-\Phi(c)}$. Since $1 - \Phi(c) \approx \frac{1}{c}\phi(c)$, thus
$$c \approx \sqrt{2\log(T)},$$
and no matter what ν is, the mean delay detection time
$$SADDT = E^*_\delta[\tau_{SH}] = \frac{1}{1-\Phi(c+\delta)}.$$
2. EWMA chart: Define $Z_n = (1-\beta)Z_{n-1} + \beta X_n$ as the EWMA process. The EWMA chart makes an alarm at
$$\tau_{EW} = \inf\left\{n > 0 : Z_n > b\left(\frac{\beta}{2-\beta}\right)^{1/2}\right\},$$
where $\frac{\beta}{2-\beta} = \lim_{n\to\infty} Var(Z_n)$.
In Srivastava and Wu (1993), the optimal values of b and β are obtained under a continuous time change point model that minimizes the SADDT for given $ARL_0 = T$ at the reference value $\mu = \delta$ where
$$\beta^* = \frac{2c^*\delta^2}{b^{*2}}; \quad b^{*2} \approx 2\log(c^*T),$$
and
$$SADDT \approx \frac{2.4554}{\delta^2}\log(c^*T),$$
where $c^* \approx 0.5117$.
3. CUSUM procedure: For the reference signal strength δ, the CUSUM procedure makes an alarm at
$$\tau_{CS} = \inf\{n > 0 : Y_n = \max(0, Y_{n-1} + X_n - \delta/2) > d\},$$
which gives $d \approx \log\left(\delta^2 T/2\right)/\delta$ for given $ARL_0 = T$ and
$$SADDT \approx \frac{2}{\delta^2}\log\left(\delta^2 T/2\right).$$
4. S-R procedure: The S-R procedure raises an alarm at
$$\tau_{SR} = \inf\left\{n > 0 : R_n = (R_{n-1}+1)e^{\delta X_n - \delta^2/2} > B\right\},$$
with $R_0 = 0$. An accurate approximation for selecting B is $ARL_0 \approx Be^{\delta\rho_+}$ for $\rho_+ = 0.5826$, that is, $B \approx Te^{-\delta\rho_+}$, and
$$SADDT \approx \frac{2}{\delta^2}\log\left(\delta^2 T/2\right).$$
5. Finite Moving Average (FMA) chart: For window length l, define
$$\tau_{MA} = \inf\left\{n > l : \overline{X}_{n-l;l} = (X_{n-l+1} + \cdots + X_n)/l > h\right\}.$$

The selection of optimal h and l for given ARL_0 can be achieved by assuming the signal strength as δ or the signal length as L. An accurate approximation for ARL_0 is given in

Noonan and Zhigljavsky (2020). In the case $h = \delta$, Wu (1996) showed that if $l \approx \frac{2}{\delta^2}\log(T)$, then

$$SADDT \approx \frac{2}{\delta^2}\log\left(\delta^2 T/2\right).$$

That means, the MA chart can be made as efficient at δ as the CUSUM and S-R charts.

When the signal strength μ is unknown, a comparison in terms of SADDT between CUSUM, EWMA, and S-R procedures in the continuous time case is given in Srivastava and Wu (1993). Knoth (2021) conducted a comparison by using the Markov chain approach in terms SADDT and quasi-stationary average delay detection time $\lim_{\nu\to\infty} E[\tau - \nu|\tau > \nu]$.

3 | FALSE DETECTION PROBABILITY AND POWER OF DETECTION

When the signal length is L, a more natural performance measure is the power of detection under the stationary state given by $P^*_\delta[\tau \le L]$. As Lai (1995) suggested, for the transient signal, the proper constraint under null hypothesis will be the false detection probability under the stationary state given by $P^*_0[\tau \le L] \le \alpha$ when $\delta = 0$. Tartakovsky et al. (2021) also noted that this constraint defines a subclass as the constraints on $ARL_0 \le T$.

For technical convenience, we shall assume that $\delta \to 0$ and $L \to \infty$. We first note that for the Shewhart chart,

$$P^*_\delta[\tau_{SH} \le L] = 1 - \Phi^L(c+\delta),$$

with false detection probability $P^*_0[\tau_{SH} \le L] = 1 - \Phi^L(c)$.

3.1 | EWMA procedure

For the EWMA procedure, at the change point ν, we assume that Z_ν follows the stationary distribution $N(0, (\beta/(2-\beta)))$. As shown in Appendix A, this stationary distribution is approximately equivalent to the stationary state distribution for the controlled process as the detecting process is positively recurrent. Thus, the power of detection can be written as

$$P^*_\delta\left[\max_{1\le k\le L} Z_k \ge b(\beta/(2-\beta))^{1/2}\right].$$

The following theorem gives the approximation for the power of detection and its proof is given in Appendix A.

Theorem 1 *Assume $\beta \to 0$, $\beta L \to \infty$ and $b - \delta/(\beta/(2-\beta))^{1/2} \to \infty$. The power of detection is approximated by*

$$\begin{aligned} &P^*_\delta\left[\max_{1\le k\le L} Z_k \ge b(\beta/(2-\beta))^{1/2}\right] \\ &= P^*_0\left[\max_{1\le k\le L}\left(Z_k + \left(1-(1-\beta)^k\right)\delta\right) \ge b(\beta/(2-\beta))^{1/2}\right] \\ &= 1 - \exp\left[-\int_0^{L\beta} m(u)^2(1-\Phi(m(u)))du\right] + o(1). \end{aligned} \tag{3.1}$$

where $m(u) = b - (1-e^{-u})\,\delta/(\beta/(2-\beta))^{1/2}$.

When $\delta = 0$, it gives the false detection probability

$$\begin{aligned} \alpha &= P^*_0\left[\max_{1\le k\le L} Z_k \ge b(\beta/(2-\beta))^{1/2}\right] \\ &\approx 1 - e^{-L\beta b^2(1-\Phi(b))} \approx L\beta b^2(1-\Phi(b)). \end{aligned} \tag{3.2}$$

Remark 1 The approximation for α can be obtained by other methods as discussed in Pickands (1969), Berman (1969), and Qualls and Watanabe (1972)). Here we used the method given in Davis (1982) where it is shown that no matter what the initial state is,

$$\alpha \approx 1 - e^{-L\beta/\int_0^b \phi^{-1}(u)du}, \tag{3.3}$$

by the recurrent property. As a matter of fact, by noting that

$$\begin{aligned} &1/\int_0^b \phi^{-1}(u)du \approx b\phi(b)\left(1 - 1/b^2 + o\left(1/b^2\right)\right) \\ &\approx b^2(1-\Phi(b))\left(1 + o\left(1/b^2\right)\right), \end{aligned}$$

we see that the approximation is equivalent to lemma 3.1 of Davis (1982) up to the second order of $1/b$.

Remark 2 When $\delta = 0$, we can also apply the classical extreme value theory and obtain an approximation for the probability in Equation (3.2). For example, applying Theorem A1 in Bickel and Rosenblatt (1973) with $C = 1$ and $\alpha = 1$, let $Q = L\beta$, we obtain for all $z \in \mathbb{R}$ that

$$\begin{aligned} &P\left((2\log Q)^{1/2}\left(\max_{0\le u\le Q} e^{-u}\mathbf{W}\left(e^{2u}\right) - B(Q)\right) \ge z\right] \\ &\quad \to 1 - \exp(-\exp(-z)), \end{aligned} \tag{3.4}$$

where the centering term

$$\begin{aligned} B(Q) &= (2\log Q)^{1/2} + (2\log Q)^{-1/2} \\ &\quad \left(2^{1/2}\log\log Q - 2^{1/2}\log\pi\right). \end{aligned} \tag{3.5}$$

Let $z = (2\log Q)^{1/2}(b - B(Q))$. Then the probability in Equation (3.1) can be approximated by the right hand side of Equation (3.4). However our numerical study in Table 1 shows that the approximation (3.4) is, as expected, quite inaccurate since the extreme value convergence is quite slow.

Remark 3 For practical implementation of the approximation, we need to correct the boundary b with the overshoot. As $\beta \to 0$, $b\sqrt{\beta/(2-\beta)} \to \infty$ such that $\beta^{3/2}b \to 0$, when Z_{k-1} is close to the boundary b,

$$Z_k \approx Z_{k-1} + \beta X_k.$$

That means, the EWMA process can be approximated by a local normal random walk with drift 0. So we can correct the boundary $b(\beta/(2-\beta))^{1/2}$ with $b(\beta/(2-\beta))^{1/2} - \beta\rho_+$ where

TABLE 1 Simulated and approximated one-sided false detection probability for $\beta L = 5$ and $\delta = 0$

	$L = 500\ (\beta = .01)$	$L = 100\ (\beta = .05)$	$L = 20\ (\beta = .25)$	$L\beta = 5$	
b	Simul (cor)	Simul (cor)	Simul (cor)	No cor	(3.4)
2.5	0.1578 (0.1636)	0.1264 (0.1316)	0.0794 (0.0810)	0.1941	0.1534
3.0	0.0482 (0.0488)	0.0384 (0.0370)	0.0207 (0.0204)	0.0607	0.0657
3.5	0.0113 (0.0109)	0.0083 (0.0078)	0.00366 (0.00385)	0.0143	0.0273
4.0	0.0019 (0.0019)	0.0014 (0.0013)	0.00064 (0.00056)	0.0025	0.0112
4.5	0.00025 (0.00024)	0.00014 (0.00015)	0.000052 (0.000062)	0.00034	0.00459

$\rho_+ \approx 0.5826$ (e.g., Siegmund (1985, [3.29]) is the mean overshoot. That means, to apply Theorem 1, we need to correct b to $b^* = b + \beta\rho_+/(\beta/(2-\beta))^{1/2}$.

Table 1 gives the numerical comparison with simulated values at $\delta = 0$ with $L = 500\,(\beta = 0.01)$, $L = 100\,(\beta = 0.05)$ and $L = 20\,(\beta = 0.25)$ and thus $m = b$ and $L\beta = 5$. The uncorrected approximation is given in the second column from last from Equation (3.2). Listed in the last column are also the approximated values based on Equation (3.4). As we noted that both approximations give overestimated values, particularly when β gets larger. The corrected approximations are given in the brackets. All simulations are replicated 50 000 times. The corrected approximation gives very satisfactory results and can be used for designing the boundary b for given false detecting probability α. Also, we note that when β increases, the false detection probability decreases significantly.

In Table 2, simulated power of detection for $\delta = 0.1,0.2,0.5$ is given for $\beta L = 5$. The approximated power is calculated by using Equation (3.1) with b being replaced with the corrected boundary $= b + \beta\rho_+/(\beta/(2-\beta))^{1/2}$. We see that the approximation with correction works very well except for large δ and small b.

Remark 4 As a referee commented, for one-sided detection, we can force $Z_k = -b\sqrt{\beta/(2-\beta)}$ whenever $Z_k < -b\sqrt{\beta/(2-\beta)}$ to avoid too negative values when change occurs. On the other hand, the EWMA chart can be naturally used to detect two-sided signals. An alarm will be made at

$$\tau_{EW2} = \inf\left\{ n > 0 : |Z_n| > b\left(\frac{\beta}{2-\beta}\right)^{1/2} \right\}.$$

The false detection probability can be approximated by

$$P_0^*\left[\max_{1\le k\le L}|Z_k| \ge b(\beta/(2-\beta))^{1/2}\right] \approx 1 - e^{-2L\beta b^2(1-\Phi(b))} \approx 2L\beta b^2(1-\Phi(b)). \quad (3.6)$$

See Dirkse (1975) for a second order approximation in the continuous time case.

Table 7 in Section 4 also gives the comparison of simulated and approximated values for the false alarm probability for two-sided EWMA chart corresponding to the case $N = 1$.

TABLE 2 Simulated power of detection for $\beta L = 5$

		$L = 500\ (\beta = .01)$	100 (.05)	20 (.25)
b	δ	Simul (cor)	Simul (cor)	Simul (cor)
2.5	0.1	0.7563 (0.7297)	0.34 (0.3542)	0.1353 (0.1352)
	0.2	0.9942 (0.2265)	0.64312 (0.6503)	0.2172 (0.2143)
	0.5	1 (1)	0.9978 (0.3643)	0.5796 (0.5720)
3.0	0.1	0.4883 (0.5032)	0.1371 (0.1381)	0.04056 (0.03845)
	0.2	0.9648 (0.3905)	0.3675 (0.3728)	0.0735 (0.0700)
	0.5	1 (1)	0.98208 (0.2746)	0.30804 (0.2981)
3.5	0.1	0.2397 (0.2433)	0.0424 (0.0400)	0.0087 (0.0082)
	0.2	0.8618 (0.6357)	0.1564 (0.1537)	0.0170 (0.0171)
	0.5	1 (1)	0.9123 (0.5550)	0.1211 (0.1131)
4.0	0.1	0.0868 (0.0865)	0.0089 (0.0087)	0.00146 (0.0013)
	0.2	0.6496 (0.5964)	0.0484 (0.0470)	0.0035 (0.0032)
	0.5	1 (1)	0.7374 (0.6293)	0.0352 (0.0320)
4.5	0.1	0.0233 (0.0230)	0.000152 (0.00015)	0.00012 (0.00017)
	0.2	0.3876 (0.3714)	0.0114 (0.0109)	0.00050 (0.00045)
	0.5	1 (1)	0.4806 (0.4544)	0.0082 (0.0069)

3.2 | MA procedure

For the MA chart, we assume that l is the window size and h is the control limit. The power of detection can be written as

$$P_0^*\left[\max_{1\le k\le L}\frac{X_{k-l+1} + \cdots + X_k + \min(k,l)\delta}{l} \ge h\right], \quad (3.7)$$

where all X_i's for $i = -l, \ldots, 0, \ldots$, are iid $N(0,1)'$s. In Noonan and Zhigljavsky (2020), approximations for the power of detection is studied given ARL_0 and compared with the CUSUM procedure. Here since we fix the false detection probability, we give a simpler approximation using the similar technique as in Theorem 1 by treating its limit as a stationary Gaussian process without proof.

Theorem 2 *Assume that* $f := \sqrt{l}(h-\delta) \to \infty$ *such that* $f\phi(f)L/l \to 0$ *and* $f \ge \sqrt{2\log(L/l)}$. *Then the probability in Equation (3.7) can be approximated by*

$$\int_0^{L/l} (\sqrt{l}h - \sqrt{l}\min(u,1)\delta)^2 \times [1 - \Phi(\sqrt{l}h - \sqrt{l}\min(u,1)\delta)]du$$

TABLE 3 Simulated false detection probability and power of detection for $L = 20$ for MA chart

	l	5	10	15	20
h	δ	Simul (cor)	Simul (cor)	Simul (cor)	Simul (cor)
1.0	0	0.1407 (0.1563)	0.00846 (0.00916)	0.00052 (0.00062)	0.00006 (0.00004)
	0.1	0.2133 (0.2273)	0.0168 (0.0187)	0.0016 (0.0016)	0.000014 (0.00013)
	0.2	0.3069 (0.3150)	0.0377 (0.0368)	0.0051 (0.0044)	0.00038 (0.000043)
	0.3	0.4158 (0.4126)	0.0762 (0.06731)	0.0124 (0.0109)	0.0020 (0.0014)
0.9	0	0.2216 (0.2379)	0.02012 (0.02243)	0.0023 (0.0024)	0.00028 (0.00028)
	0.1	0.3157 (0.3265)	0.0424 (0.0416)	0.0058 (0.0056)	0.00072 (0.00072)
	0.2	0.4324 (0.4248)	0.0784 (0.0735)	0.0143 (0.0130)	0.0023 (0.0020)
	0.3	0.5520 (0.5194)	0.1425 (0.1195)	0.0333 (0.0277)	0.0077 (0.0052)
0.8	0	0.3297 (0.3396)	0.0498 (0.0490)	0.0081 (0.0081)	0.0014 (0.0014)
	0.1	0.4385 (0.4383)	0.0892 (0.0825)	0.0177 (0.0167)	0.0035 (0.0032)
	0.2	0.5665 (0.5331)	0.1533 (0.1302)	0.0392 (0.0330)	0.0095 (0.0074)
	0.3	0.6811 (0.6045)	0.2467 (0.1872)	0.0818 (0.05945)	0.0240 (0.0160)
0.7	0	0.4556 (0.4526)	0.1020 (0.0948)	0.0241 (0.0230)	0.0059 (0.0059)
	0.1	0.5737 (0.5471)	0.1691 (0.1444)	0.0506 (0.0414)	0.0127 (0.0116)
	0.2	0.6958 (0.6178)	0.2643 (0.2028)	0.0942 (0.0704)	0.0315 (0.0223)
	0.3	0.7928 (0.6437)	0.3848 (0.2558)	0.1705 (0.1069)	0.0682 (0.0397)

$$= \int_0^1 (\sqrt{l}h - \sqrt{l}u\delta)^2[1 - \Phi(\sqrt{l}h - \sqrt{l}u\delta)]du + (L-l)(h-\delta)^2(1 - \Phi(\sqrt{l}h - \sqrt{l}\delta)), \text{ if } L \geq l. \quad (3.8)$$

When $\delta = 0$, the false alarm probability is approximately

$$P_0^*\left[\max_{1\leq k\leq L} \frac{X_{k-l+1} + \cdots + X_k}{l} \geq h\right] \approx Lh^2(1 - \Phi(h\sqrt{l})). \quad (3.9)$$

Remark 5 For practical use, we can use the continuous correction by adding the mean overshoot to h. When l is large, $(X_{k-l+1} + \cdots + X_k)/l$ behaves like a random walk with increment X_k/l. So we can change h to $h^* = h + \rho_+/l$ with $\rho_+ = 0.5826$. As for the EWMA process, the MA chart can also be naturally used for two-sided test.

To show how well the approximation behaves, we take $L = 20$, $l = 5,10,15,20$, and $h = 0.7,0.8,0.9,1.0$. Table 3 gives the simulated false detection probability and the power of detection for $\delta = 0.1,0.2$ and 0.3. The approximated values from Equations (3.8) and (3.9) with corrected boundary are given in the bracket. Again, we see that the approximation with corrected boundary works quite well except for large δ.

3.3 | CUSUM chart

For the CUSUM procedure, for given signal strength δ, the control limit d is typically chosen based on the approximation for ARL_0. The power of the detection can be written as

$$P_\delta^*(\tau_{CS} \leq L | Y_0 =_d Y_\infty) = P_\delta^*\left(\max_{0\leq k\leq L} Y_k > d | Y_0 =_d Y_\infty\right),$$

where Y_∞ follows the stationary distribution.

Note that $P_0^*[\tau_{CS} \leq L | Y_0 =_d Y_\infty]$ is essentially the same as the significance level for the likelihood ratio test for H_a: $X_1, \ldots, X_k$ and $X_{n+1}, \ldots, X_L$ follow $N(0,1)$ and $X_{k+1}, \ldots, X_n$ follow $N(0,\delta)$, where $Y_0 = 0$. The power of detection corresponds to the power for testing H_a under the special alternative $k = 0$ and $n = L$.

An approximation for $P_0(\tau_{CS} \leq L | Y_0 = 0) = P_0(\max_{0\leq k\leq n\leq L} (S_n - S_k - (n-k)\delta/2) > d)$ is given in Siegmund (1985, p. 240) by using the boundary crossing probability approximation for a normal random walk $\{S_n\}$ with drift zero. By assuming $\delta L/2 > d$,

$$P_0(\tau_{CS} \leq L | Y_0 = 0) \approx \delta(\delta L/2 - d)e^{-\delta(d+2\rho_+)},$$

where $\rho_+ = 0.5826$.

3.4 | Comparison of power of detection

In this subsection, we compare the EWMA, MA and CUSUM charts in terms of power of detection for fixed false detection probability α at the stationary state.

For $L = 20$, $\alpha = .01$, and $\beta = .01, .025, .05, .10$, the approximated value for the boundary b are calculated as 2.2874, 2.6713, 2.8914 and 3.0636 from Equation (3.6), respectively, by using corrected approximation given in Theorem 1. Table 4 gives the power of detection. We see that $\beta = .05$ and 0.10 should be recommended as it has the largest power no matter what the strength of signal is.

For the MA chart, for $L = 20$ and $\alpha = .01$, and $l = 5,10,15,20$, we use the corrected approximation by doubling the value given in Equation (3.9) and find the corresponding corrected h as 1.2578, 0.9136, 0.7548, and 0.6582. Table 5 gives the simulated power of detection as in Table 5. Comparing the two tables, we see that two charts differ very little by considering the effect of approximation.

TABLE 4 Simulated power of detection with $L = 20$ and $\alpha = .01$ for EWMA chart

δ/β	0.01	0.025	0.05	0.1
0.25	0.09923	0.0843	0.0851	0.0791
0.5	0.2922	0.3200	0.3442	0.3315
0.75	0.5926	0.6906	0.7367	0.7124
1.00	0.8474	0.9274	0.9523	0.9461
1.25	0.9673	0.9923	0.9966	0.9961
1.5	0.9958	0.9998	0.9998	0.99998
1.75	0.9997	1	1	1
2.0	1	1	1	1

TABLE 5 Simulated power of detection with $L = 20$ and $\alpha = .01$ for MA chart

$\delta/1$	5	10	15	20
0.25	0.1221	0.09700	0.0863	0.0682
0.5	0.3369	0.3405	0.3335	0.3180
0.75	0.6470	0.6925	0.7158	0.7186
1.00	0.8897	0.9293	0.9432	0.9522
1.25	0.9829	0.9938	0.9956	0.9971
1.5	0.9989	0.9998	0.9998	0.9999
1.75	0.9999	1	1	1
2.0	1	1	1	1

TABLE 6 Simulated power of detection for $L = 20$ and $\alpha = .01$ for CUSUM chart

δ/δ_0	0.5	1.0
0.25	0.0813	0.06082
0.5	0.3577	0.2694
0.75	0.7546	0.6425
1.00	0.9600	0.9181
1.25	0.9978	0.9926
1.5	0.99996	0.99966
1.75	1	1
2.0	1	1

For the CUSUM chart, we take two popular reference value $\delta_0 = 0.5$ and 1.0. For $\alpha = .01$ and $L = 20$, the simulated values for boundary are $d = 8.5$ and 5.9 respectively (Table 6).

We see that although the CUSUM chart gives slightly larger power of detection for larger values of δ, it is too sensitive to the choice of the reference value of δ.

4 | MULTI-DIMENSIONAL EWMA CHART

In this section, we consider a natural multivariate EWMA procedure that can detect a change without knowing the direction of signal. Let $X_i = (X_i(1), \dots, X_i(N))^T$ for $i \geq 1$ be N independent normal vector sequences that follows $N(0, \Sigma)$ for $i \leq \nu$ and $N(\delta, \Sigma)$ for $i = \nu + 1, \dots, \nu + L$, where $\delta = (\delta_1, \dots, \delta_N)^T$ is the reference signal with the strength of the signal being defined as $\|\delta\|_\Sigma = \sqrt{\delta^T \Sigma^{-1} \delta}$ and $\|.\|_\Sigma$ denotes the Mahalanobis-norm.

Define

$$Z_j = (1 - \beta) Z_{j-1} + \beta X_j.$$

For the control limit b^2, an alarm will be raised at

$$\tau_{MEW} = \inf \left\{ j > 0 : Z_j^T \Sigma^{-1} Z_j > b^2 (\beta/(2 - \beta)) \right\}.$$

See Lowry et al. (1992) for the numerical evaluation of ARL_0 and ARL_1.

Note that $Z_j^T \Sigma^{-1} Z_j$ is simply a chi-square process as a sum of squared EMWA processes. Similar to Theorem 1, we can show that as $\beta \to 0$ and $L \to \infty$ such that $\beta L \to \infty$,

$$Z_j^T \Sigma^{-1} Z_j / \sqrt{\beta/(2 - \beta)} \Rightarrow \left\{ e^{-u} \left\| \mathbf{W}\left(e^{2u}\right) \right\|, 0 \leq u \leq L\beta \right\},$$

where $\mathbf{W}(t)$ is a N-dimensional Brownian motion.

The following theorem gives the approximation for the false detection probability $P\left[\max_{0 \leq u \leq L\beta} e^{-u} \left\| \mathbf{W}\left(e^{2u}\right) \right\| > b\right]$.

Theorem 3 *As $\beta L \to \infty$ and $b \to \infty$, the false probability of detection $P_0^*(\tau_{MEW} \leq t)$ is equal to*

$$\begin{aligned} & P\left(\max_{0 \leq t \leq L\beta} Y_t \geq \frac{b^2}{2} \right) \\ & \approx 1 - exp\left(- \frac{2L\beta}{\Gamma(N/2) \int_0^{b^2/2} u^{-N/2} e^u du} \right) \\ & \approx 1 - exp\left(- \frac{2L\beta \left(b^2/2\right)^{N/2}}{\Gamma(N/2)} e^{-b^2/2} \left(1 - N/b^2\right) \right). \quad (4.1) \end{aligned}$$

Remark 6 A second order approximation is also given in De Long (1981) where it is shown that

$$\begin{aligned} P\left[\max_{0 \leq u \leq L\beta} e^{-u} \left\| \mathbf{W}\left(e^{2u}\right) \right\| > b \right] \approx & \frac{b^N}{2^{N/2} \Gamma(N/2)} e^{-b^2/2} \\ & \times \left[2L\beta \left(1 - \frac{N}{b^2}\right) + \frac{4}{b^2} + O\left(\frac{1}{b^4}\right) \right]. \end{aligned}$$

We note that the difference is only a second order term in $1/b^2$. Since we use the continuous model as an approximation for the discrete time case, this term can be ignored.

Remark 7 To apply the above approximation, we need a continuous correction for the overshoot. We first note that

$$\begin{aligned} Z_k^T \Sigma^{-1} Z_k &= ((1 - \beta) Z_{k-1} + \beta X_k)^T \Sigma^{-1} ((1 - \beta) Z_{k-1} + \beta X_k) \\ &= (1 - \beta)^2 Z_{k-1}^T \Sigma^{-1} Z_{k-1} + 2\beta(1 - \beta) Z_{k-1}^T \Sigma^{-1} X_k > \\ &\quad + \beta^2 X_k^T \Sigma^{-1} X_k \\ &= (1 - \beta)^2 Z_{k-1}^T \Sigma^{-1} Z_{k-1} + 2\beta(1 - \beta) \left(Z_{k-1}^T \Sigma^{-1} Z_{k-1}\right)^{1/2} \\ &\quad \times \frac{Z_{k-1}^T \Sigma^{-1} X_k}{\left(Z_{k-1}^T \Sigma^{-1} Z_{k-1}\right)^{1/2}} \end{aligned}$$

TABLE 7 Simulated and approximate false detection probability for $N = 1,2,10,100$

		$L = 500$ ($\beta = .01$)	100 (.05)	20 (.25)	$L\beta = 5$
N	b	Simul (cor)	Simul (cor)	Simul (cor)	No cor
1	2.5	0.3059 (0.3272)	0.2519 (0.2632)	0.1564 (0.1620)	0.3882
	3.0	0.0957 (0.0976)	0.0736 (0.0740)	0.0404 (0.0408)	0.1215
	3.5	0.0220 (0.0218)	0.0158 (0.0156)	0.0077 (0.0077)	0.0285
	4.0	0.0037 (0.0038)	0.0026 (0.0026)	0.00114 (0.00122)	0.0051
	4.5	0.00049 (0.00048)	0.00037 (0.00031)	0.00011 (0.00012)	0.00069
2	3	0.3006 (0.2771)	0.2434 (0.2268)	0.1400 (0.1424)	0.4110
	3.5	0.0863 (0.0848)	0.0652 (0.0637)	0.0328 (0.0339)	0.1165
	4	0.0177 (0.0175)	0.0121 (0.0123)	0.0056 (0.0057)	0.0242
	4.5	0.0026 (0.0026)	0.0015 (0.0017)	0.00074 (0.00072)	0.00374
	5	0.00020 (0.00029)	0.00018 (0.00018)	0.00008 (0.00007)	0.00044
10	4.5	0.5655 (0.5359)	0.4818 (0.4636)	0.3056 (0.3194)	0.9342
	5	0.2010 (0.2022)	0.1531 (0.1550)	0.0778 (0.0844)	0.2919
	5.5	0.0441 (0.0435)	0.0299 (0.0303)	0.0138 (0.0138)	0.0602
	6	0.0056 (0.0061)	0.0042 (0.0039)	0.0017 (0.0016)	0.0087
	6.5	0.00068 (0.000060)	0.00048 (0.00036)	0.00040 (0.00012)	0.00090
100	11.5	0.46904 (0.4832)	0.3800 (0.4007)	0.2210 (0.2488)	0.8108
	12	0.1337 (0.1399)	0.0943 (0.1005)	0.0425 (0.0475)	0.2006
	12.5	0.0204 (0.0211)	0.0135 (0.0136)	0.0048 (0.0052)	0.0306
	13	0.0019 (0.0019)	0.0011 (0.0011)	0.00046 (0.00037)	0.00298
	13.5	0.00004 (0.00012)	0.00012 (0.000063)	0.00004 (0.00002)	0.000190

$$+\ \beta^2\frac{\left(Z_{k-1}^T\Sigma^{-1}X_k\right)^2}{Z_{k-1}^T\Sigma^{-1}Z_{k-1}}+\beta^2\left(X_k^T\Sigma^{-1}X_k-\frac{\left(Z_{k-1}^T\Sigma^{-1}X_k\right)^2}{Z_{k-1}^T\Sigma^{-1}Z_{k-1}}\right)$$
$$=\left((1-\beta)\left(Z_{k-1}^T\Sigma^{-1}Z_{k-1}\right)^{1/2}+\beta\frac{Z_{k-1}^T\Sigma^{-1}X_k}{\left(Z_{k-1}^T\Sigma^{-1}Z_{k-1}\right)^{1/2}}\right)^2$$
$$+\ \beta^2\left(X_k^T\Sigma^{-1}X_k-\frac{\left(Z_{k-1}^T\Sigma^{-1}X_k\right)^2}{Z_{k-1}^T\Sigma^{-1}Z_{k-1}}\right).$$

Note that $\varepsilon_k = Z_{k-1}^T\Sigma^{-1}X_k/\left(Z_{k-1}^T\Sigma^{-1}Z_{k-1}\right)^{1/2}$ is $N(0,1)$, and $Y_k = X_k^T\Sigma^{-1}X_k - \left(Z_{k-1}^T\Sigma^{-1}X_k\right)^2/Z_{k-1}^T\Sigma^{-1}Z_{k-1}$ is a χ^2_{N-1} random variable, and the two are mutually independent and both are also independent of Z_{k-1}. That means, we can write $Z_k^T\Sigma^{-1}Z_k$ as

$$Z_k^T\Sigma^{-1}Z_k=\left((1-\beta)\left(Z_{k-1}^T\Sigma^{-1}Z_{k-1}\right)^{1/2}+\beta\varepsilon_k\right)^2+\beta^2Y_k=\beta^2$$
$$\times\left[\left(\frac{1-\beta}{\beta}\left(Z_{k-1}^T\Sigma^{-1}Z_{k-1}\right)^{1/2}+\varepsilon_k\right)^2+Y_k\right].$$

Thant means, $Z_k^T\Sigma^{-1}Z_k/\beta^2$ is not only a Markov chain, but is distributed as a non-central chi-square with non-central parameter $(1-\beta)Z_{k-1}^T\Sigma^{-1}Z_{k-1})^{1/2}/\beta$ given $Z_{k-1}^T\Sigma^{-1}Z_{k-1}$.

As $\beta \to 0$ and $\left(Z_{k-1}^T\Sigma^{-1}Z_{k-1}\right)^{1/2} \approx b\sqrt{\beta/(2-\beta)} \to \infty$ such that $\beta^{3/2}b \to 0$,

$$\left(Z_k^T\Sigma^{-1}Z_k\right)^{1/2}=(1-\beta)\left(Z_{k-1}^T\Sigma^{-1}Z_{k-1}\right)^{1/2}+\beta\varepsilon_k+O_p\left(\beta^2\right).$$

Therefore, just like in the one-dimensional case, $\left(Z_k^T\Sigma^{-1}Z_k\right)^{1/2}$ behaves locally as a normal random walk as $\beta\left(Z_k^T\Sigma^{-1}Z_k\right)^{1/2}$ is at the lower order. Thus, we can correct b to $b^* = b + \rho_+\beta/\sqrt{\beta/(2-\beta)}$.

Table 7 gives the simulated value for the false alarm probability along with the corrected values for $N = 1,2,10,100$ and $\beta L = 5$ with $\delta = 0$. For the corrected approximation we used the approximation given in Equation (4.1). We see that the approximation without correction significantly overestimates the true value. The corrected approximation works very well, even for large N and β.

Remark 8 To study the power of detection, we assume that the signal is $\delta = (\delta_1, \ldots, \delta_N)$ with strength $\|\delta\|_\Sigma$. With a simple transformation, we can assume that the signal only appear in the first panel with strength $\|\delta\|_\Sigma$. By denoting $\widetilde{\mathbf{B}}(t)$ as the $(N-1)$-dimensional standard Brownian motion that is independent of a Brownian motion $W(t)$, as in the proof for Theorem 3, we can write the power of detection as

$$P\left(\max_{0\le u\le L\beta}e^{-2u}\left(\left(W\left(e^{-2u}\right)+\|\delta\|_\Sigma\left(1-e^{-u}\right)/\sqrt{\beta/(2-\beta)}\right)^2\right.\right.$$
$$\left.\left.+\left\|\widetilde{\mathbf{B}}\left(e^{-2u}\right)\right\|^2\right)>b^2\right)$$

By using the same technique as in the proof for Theorem 1 and noting $e^{-2u}\left\|\widetilde{\mathbf{B}}\left(e^{-2u}\right)\right\|^2$ is a stationary chi-square process,

TABLE 8 Comparison of false detection probability and power of detection under intra-class model

$\delta_i = \delta$	$\rho = 0$	0.1	0.2	0.3	0.4	0.5
0.0	0.0302 (0.0294)	0.0303 (0.2738)	0.0306 (0.4228)	0.0292 (0.5235)	0.0308 (0.5936)	0.0309 (0.6458)
0.25	0.1091 (0.1089)	0.1491 (0.3643)	0.1940 (0.4997)	0.2466 (0.5927)	0.3241 (0.6535)	0.4343 (0.6983)
0.5	0.5947 (0.5948)	0.7860 (0.6954)	0.8937 (0.7546)	0.9603 (0.7955)	0.9902 (0.8289)	0.9989 (0.8472)
0.75	0.9802 (0.9802)	0.9987 (0.9834)	0.99996 (0.9860)	1.00 (0.9869)	1.00 (0.9884)8	1.00 (0.9886)
1.00	1.00 (1.00)	1.00 (1.00)	1.0 (1.00)	1.0 (1.00)	1.0 (1.00)	1.0 (1.00)

TABLE 9 False detection probability and power of detection for $\beta = .05$, $L = 20$, $N = 100$ and $K = 10$

δ						
Original	$b = 11.5$	11.6	11.7	11.8	11.9	12
0.0	0.1088	0.0804	0.0599	0.0441	0.0309	0.0217
0.25	0.2214	0.1777	0.1380	0.1086	0.0829	0.0624
0.50	0.7713	0.7235	0.6752	0.6221	0.5721	0.5217
0.75	0.9983	0.9974	0.9961	0.9947	0.9926	0.9890
1.00	1	1	1	1	1	1
Oracle ($N = K = 10$)	$b = 4.6$	4.7	4.8	4.9	5.0	5.1
0.0	0.1159	0.0889	0.0688	0.0511	0.0375	0.0269
0.25	0.5623	0.5054	0.4565	0.4031	0.3574	0.3112
0.50	0.9976	0.9964	0.9955	0.9936	0.9911	0.9873
0.75	1	1	1	1	1	1
1.00	1	1	1	1	1	1
max-K (=10)	b=7	7.1	7.2	7.3	7.4	7.5
0.0	0.1083	0.0798	0.0574	0.0400	0.0275	0.0190
0.25	0.3123	0.2541	0.2048	0.1593	0.1237	0.0956
0.5	0.9226	0.9004	0.8686	0.8350	0.7978	0.7520
0.75	0.99998	0.99998	99 988	0.99974	0.99968	0.99974
1	1	1	1	1	1	1
Min-δ (=0.25)	$b = 7$	7.1	7.2	7.3	7.4	7.5
0.0	0.127	0.1066	0.0855	0.0693	0.0556	0.0429
0.25	0.3770	0.3320	0.2917	0.2570	0.218	0.1891
0.5	0.9387	0.9250	0.9076	0.8919	0.8683	0.8467
0.75	0.99994	0.99994	0.99994	0.99992	0.99994	0.99976
1.0	1	1	1	1	1	1

we can approximate the above probability as

$$\int_0^{L\beta} \mathbf{E}\left[\left(\left(b^2 - \widetilde{Y}^2(\infty)\right)^{1/2} - \frac{\|\delta\|_\Sigma(1 - exp(-u))}{\sqrt{\beta/(2-\beta)}}\right)^2 \times\left(1-\Phi\left(\left(b^2-\widetilde{Y}^2(\infty)\right)^{1/2} - \frac{\|\delta\|_\Sigma(1-exp(-u))}{\sqrt{\beta/(2-\beta)}}\right)\right)\right] du$$

where $\widetilde{Y}^2(\infty)\big)$ is a $(N-1)$-dimensional chi-square random variable as in the proof for Theorem 3.

Example 1 Under the intra-class correlation model, $X_{ij} = a_j + \varepsilon_{ij}$ for $i = 1, \ldots, N$ and $j = 1, 2, \ldots$ before the change with a_j follows $N\left(0, \sigma_a^2\right)$ and e_{ij} follows $N\left(0, \sigma_e^2\right)$ and a_j and e_{ij} are independent crossing i and j. We can find

$$\Sigma = \sigma^2((1-\rho)I + \rho J),$$

where $I_{N\times N}$ is the identity matrix, $J_{N\times N}$ is the matrix with elements being 1, and $\sigma^2 = \sigma_a^2 + \sigma_e^2$ and $\rho = \sigma_a^2/\sigma^2$. From Sylvester' theorem, we know

$$\Sigma^{-1} = \frac{1}{\sigma^2(1-\rho)}\left(I - \frac{\rho}{1+\rho(N-1)}J\right).$$

Thus,

$$\tau_{MEW} = \inf\left\{t > 0 : Y_t^T \Sigma^{-1} Y_t > b^2\left(\frac{\beta}{2-\beta}\right)\right\}$$

$$= \inf\left\{t > 0 : \frac{1}{\sigma^2(1-\rho)}\left(\sum_{i=1}^N Y_{it}^2 - \frac{\rho}{1+\rho(N-1)}\left(\sum_{i=1}^N Y_{it}\right)^2\right) > b^2\left(\frac{\beta}{2-\beta}\right)\right\}. \quad (4.2)$$

TABLE 10 Simulated power of detection of Min-δ chart for $\beta = .05$ and $L = 20$ with $N = 100$ and $\delta_0 = 0.25$

δ	b = 7	7.1	7.2	7.3	7.4	7.5
$\delta = 0$	0.127	0.1066	0.08546	0.06932	0.055556	0.04294
K = 1						
0.25	0.1459	0.1237	0.1029	0.0806	0.0637	0.0509
0.5	0.1940	0.1624	0.1349	0.1116	0.0925	0.0742
0.75	0.2966	0.2560	0.2176	0.1884	0.1565	0.1317
1.0	0.4679	0.4207	0.3757	0.3317	0.2938	0.2542
K = 5						
0.25	0.2294	0.1978	0.1710	0.1402	0.1139	0.0934
0.5	0.6069	0.5589	0.5164	0.4791	0.4304	0.3889
0.75	0.9639	0.9527	0.9427	0.9281	0.9160	0.8958
1.0	0.9992	0.9998	0.9997	0.9994	0.9991	0.9988
K = 10						
0.25	0.3770	0.3320	0.2917	0.2570	0.2180	0.1891
0.5	0.9387	0.9250	0.9076	0.8919	0.8683	0.8467
0.75	0.99994	0.99994	0.99994	0.99992	0.99994	0.99976
1.0	1	1	1	1	1	1

TABLE 11 Simulated power of detection of two-sided Min-δ chart for $\beta = .05$ and $L = 20$ with $N = 100$ and $\delta_0 = 0.25$

δ	b = 7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
$\delta = 0$	0.1314	0.10772	0.08304	0.06616	0.0524	0.03946	0.02904	0.02244
K = 1								
0.25	0.1422	0.1135	0.0908	0.0699	0.0555	0.0409	0.03216	0.0236
0.5	0.1740	0.1377	0.1142	0.0899	0.0703	0.0539	0.0414	0.0324
0.75	0.2409	0.2033	0.1683	0.1342	0.1110	0.0880	0.0701	0.0524
1.0	0.3816	0.3352	0.2837	0.2466	0.2066	0.1730	0.1450	0.01233
K = 5								
0.25	0.1892	0.1559	0.1270	0.1019	0.0805	0.0627	0.0478	0.0371
0.5	0.5034	0.4552	0.4076	0.3637	0.3222	0.2839	0.2435	0.2077
0.75	0.9317	0.9177	0.9013	0.8792	0.8571	0.8317	0.8008	0.7712
1.0	0.9995	0.9993	0.9990	0.9988	0.9979	0.9973	0.9959	0.9948
K = 10								
0.25	0.2873	0.2471	0.2109	0.1785	0.1477	0.1208	0.1023	0.0803
0.5	0.8951	0.8734	0.8508	0.8231	0.7953	0.7649	0.7253	0.6923
0.75	0.9999	0.9999	0.9998	0.9996	0.9996	0.9994	0.9992	0.9991
1.0	1	1	1	1	1	1	1	1

We can use the intra-class model to check how the cross-correlation affects the power of the regular control chart τ_0 given

$$\tau_0 = \inf\left\{t > 0 : Y_t^T Y_t > b^2(\beta/(2-\beta))\right\} \qquad (4.3)$$

by ignoring the cross-correlation.

For $N = 10$, Table 8 gives the simulated power of detection when $ARL_0 = 1000$, $\nu = 100$, $\beta = .05$, and $\sigma_a^2 = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5$ and $\sigma_e^2 = 1,0.9,0.8,0.7,0.6,0.5$ correspondingly to make $\sigma^2 = 1$. The control limit b is selected as 5.5 that corresponds to false detection probability 0.03. For signal $\delta_1 = \cdots = \delta_N$, at the stationary state we assume Y_0 follows $N\left(0, \frac{\beta}{2-\beta}\Sigma\right)$. The numbers in the brackets are the corresponding values for τ_0.

From Table 8, we can see that the intra-calls correlation dramatically increases the false detection probability under the regular control chart. In addition, as δ gets larger, it has even lower power of detection.

5 | MODIFIED MEWMA CHART FOR DETECTING SPARSE TRANSIENT SIGNALS

When the change or signal only appears in a small portion of the N panels, called sparse signal, the power of detection will be low if we use the original MEWMA chart without considering the sparsity of the signal. To show how much power can be improved, we first conduct a simulation study in the oracle case, that means, we know exactly which portion of the panels changes. Without loss of generality, we can assume that X_j follows $N(0, I)$ without change.

We choose $\beta = 0.05$ and $L = 20$. Suppose $N = 100$ and the common change occurs in the first $K = 10$ panels with signal strength $\delta = 0$ (no change), 0.25, 0.5, 0.75, 1.0. For the regular MEWMA chart, we choose $b = 11.5, 11.6$ 11.7, 11.8, 11.9, 12. For the oracle case, we just use the first 10 panels to form the MEWMA chart with $N = 10$ and $b = 4.6, 4.7, 4.8, 4.9, 5.0, 5.1$. Table 9 gives the simulated significance level and power of detection.

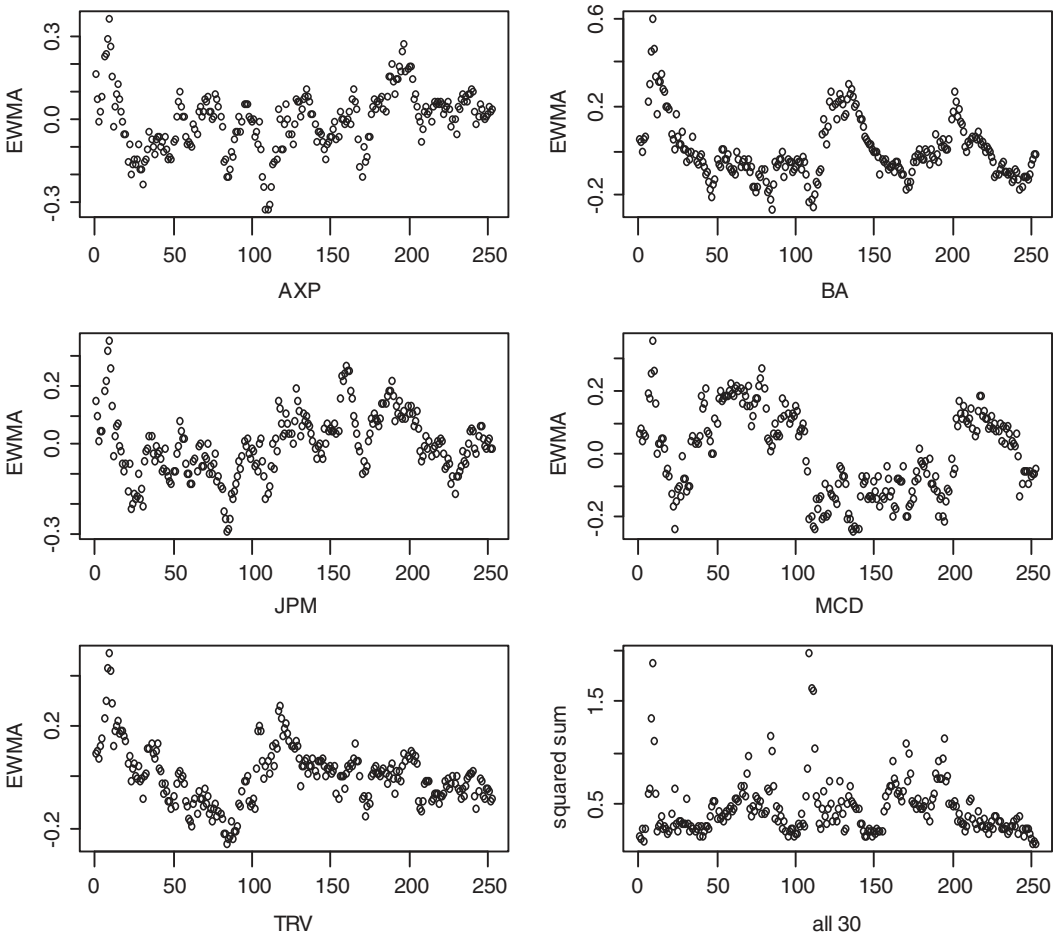

FIGURE 1 EWMA and sum of squared EWMA processes for 30 Dow Jones industrial stocks

From Table 9, we see that the power of detection increases drastically if we know exactly which panels have changes. However, as we do not know which panels the common change occurs, we need to trim the panels without change. We first assume that we know the post-change mean is positive among those changed panels.

From isolating changed panel point of view, we can use the K maximum EWMA processes at each time, that is. we make an alarm at

$$\tau_K = \inf\left\{j > 0 : Z_j^2[1] + \cdots + Z_j^2[K] > b_1^2\big(\beta/(2-\beta)\big)\right\},$$

where $Z_j[1], \ldots, Z_j[K]$ are the largest K EWMA processes at time j. Here, the reference value for K needs to be given.

From detecting signal point of view, we can use those $Z_j(i)$'s such that $Z_j(i) > \delta_0$, say, where δ_0 is the minimum signal strength to be detected, that is, we make an alarm at

$$\tau_{\delta_0} = \inf\left\{j > 0 : \sum_{k=1}^{N} Z_j^2(i) I_{[Z_j(i)>\delta_0]} > b_2^2(\beta/(2-\beta)\right\}$$

In Table 9, we also presented the simulated results for τ_K and τ_2 for $b_1 = b_2 = 7, 7.1, 7.2, 7.3, 7.4, 7.5$ where $K = 10$ and $\delta_0 = 0.25$. From the simulation results, we see that the max-K and min-δ procedures give much improved power of detection. However, from signal detection point of view, τ_{δ_0} gives larger power, although the number of panels used at every step is random. So we recommend to use the Min-δ procedure.

To show how the number of changed panels affects the power of detection, Table 10 gives the simulated powers for $K = 1,5,10$ under the same design as in Table 9 for *Min-δ* chart with $\delta_0 = 0.25$. Theoretically, there is an optimization problem on how to choose the optimal minimum signal strength δ_0 for detecting a signal with strength δ.

When the sign of post-change mean is unknown, we need to use a pair of truncated MEWMA charts by making an alarm at $\min\left(\tau_{\delta_0}, \tau_{-\delta_0}\right)$, where for $-\delta_0$,

$$\tau_{-\delta_0} = \inf\left\{j > 0 : \sum_{k=1}^{N} Z_j^2(i) I_{[Z_j(i)<-\delta_0]} > b_2^2(\beta/(2-\beta)\right\}$$

Table 11 gives the corresponding results as in Table 10.

Comparing Tables 10 and 11, we see that although the power of detection for the two-sided case is lower than that for the one-sided case, it still has significant improved power comparing with the original one.

6 | APPLICATION WITH A REAL DATA EXAMPLE

To detect a transient signal in a single panel, we can run the EWMA process and initiate the monitoring for any segment

of length L or segment by segment. For multiple panels, we are more interested in detecting a common transient signal. We can run all individual EWMA processes simultaneously and use the squared sum as the detecting process in a segment with length L. If an alarm is raised, we can inspect the individual EWMA processes and isolate those panels with larger (absolute) values.

For a demonstration, we use Dow Jones 30 industrial stock closing prices from May 26, 2020 to May 26, 2021 as a demonstration. The data are downloaded from finance.yahoo.com. For total 244 days, we use the simple exponential normal random walk model for daily closing price P_t. That means, we use the differences of the logarithms $\ln(P_t/P_{t-1})$ as the raw data. As we are only interested in detecting transient segment change, so we first modify the data by truncating the outliers with $\pm 3\sigma$-rule and then standardize by subtracting the sample mean and dividing the standard deviation to satisfying our model assumptions. The ACF plots show that there are no significant correlations. The 30 stocks are

AXP, AMGN, AAPL, BA, CAT, CSCO, CVX, GS, HD, HON, IBM, INTC, JNJ, KO, JPM,
MCD, MMM, MRK, MSFT, NKE, PG, TRV, UHN, CRM, VZ, V, WBA, WMT, DIS, DOW.

We first calculated the EWMA processes for each stock. Figure 1 gives the plots for AXP, BA, JPM, MCD, and TRV. The sum of squared EWMA processes is also plotted in the last graph. For $N = 30$, $\alpha = .05$, $\beta = .05$ and $L = 20$, the approximation by using Theorem 3 gives $b^* \approx 7.2$ after continuous correction. This gives $b^*\beta/(2-\beta) \approx 1.33$. The two boundary crossing segments are detected as [8,9] and [109–111].

To identify which stocks have the increment segment signal, we also run each individual EWMA process with $\alpha = 0.10$. By Theorem 2, this gives $b^* \approx 2$ and $b^*\sqrt{\beta/(2-\beta)} \approx 0.32$. The individual charts show that the five stocks plotted here show increment in mean around 9; while no stock shows increment in the second segment. Therefore, we can claim that the a common change segment is [1–9], that is, from May 26, 2020 to June 5, 2020.

7 | CONCLUSION AND DISCUSSIONS

In this communication, we studied the performance of EWMA and MA processes as detection procedures in terms of power of detection given false detecting probability for a transient signal. The EWMA and MA charts have the advantage that they can be conveniently generalized to multi-dimensional case that can also be modified to deal with common transient signal case. In particular, the continuous time analog of EWMA process is a diffusion process and accurate approximations for false detecting probability and power of detection with a continuous correction. The same techniques developed here may help to deal with more complex models with serial dependent observations such as AR(1) model. Future research may also consider the following aspects.

1. One can consider different signal patterns such as variance change or clustered signals. This is typical in network or spatial data monitoring where the signal may appear as clusters. An EWMA chart for variance change based on log-transformation is considered in Crowder and Hamilton (1993). A recent discussion on the selection of variables is given in Capizzi (2015).
2. A more general measure of performance for a transient signal may include both the power of detection and the truncated delay detection time $E^*[\min(\tau, m)]$ for a stopping time τ. A recent discussion for the redesign of control chart is given in Woodall and Faltin (2019).
3. The investigation of multivariate MA chart and comparison with MEWMA chart will be one of the future objectives.

ACKNOWLEDGMENTS

This work was supported in part from an NSF grant. The authors thank the Editor, an Associate Editor and two anonymous referees for their many constructive comments.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Yanhong Wu https://orcid.org/0000-0002-9137-503X

REFERENCES

Bauer, P., & Hackel, P. (1978). The use of MOSUMS for quality control. *Technometrics*, *20*(4), 431–436.

Berman, S. M. (1969). Limiting distribution of the maximum of a diffusion process. *Annals of Mathematical Statistics*, *35*, 319–329.

Bickel, P. S., & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics*, *6*, 1071–1095.

Capizzi, G. (2015). Recent advances in process monitoring: Nonparametric and variable selection methods for Phase I and Phase II. *Quality Engineering*, *27*(1), 44–67.

Crowder, S. V., & Hamilton, M. D. (1993). An EWMA for monitoring a standard deviation. *Journal of Quality Technology*, *24*(1), 12–21.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica*, *53*(2), 363–384.

Davis, R. A. (1982). Maximum and minimum of one-dimensional diffusions. *Stochastic Processes and their Applications*, *13*, 1–9.

De Long, D. M. (1981). Crossing probabilities for a square root boundary by a Bessel process. *Communications in Statistics - Theory and Methods*, *10*(21), 2197–2213.

Dirkse, J. P. (1975). An absorption probability for the Ornstein-Uhlenbeck process. *Journal of Applied Probability*, *12*(3), 595–599.

Frisén, M. (1992). Evaluation of methods for statistical surveillance. *Statistics in Medicine*, *11*, 1489–1502.

Gueppie, B. K., Fillatre, L., & Nikiforov, I. V. (2012). Sequential detection of transient changes. *Sequential Analysis*, *31*(4), 528–547.

Knoth, S. (2021). Steady-state average run length(s) - methodology, formulas and numerics. *Sequential Analysis*, *40*(3), 405–426.

Lai, T. L. (1974). Control charts based on weighted sums. *Annals of Statistics*, *2*, 134–147.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems (with discussions). *Journal of the Royal Statistical Society. Series B*, *57*(4), 613–658.

Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, *34*(1), 46–53.

Margavio, T. M., Conerly, M. D., Woodall, W. H., & Drake, L. G. (1995). Alarm rates for quality control charts. *Statistics & Probability Letters*, *24*(3), 219–224.

Moustakides, G. V. (2014). Multiple optimality properties of the Shewhart test. *Sequential Analysis*, *33*(3), 318–344.

Noonan, J., & Zhigljavsky, A. (2020). Power of MOSUM test for online detection of transient change in mean. *Sequential Analysis*, *39*(2), 269–293.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*, 100–114.

Pickands, J. (1969). Upcrossing probabilities for stationary Gaussian processes. *Transactions of the American Mathematical Society*, *145*, 51–73.

Pollak, M., & Krieger, A. M. (2013). Shewhart revisited. *Sequential Analysis*, *32*, 230–242.

Qualls, C., & Watanabe, H. (1972). Asymptotic properties of Gaussian processes. *Annals of Mathematical Statistics*, *43*, 580–596.

Reynolds, M. R., Jr., & Stoumbos, Z. G. (2004). Control charts and the efficient allocation of sampling resources. *Technometrics*, *46*(2), 200–214.

Roberts, S. W. (1959). Control charts based on geometric moving average. *Technometrics*, *1*, 239–250.

Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics*, *8*, 411–430.

Shewshart, W. (1931). Economic control of quality of manufactured products. Van Nostrand.

Shiryayev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, *13*, 22–46.

Siegmund, D. (1985). Sequential analysis: Tests and confidence intervals. Springer.

Srivastava, M. S., & Wu, Y. (1993). Comparison of CUSUM, EWMA, and Shiryayev-Roberts procedures for detection of a shift in the mean. *Annals of Statistics*, *21*(2), 645–670.

Tartakovsky, A. G., Berenkov, N. R., Kolessa, A. E., & Nikiforov, I. V. (2021). Optimal sequential detection of signals with unknown appearance and disappearance points in time. *IEEE Transactions on Signal Processing*, *69*, 2653–2662. https://doi.org/10.1109/tsp.2021.3071016

Woodall, W. H., & Faltin, F. W. (2019). Rethinking control chart design and evaluation. *Quality Engineering*, *31*(4), 596–605.

Wu, Y. (1996). A less sensitive linear detector for the change point based on kernel smoothing method. *Metrika*, *43*, 43–55.

How to cite this article: Wu, Y., & Wu, W. B. (2022). Sequential detection of common transient signals in high dimensional data stream. *Naval Research Logistics (NRL)*, *69*(4), 640–653. https://doi.org/10.1002/nav.22034

APPENDIX A

A.1 | Proof of Theorem 1

Let $\{\mathbf{W}(t)\}, t \geq 0$, be the standard Brownian motion. Standard weak convergence theory shows that, as $\beta \to 0$ and $L\beta \to \infty$,

$$\left\{Z_k/(\beta/(2-\beta))^{1/2}, 1 \leq k \leq L\right\} \Rightarrow \left\{e^{-u}\mathbf{W}\left(e^{2u}\right), 0 \leq u \leq L\beta\right\}. \tag{A1}$$

Note that $\left\{e^{-u}\mathbf{W}\left(e^{2u}\right), u \geq 0\right\}$ is a stationary Gaussian process and an Ornstein–Uhlenbeck process. Since $\max_{a\geq 0}|(1-\beta)^{a/\beta} - e^{-a}| = O(\beta)$ as $\beta \to 0$, by Equation (A1), we have

$$P_0^*\left[\max_{1\leq k\leq L}\left(Z_k + \left(1-(1-\beta)^k\right)\delta\right) \geq b(\beta/(2-\beta))^{1/2}\right] = P\left(\max_{0\leq u\leq L\beta}\left[e^{-u}\mathbf{W}\left(e^{2u}\right) + \frac{(1-e^{-u})\,\delta}{(\beta/(2-\beta))^{1/2}}\right] > b\right] + o(1)$$

as $\beta \to 0$ and $L\beta \to \infty$.

Next we shall apply lemma 3.1 in Davis (1982) and show that

$$P\left(\max_{0\leq u\leq L\beta}\left[e^{-u}\mathbf{W}\left(e^{2u}\right) + \frac{(1-e^{-u})\,\delta}{(\beta/(2-\beta))^{1/2}}\right] > b\right) = 1 - \exp\left[-\int_0^{L\beta} m(u)^2(1-\Phi(m(u)))du\right] + o(1). \tag{A2}$$

Let $P^x(\cdot)$ be the probability measure of the Ornstein-Uhlenbeck process given $e^{-u}\mathbf{W}\left(e^{2u}\right)\Big|_{u=0} = x$. By lemma 3.1 in Davis (1982), for v_t with $\lim_{t\to\infty} v_t = \infty$ we have for all x that

$$P^x\left[\max_{0\leq u\leq t} e^{-u}\mathbf{W}\left(e^{2u}\right) \leq v_t\right] = \exp\left[-tv_t^2\left(1-\Phi\left(v_t\right)\right)\right] + o(1) \tag{A3}$$

as $t \to \infty$. Let $J \in \mathbb{N}$ be fixed and $I_j = [(j-1)t/J, jt/J]$, $j = 1, \ldots, J$. Assume $\min_{j\leq J} v_{t,j} \to \infty$. By Equation (A3), we have

$$P\left[\max_{u\in I_j} e^{-u}\mathbf{W}\left(e^{2u}\right) \leq v_{t,j} \text{ for } j = 1, \ldots, J\right] = \prod_{j=1}^{J}\exp\left[-J^{-1}tv_{t,j}^2\left(1-\Phi\left(v_{t,j}\right)\right)\right] + o(1) \tag{A4}$$

To see the above, without loss of generality let $J = 2$. Let event $A_j = \left\{\max_{u\in I_j} e^{-u}\mathbf{W}\left(e^{2u}\right) \le v_{t,j}\right\}$. Then by Markovian property

$$P(A_1 \cap A_2) = E\left[\mathbf{1}_{A_1} P\left(A_2 | e^{-u}\mathbf{W}\left(e^{2u}\right), u \in I_1\right)\right] = E\left[\mathbf{1}_{A_1} P\left(A_2 | e^{-t/2}\mathbf{W}\left(e^{t}\right)\right)\right]. \quad \text{(A5)}$$

By Equation (A3) and the Lebesgue dominated convergence theorem,

$$\lim_{t\to\infty} E|P\left(A_2 | e^{-t/2}\mathbf{W}\left(e^{t}\right)\right) - \exp\left[-2^{-1}tv_{t,2}^2\left(1-\Phi\left(v_{t,2}\right)\right)\right]| = 0. \quad \text{(A6)}$$

Then Equation (A4) follows. Now let $t = L\beta$, $\bar{v}_{t,j} = \max_{u\in I_j} m(u)$ and $\underline{v}_{t,j} = \min_{u\in I_j} m(u)$. Then

$$\begin{aligned} &P\left[\max_{u\in I_j} e^{-u}\mathbf{W}\left(e^{2u}\right) \le \underline{v}_{t,j} \text{ for } j = 1, \dots, J\right] \\ &\le P\left(\max_{0\le u\le L\beta}\left[e^{-u}\mathbf{W}\left(e^{2u}\right) + \frac{(1-e^{-u})\,\delta}{(\beta/(2-\beta))^{1/2}}\right] \le b\right] \\ &\le P\left[\max_{u\in I_j} e^{-u}\mathbf{W}\left(e^{2u}\right) \le \bar{v}_{t,j} \text{ for } j = 1, \dots, J\right], \end{aligned} \quad \text{(A7)}$$

by the monotonicity of cumulative distribution functions. By elementary calculations,

$$\lim_{J\to\infty}\left|\exp\left[-J^{-1}\sum_{j=1}^{J} t\bar{v}_{t,j}^2\left(1-\Phi\left(\bar{v}_{t,j}\right)\right)\right] - \exp\left[-J^{-1}\sum_{j=1}^{J} t\underline{v}_{t,j}^2\left(1-\Phi\left(\underline{v}_{t,j}\right)\right)\right]\right| = 0. \quad \text{(A8)}$$

Therefore Equation (A2) follows from

$$\lim_{J\to\infty}\left|\exp\left[-J^{-1}\sum_{j=1}^{J} t\bar{v}_{t,j}^2\left(1-\Phi\left(\bar{v}_{t,j}\right)\right)\right] - \exp\left[-\int_0^{L\beta} m(u)^2(1-\Phi(m(u)))du\right]\right| = 0. \quad \text{(A9)}$$

A.2 | Proof of Theorem 3

Here we give a proof based on Davis (1982). Since an O-U process defined by $dX_t = -X_t dt + dW_t$ has asymptotic variance 1/2, and if the initial state is stationary, then it is in distribution equivalent to $2^{-1/2}e^{-t}\mathbf{B}\left(e^{2t}\right)$ for a standard Brownian motion $\mathbf{B}(t)$. By treating $\|Z_k\|^2/(2\beta/(1-\beta))$ as the sum of N squared O-U processes, we get the well-known CIR (Cox-Ingersoll-Ross, 1985) process. Indeed, if we denote $X_t^{(i)}$ are N independent O-U processes satisfying

$$dX_t^{(i)} = -X_t^{(i)}dt + dW_t^{(i)},$$

where $W_t^{(i)}$ are independent standard Brownian motions. Since

$$d\left(X_t^{(i)}\right)^2 = \left(-2\left(X_t^{(i)}\right)^2 + 1\right)dt + 2X_t^{(i)}dW_t^{(i)}.$$

Thus, $Y_t = \sum_{i=1}^N \left(X_t^{(i)}\right)^2$ satisfies

$$dY_t = (N - 2Y_t)\,dt + 2\sum_{i=1}^{N} X_t^{(i)}dW_t^{(i)}.$$

Note that $\int_0^t \sum_{i=1}^N X_u^{(i)}dW_u^{(i)}$ has quadratic variation $\int_0^t Y_u du$. Thus, by Levy's characterization theorem, the process

$$\widetilde{W}_t = \int_0^t \frac{1}{\sqrt{Y_u}}\sum_{i=1}^{N} X_u^{(i)}dW_u^{(i)}$$

is a Brownian motion, that is,

$$dY_t = (N - 2Y_t)\,dt + 2\sqrt{Y_t}d\widetilde{W}_t.$$

That means, Y_t is a diffusion process with drift $\mu(x) = N - 2x$ and diffusion $\sigma^2(x) = 4x$.

Thus, the probability in (4.10) is equivalent to $P\left(\max_{0\le t\le L\beta} Y_t \ge b^2/2\right)$. Following the standard notation of diffusion process, since

$$s(x) = exp\left(-\int^x \frac{2\mu(u)}{\sigma^2(u)}du\right) = exp\left(-\int^x\left(\frac{N}{2u} - 1\right)du\right) = x^{-N/2}e^x,$$

so the scale function $S(x)$ is given by

$$S(x) = \int_0^x s(u)du = \int_0^x u^{-N/2}e^u du.$$

On the other hand, the speed function $M(x) = \int_0^x m(u)du$ where the speed density $m(x)$ is given by

$$m(x) = \frac{2}{\sigma^2(x)s(x)} = \frac{1}{2}x^{N/2-1}e^{-x},$$

with total mass

$$|m| = M(\infty) = \int_0^\infty m(x)dx = \Gamma(N/2)/2.$$

From theorem 3.2 of Davis (1982), we have the following approximation.

$$P\left(\max_{0\le t\le L\beta} Y_t \ge \frac{b^2}{2}\right) \approx 1 - exp\left(-\frac{L\beta}{S\left(b^2/2\right)|m|}\right) = 1 - exp\left(-\frac{2L\beta}{\Gamma(N/2)\int_0^{b^2/2} u^{-N/2}e^u du}\right).$$

As $x \to \infty$,

$$x^{N/2}e^{-x}\int_0^x u^{-N/2}e^u du = \int_0^x\left(1-\frac{u}{x}\right)^{-N/2}e^{-u}du \approx \int_0^x e^{-u(1-N/2x)}dx = \frac{1}{1-N/2x},$$

that is,

$$\int_0^x u^{-N/2}e^u du \approx \frac{1}{1-N/2x}x^{-N/2}e^x.$$

Thus,

$$P\left(\max_{0\le t\le L\beta} Y_t \ge \frac{b^2}{2}\right) \approx 1 - exp\left(-\frac{2L\beta\left(b^2/2\right)^{N/2}}{\Gamma(N/2)} e^{-b^2/2}\left(1 - N/b^2\right)\right).$$

A.3 | Asymptotic equivalence of uncontrolled and controlled stationary distribution

The result for single O-U process is well known. For the CIR process defined in the proof for Theorem 3, the uncontrolled stationary density $\pi(x)$ follows Gamma (N/2). Since 0 is an attainable boundary, we fist let $\varepsilon > 0$ be the initial state and the Green function can be written as

$$G(\varepsilon, x) = \frac{1}{2} x^{N/2-1} e^{-x} \frac{\int_{max(\varepsilon,x)}^{b^2/2} u^{-N/2} e^u du \int_0^{min(\varepsilon,x)} u^{-N/2} e^u du}{\int_0^{b^2/2} u^{-N/2} e^u du}.$$

So the controlled stationary density function by letting $\varepsilon \to 0$ is obtained as

$$\begin{aligned}\pi^*(x) &= \lim_{\varepsilon\to 0} \frac{G(\varepsilon, x)}{\int_0^{b^2/2} G(\varepsilon, x)dx} \\ &= \frac{x^{N/2-1} e^{-x} \int_x^{b^2/2} u^{-N/2} e^u du}{\int_0^{b^2/2} x^{N/2-1} e^{-x} \int_x^{b^2/2} u^{-N/2} e^u du dx} \\ &= \frac{x^{N/2-1} e^{-x} \int_x^{b^2/2} u^{-N/2} e^u du}{\int_0^{b^2/2} u^{-N/2} e^u \int_0^u x^{N/2-1} e^{-x} dx du}\end{aligned}$$

As $b \to \infty$,

$$\pi^*(x) \approx \frac{x^{N/2-1} e^{-x}}{\int_0^{b^2/2} x^{N/2-1} e^{-x} dx},$$

for $x \le b^2/2$, which is the truncated Gamma $(N/2)$ density function.