



EWMA control charts for monitoring correlated counts with finite range

Maria Anastasopoulou and Athanasios C. Rakitzis

Department of Statistics and Actuarial-Financial Mathematics, University of Aegean, Karlovasi, Greece

ABSTRACT

In this work, we develop and study upper and lower one-sided EWMA control charts for monitoring correlated counts with finite range. Often in practice, data of that kind can be adequately described by a first-order binomial or beta-binomial autoregressive model. Especially, when there is evidence that data demonstrate extra-binomial variation, the latter model is preferable than the former. The proposed charts can be used for detecting upward or downward shifts in process mean level. Practical guidelines concerning the statistical design of the proposed charts are given, while the effect of the extra-binomial variation is investigated as well. Comparisons with existing control charting procedures are also provided. Finally, an illustrative real-data example is also given.

ARTICLE HISTORY

Received 18 March 2020
Accepted 2 September 2020



KEYWORDS

Average run length; BAR(1) model; beta-BAR(1) model; extra binomial variation; rounding operation; *s*-EWMA chart; statistical process control

1. Introduction

The statistical process monitoring (SPM) methods consist of the basic tools of statistical process control (SPC). Among them, the control chart is the most widely used SPM tool. It can be used for the quick and accurate detection of abnormal (usually unwanted) situations. Traditionally, control charts are used in industry to monitor production processes. The basic aim is to detect a possible increase in the percentage of defective items, which is related to process deterioration. In this case, the Shewhart p and np charts (Montgomery [11]) are widely used to monitor the proportion and the number of nonconforming units, respectively, within a sample of finite size. These monitoring schemes are developed under the assumption that the number of nonconforming units follows a binomial distribution $B(n, \pi)$, where n is the sample size and π is the success probability (i.e. the probability for a unit to be nonconforming). Moreover, a common assumption when the p and np charts are applied in practice is that the successive counts are independent and identically distributed (iid) binomial random variables (rv).

It is well known that Shewhart control charts are not sensitive in the detection of small and medium shifts in the mean of the process (e.g. average number of defective items). The cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) control charts, as control charts with memory, detect such changes more quickly than Shewhart

CONTACT Athanasios C. Rakitzis  anastasopoulou@aegean.gr; arakitz@aegean.gr  Department of Statistics and Actuarial-Financial Mathematics, University of Aegean, Karlovasi, Samos 83200, Greece

charts (Montgomery [11]). In addition to the p and np charts, Gan [5], Gan [6], Chang and Gan [4], Wu *et al.* [28] and Yeh *et al.* [29] studied CUSUM and EWMA control charts for monitoring binomial counts. All the above-mentioned control charts are based on the assumption of iid binomial rv. Even though the iid assumption is a common assumption, observations on a process will be autocorrelated when the sampling rate is very high, which, in turn, commonly happens because of the technological progress in automated sampling [13]. Clearly, in a variety of real-life problems, the iid assumption is violated.

In that case, the previously mentioned control charts cannot be used because they demonstrate an increased false alarm rate (FAR). One solution to this problem is to select first an appropriate model of integer-valued time series, and then to develop control charts based on this model. In particular, if it is of interest to monitor the number X of defects in a sample of n objects, then there is a finite number of possible values for X . Therefore, the appropriate integer-valued time series model must be such that X takes a finite number of possible values. Consequently, an appropriate model for correlated binomial counts needs to be selected first.

The monitoring of correlated binomial counts has been considered by Weiß [20] who developed and studied Shewhart and Moving Average (MA) control charts for monitoring a process that is properly described by the first-order binomial autoregressive model (binomial AR(1) or BAR(1)) of McKenzie [10] and Al-Osh and Alzaid [1].

However, in various real-life problems, the counts in the sample are from a population that contains inhomogeneous units. This is, for example, the case of the monthly number of districts in an area that are infected by a specific virus (Weiß and Pollett [26], Ristić *et al.* [16]). Clearly, it is more realistic to assume that the probability of the occurrence of the virus is not the same across the districts due to differences attributed to structure, quality of life and economic status. Another case (see Weiß and Kim [25]) is the monthly number of country members of the Eurozone which demonstrate price stability (in terms of inflation rates). Again, it is more realistic to assume that all the country members do not have the same probability to achieve a monthly inflation rate below a specific limit, due to the structural differences of their economies.

In those cases, the fixed relation between the mean and variance of the binomial distribution is violated. This relation is expressed in terms of the binomial index of dispersion (see, e.g. [25]) which is defined for an rv $X \in \{0, 1, \dots, n\}$ with mean μ and variance σ^2 as

$$I_d = \frac{n\sigma^2}{\mu(n - \mu)} \in (0, \infty),$$

and for the case of binomial distribution it is $I_d = 1$, for every $\pi \in (0, 1)$. For finite-range count data rv satisfying $I_d > 1$, there is an indication that this rv shows extra-binomial variation (overdispersion with regard to the binomial distribution). Therefore, when extra-binomial variation is present, alternative models to the BAR(1) model must be considered and the necessary control charts should be designed under the assumption of these models.

A popular choice to that direction is the first-order beta-binomial autoregressive model (beta-binomial AR(1) or BBAR(1)), proposed by Weiß and Kim [25]. The BBAR(1) model is an adequate choice for practical applications in which heterogeneity is observed between the units. Upper and lower one-sided Shewhart and CUSUM control charts for monitoring

a process that is properly described by a BBAR(1) process were proposed and studied by Rakitzis *et al.* [15].

Motivated by the work of Weiß [21], in this work, we propose and study upper and lower one-sided EWMA control charts for monitoring correlated counts that are described by a BAR(1) and a BBAR(1) model. The proposed chart can be considered as an alternative monitoring scheme to the CUSUM chart. The aim of this study is to investigate its statistical design as well as its performance in the detection of increasing and decreasing shifts in process mean level. Note also that Weiß [21] considered only the case of increases (upper one-sided charts). To the best of our knowledge, EWMA-type charts for monitoring BAR(1) and BBAR(1) processes have not been considered so far in the literature.

This paper is organized as follows. In Section 2, we present the essential properties of the BAR(1) (Section 2.1) and BBAR(1) (Section 2.1) models. Next, in Section 3 we present the upper and the lower one-sided EWMA charts for BAR(1) and BBAR(1) processes. Section 4 consists of the results of an extensive numerical study on the performance of the upper and lower one-sided EWMA control charts for monitoring BAR(1) and BBAR(1) processes. In Section 5, the practical application of the proposed schemes is illustrated via a real-data example. Finally, conclusion is summarized in Section 6.

2. Models of counts with finite range

2.1. The BAR(1) model

The BAR(1) model, proposed by McKenzie [10], is a simple model for autocorrelated processes of counts with a finite range. This model is based on the binomial thinning operator ‘ \circ ’, see [18]. More specifically, if X is a non-negative discrete rv and $\alpha \in (0, 1)$ then, by using the binomial thinning operator, it is possible to define the rv $\alpha \circ X = \sum_{i=1}^X Y_i$, as an alternative to the usual multiplication $\alpha \cdot X$. However, the result of $\alpha \circ X$ will always be integer. The rv $Y_i, i = 1, 2, \dots$, are iid Bernoulli rv with success probability α , independent also of the count data rv X . Therefore, the conditional distribution of $\alpha \circ X$, given $X = x$, is the binomial distribution $B(x, \alpha)$. We will refer to a process $\{X_t\}_{t \in \mathbb{N}}$, where $\mathbb{N} = \{1, 2, \dots\}$, as a BAR(1) process if it is of the form

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}), \tag{1}$$

where $\beta = \pi \cdot (1 - \rho), \alpha = \beta + \rho, \pi \in (0, 1), \rho \in (\max\{-\pi/(1 - \pi), -(1 - \pi)/\pi\}, 1)$ and $n \in \mathbb{N}$ is fixed. The condition on ρ guarantees that $\alpha, \beta \in (0, 1)$. Moreover, all thinnings are performed independently of each other and the thinnings at time t are independent of $X_s, s < t$, as well.

It is known (see, e.g. [20]) that the process $\{X_t\}_{t \in \mathbb{N}_0}$, where $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, is a stationary Markov chain with marginal distribution $B(n, \pi)$. Clearly, the marginal mean and variance are, respectively, equal to

$$\mathbb{E}(X_t) = n\pi, \quad \mathbb{V}(X_t) = n\pi(1 - \pi). \tag{2}$$

Moreover, the transition probabilities are

$$\begin{aligned}
 p_{k|l} &= P(X_t = k | X_{t-1} = l) \\
 &= \sum_{m=\max\{0, k+l-n\}}^{\min\{k, l\}} \binom{l}{m} \binom{n-l}{k-m} \alpha^m (1-\alpha)^{l-m} \beta^{k-m} (1-\beta)^{n-l+m-k}, \quad (3)
 \end{aligned}$$

for $k, l \in \{0, 1, 2, \dots, n\}$.

The conditional mean and variance, respectively, are equal to (see Weiß [24])

$$\mathbb{E}(X_t | X_{t-1}) = \rho \cdot X_{t-1} + n\beta, \quad \mathbb{V}(X_t | X_{t-1}) = \rho(1-\rho)(1-2\pi) \cdot X_{t-1} + n\beta(1-\beta), \quad (4)$$

while the autocorrelation function is given by $\rho(k) = \rho^k$ for $k \geq 0$.

Parameters π and ρ of the BAR(1) model can be estimated via the method of Conditional Maximum Likelihood (CML), when time series data are available. Let us assume that $x_1, \dots, x_T, T \in \mathbb{N}$, is a segment from a stationary BAR(1) process. Then the conditional likelihood function equals

$$L(\pi, \rho) = \binom{n}{x_1} \pi^{x_1} (1-\pi)^{n-x_1} \prod_{t=2}^T p_{x_t|x_{t-1}}, \quad (5)$$

where the probabilities $p_{x_{t-1}|x_t}$ are given in Equation (3). There is no closed-form formula for the maximum likelihood (ML) estimators $\hat{\pi}_{ML}, \hat{\rho}_{ML}$ of π, ρ and therefore, they are obtained by maximizing numerically the log-likelihood $l(\pi, \rho) = \log L(\pi, \rho)$. The corresponding standard errors can be computed from the observed Fisher’s Information matrix. Further details on the estimation of the parameters of a BAR(1) process are given in Weiß and Kim [24].

2.2. The BBAR(1) model

The Beta-binomial AR(1) model (BBAR(1)) is a simple and discrete model which can be used for correlated counts with a finite range. This model is appropriate for capturing the extra-binomial variation in the data as well as the heterogeneity among the n items. The BBAR(1) model is based on the concept of beta-binomial thinning, a generalization of the binomial thinning operation discussed previously.

Specifically, let α_ϕ be an rv which follows the beta distribution $Beta\left(\frac{(1-\phi)\cdot\alpha}{\phi}, \frac{(1-\phi)\cdot(1-\alpha)}{\phi}\right)$, where $\alpha, \phi \in (0, 1)$. The α_ϕ is independent of X . We will say that $\alpha_\phi \circ X$ is obtained from X by beta-binomial thinning if the operator “ \circ ” is the binomial thinning operator, performed independently of X and α_ϕ . Hence, the distribution of the (conditional) rv $\alpha_\phi \circ X$, given $X = x$, is the beta-binomial distribution $BB(x; \alpha, \phi)$ with probability mass function (pmf) given by

$$P(\alpha_\phi \circ X = w | X = x) = \binom{x}{w} \frac{B\left(w + \frac{1-\phi}{\phi} \cdot \alpha, x - w + \frac{1-\phi}{\phi} \cdot (1-\alpha)\right)}{B\left(\frac{1-\phi}{\phi} \cdot \alpha, \frac{1-\phi}{\phi} \cdot (1-\alpha)\right)},$$

where $B(\theta_1, \theta_2) = \int_0^1 u^{\theta_1-1}(1-u)^{\theta_2-1} du$ is the beta function. Next, we present the BBAR(1) model, which has one additional model parameter $\phi \in (0, 1)$, compared to the BAR(1) model. Further details and properties of the BBAR(1) model can be found in Weiß and Kim [25].

Similar to the case of the BAR(1) process, let $(\pi, \phi) \in (0, 1)^2$, $\rho \in (\max\{-\pi/(1-\pi), -(1-\pi)/\pi\}, 1)$, $n \in \mathbb{N}$ fixed, $\beta = \pi \cdot (1-\rho)$ and $\alpha = \beta + \rho$. We will refer to a process $\{X_t\}_{t \in \mathbb{N}}$ as the BBAR(1) process when it is defined by the recursion

$$X_t = \alpha_\phi \circ X_{t-1} + \beta_\phi \circ (n - X_{t-1}), \tag{6}$$

where all α_ϕ, β_ϕ and all thinnings are performed independently of each other. Also α_ϕ, β_ϕ and the thinnings at time t are independent of $X_s, s < t$.

The process $\{X_t\}_{t \in \mathbb{N}_0}$ is a homogeneous and ergodic Markov chain, with the one-step transition probability given by (for $k, l \in \{0, 1, 2, \dots, n\}$)

$$\begin{aligned} p_{k|l} &= P(X_t = k | X_{t-1} = l) \\ &= \sum_{m=\max\{0, k+l-n\}}^{\min\{k,l\}} \binom{l}{m} \binom{n-l}{k-m} \frac{B\left(m + \frac{(1-\phi)\alpha}{\phi}, l - m + \frac{(1-\phi)(1-\alpha)}{\phi}\right)}{B\left(\frac{(1-\phi)\alpha}{\phi}, \frac{(1-\phi)(1-\alpha)}{\phi}\right)} \\ &\quad \times \frac{B\left(k - m + \frac{(1-\phi)\beta}{\phi}, n - l - k + m + \frac{(1-\phi)(1-\beta)}{\phi}\right)}{B\left(\frac{(1-\phi)\beta}{\phi}, \frac{(1-\phi)(1-\beta)}{\phi}\right)}. \end{aligned} \tag{7}$$

The corresponding conditional mean and variance are

$$\mathbb{E}(X_t | X_{t-1}) = \alpha \cdot X_{t-1} + \beta \cdot (n - X_{t-1}) = \rho \cdot X_{t-1} + n\beta, \tag{8}$$

$$\begin{aligned} \mathbb{V}(X_t | X_{t-1}) &= \phi \cdot (\alpha(1-\alpha)) + (\beta(1-\beta)) \cdot X_{t-1}^2 + n\beta(1-\beta) \cdot (1 + \phi(n-1)) \\ &\quad + X_{t-1} \cdot (\rho(1-\rho)(1-2\pi)(1-\phi) - 2n\beta(1-\beta) \cdot \phi), \end{aligned} \tag{9}$$

while the stationary mean $\mu = \mathbb{E}(X_t)$ and variance $\sigma^2 = \mathbb{V}(X_t)$ are equal to

$$\mathbb{E}(X_t) = n\pi, \quad \mathbb{V}(X_t) = n\pi(1-\pi) \cdot \underbrace{\frac{(1-\phi)(1+\rho) + n\phi(1-2\pi(1-\pi)(1-\rho))}{(1-\phi)(1+\rho) + \phi(1-2\pi(1-\pi)(1-\rho))}}_{I_\phi} \tag{10}$$

It is not difficult to realize that BAR(1) and BBAR(1) models have the same marginal mean μ while the marginal variance of BBAR(1) model is the usual binomial variance, multiplied by a factor I_ϕ , which determines the deviation of the true stationary variance from the binomial one. As $\phi \rightarrow 0$, the $I_\phi \rightarrow 0$.

It should be noted that there is not a closed-form expression for the stationary marginal distribution of X_t for the BBAR(1) model. However, we can determine it numerically as follows: First, if the initial count X_0 follows the stationary marginal distribution, then the whole process $\{X_t\}_{t \in \mathbb{N}}$ becomes stationary. Therefore, the vector $\mathbf{p} = (p_0, \dots, p_n)^\top$ of the marginal probabilities $p_x = P(X_t = x)$, $x \in \{0, 1, \dots, n\}$, is the solution of the equation

$\mathbf{P}\mathbf{p} = \mathbf{p}$, where \mathbf{P} is the $(n + 1) \times (n + 1)$ transition probability matrix with conditional probability $p_{k|l}$ (given in Equation (7)) at the (l, k) entry.

Similar to the case of BAR(1) process, when a segment $x_1, x_2, \dots, x_T, T \in \mathbb{N}$ is available from a BBAR(1) process, the parameters π, ρ and ϕ can be estimated via the method of CML. The log-likelihood function, conditioned on x_1 , equals

$$\ell(\pi, \rho, \phi) = \sum_{t=2}^T \log p_{x_t|x_{t-1}},$$

where the transition probabilities are now given in (7). Again, the ML estimators $\hat{\pi}_{ML}, \hat{\rho}_{ML}$ and $\hat{\phi}_{ML}$ of π, ρ and ϕ are obtained by maximizing numerically the $\ell(\pi, \rho, \phi)$. Further details on estimation methods for the parameters of a BBAR(1) model can be found in Weiß and Kim [25].

3. Methodology

In this section, we develop one-sided EWMA control charts for monitoring BAR(1) and BBAR(1) processes. The aim is to detect quickly and accurately a change in the mean level $\mu \equiv \mu_X = E(X_t) = n\pi$ of the process. When the process is in-control (IC), we will denote it as $\mu_{0,X}$ while in the out-of-control state (OoC) it is denoted as $\mu_{1,X}$. In a similar manner, the IC (OoC) parameter values of the BAR(1) model are denoted as π_0 and ρ_0 (π_1 and ρ_1) while for the BBAR(1) model, the IC (OoC) parameter values are denoted as π_0, ρ_0 and ϕ_0 (π_1, ρ_1 and ϕ_1).

In several applications, practitioners are interested in detecting increases in the process mean level, from an IC value $\mu_{0,X}$ to an OoC value $\mu_{1,X} > \mu_{0,X}$. This is important because with this displacement, if X is the number of non-conforming items produced by a process, there is an increase on the average number of them. On the contrary, when there is a decrease in the mean level of the process, i.e. when $\mu_{1,X} < \mu_{0,X}$, then less non-conforming items are produced, which is an indication of process improvement. In this work we consider both cases.

The exponentially weighted moving average (EWMA) control chart was introduced by Roberts [17]. For $t = 1, 2, \dots$, the values of the following statistic are plotted on the chart

$$Q_t = \lambda X_t + (1 - \lambda)Q_{t-1}, \quad Q_0 = q_0, \tag{11}$$

where λ is a smoothing parameter such that $0 < \lambda \leq 1$. For small values of λ less weight is given to the most recent observation X_t and more weight is given to all the available observations since the beginning of process monitoring. This is a control chart with memory and it is more capable of than a Shewhart control chart in detecting shifts of small or medium magnitude in the mean level of the process. For $\lambda = 1$, the EWMA chart coincides with the usual Shewhart chart.

Before proceeding with the presentation of EWMA charts for BAR(1) and BBAR(1) processes, it should be noted that a special characteristic of these processes is that the values of $X_t, t \geq 1$, are integers. Clearly, by applying (11) directly to the data from a BAR(1) or a BBAR(1) process, the Q_t values do not remain integers. In this case, the performance of the chart can only be calculated approximately, since (see, e.g. [19]) the range of the

possible values for Q_t changes at each time t . Weiß [21] proposed and studied the upper one-sided s -EWMA control charts, $s \in \{1, 2, 4, \dots\}$ for monitoring a first-order integer-valued Poisson autoregressive (PINAR(1)) process. Specifically, for $s = 1$, the EWMA chart statistic is given by

$$Q_t = \text{round}(\lambda X_t + (1 - \lambda)Q_{t-1}), \quad Q_0 = q_0, \quad t = 1, 2, \dots, \tag{12}$$

where the Q_t values remain integers and $q_0 \in \mathbb{N}_0$. Clearly, this is achieved by applying on the result of (11) the function of rounding the number x to the nearest integer, i.e. the $\text{round}(x) = z$, if $x \in [z - 0.5, z + 0.5]$. We will refer to this chart as the 1-EWMA, which gives an OoC signal when for the first time $Q_t \geq UCL_1$. The UCL_1 is an appropriate upper control limit. In a similar manner, for the lower one-sided 1-EWMA, with a lower control limit LCL_1 , an OoC signal is given when $Q_t \leq LCL_1$. It is also noteworthy that in the case of serial independence of binomial rv, the 1-EWMA chart coincides with the modified EWMA chart in Gan [6].

In order to evaluate the performance of 1-EWMA chart, it is necessary to determine its run length distribution, which is defined as the distribution of the rv $L = \min\{j : Q_j \geq UCL_1\}$. The rv L is defined as the number of points plotted on the chart until it gives for the first time an OoC signal. Since the possible values of the plotted statistic of the 1-EWMA chart are integers, by using the Markov chain method, it is possible to compute exactly the entire distribution of rv L . Note also that Gan [5] used the modified EWMA statistic in (12) in order to preserve the integer character of X_t values (they are integers and the Q_t values remain integers, as well) and therefore, compute exactly the run length properties of the proposed chart. In addition, in an attempt to extend Gan’s work, we investigate the performance of the modified EWMA statistic in the case of serially dependent binomial observations. The Markov chain method is very well documented in the related literature (see, e.g. Brook and Evans [2], Yontay *et al.* [30], Knoth [7], Weiß [21], Weiß [22] and references therein) and we do not provide further details about it. Here, we used the method as described in Weiß [21].

The expected value $\mathbb{E}(L)$, also known as average run length (ARL), is the most common performance measure of a control chart. The ARL expresses the average number of points to be plotted on the chart until it gives for the first time an OoC signal. Following Rakitzis *et al.* [15] (see also Weiß [23]), in this work, the IC performance of the proposed schemes is evaluated in terms of the zero-state ARL ($zsARL$) which is the expected number of points plotted on the chart until the first (false) alarm is given. For an OoC process, the performance of the proposed schemes is evaluated in terms of the steady-state ARL ($ssARL$) which gives an approximation of the true mean delay for detection after a change in the process, from the IC state to the OoC state. We assume that a change in process happens at an (unknown) change-point $\tau \in \{1, 2, \dots\}$. Specifically, for $t < \tau$, the process is in the IC state while for $t \geq \tau$, the process has shifted to the OoC state. Therefore, the $ssARL$ expresses the expected number of points to be plotted on the chart until it gives for the first time an indication of an OoC process, given that the process has been operated for “sufficient time” in control. According to Weiß and Testik [27], the $zsARL$ and the $ssARL$ are substantially different in the case of monitoring processes with correlated counts.

The statistical design of the 1-EWMA (either for BAR(1) or BBAR(1) processes) requires the determination of the values for the design parameters, λ and UCL_1 . These values are generally different for different processes. We follow the steps below to design an

upper one-sided 1-EWMA chart with the desired ARL_0 (see also [21]); the case of the lower one-sided 1-EWMA chart is treated in a similar manner (after some necessary but straightforward modifications) and, due to space economy, the details are omitted.

- Step 1 We choose the IC values of the design parameters n, π_0, ρ_0 (for BAR(1) process) or n, π_0, ρ_0, ϕ_0 (for BBAR(1) process) and the desired in-control ARL_0 value for the $zsARL$.
- Step 2 We choose an initial value for the UCL_1 such that the IC $zsARL$ of the corresponding upper one-sided Shewhart chart is lower than the desired value ARL_0 .
- Step 3 We decrease the value of $\lambda \in (0, 1]$ to adjust the IC $zsARL$ until we get a value as close as possible to the desired value ARL_0 . So it turns out the pair of the values (λ, UCL_1) .

The design procedure continues, by reducing further the initial value of UCL_1 and then searching again for the value of λ until the IC $zsARL$ value attains a value as close as possible to the desired value ARL_0 . The UCL_1 decreases so that it is always $UCL_1 \geq \mu_{0,X}$. Clearly, there is at least one pair of (λ, UCL_1) values that satisfies the desired IC performance. Finally, among all the possible designs (λ, UCL_1) , it turns out the optimal pair of the values (λ, UCL_1) for the 1-EWMA chart, which satisfies the following two conditions: (i) its ARL performance is as close as possible to the desired ARL_0 value and (ii) it attains the minimum possible OoC $ssARL$ value for a pre-defined OoC value $\mu_{1,X} > \mu_{0,X}$. It should be mentioned that since it is not possible, because of the discrete nature of BAR(1) and BBAR(1) processes, to achieve exactly the desired IC performance, we applied the following criterion (see Castagliola *et al.* [3]) to consider a design (λ, UCL_1) as such that gives IC $zsARL$ as close as possible to the desired ARL_0 value:

$$\frac{|zsARL - ARL_0|}{ARL_0} \leq 0.05$$

In a similar manner, the statistical design of the lower one-sided 1-EWMA chart requires the determination of the optimal pair (λ, LCL_1) such that the achieved IC $zsARL$ is as close as possible to the desired ARL_0 value and the OoC $ssARL$ value is the minimum possible for an OoC value $\mu_{1,X} < \mu_{0,X}$.

It is worth mentioning that we may achieve a better approximation of the desired value ARL_0 by varying appropriately the starting value q_0 . However, in this work, we assume that $q_0 = 0$ so as not to give the Fast Initial Response feature [8] in the EWMA charts.

Apart from the 1-EWMA chart, Weiß [21] proposed also the s -EWMA chart, $s \in \mathbb{N}$, where the values of the statistic

$$Q_t^{(s)} = s - \text{round}(\lambda X_t + (1 - \lambda)Q_{t-1}^{(s)}), \quad Q_0^{(s)} = q_0^{(s)}, \tag{13}$$

are plotted on the chart, for $t = 1, 2, \dots$ and $q_0^{(s)} \in \mathbb{N}$. Likewise the case of 1-EWMA chart, we assume $q_0^{(s)} = 0$. The s -round function is defined as $s - \text{round}(x) = z$ if-f $x \in [z - 0.5 \cdot s, z + 0.5 \cdot s]$ and can be viewed as an extension of the usual rounding function, since it maps x to the nearest fraction with denominator s . In this way, the exact calculation of the run length distribution of s -EWMA chart remains feasible via the Markov chain method.

Furthermore, the $s - \text{round}(\dots)$ is a fractional operation rounding; by using larger s values, it is possible to correct the problem of oversmoothing that arises on the 1-EWMA chart, due to the rounding of the Q_t values, especially for low process mean level and small λ values. This adds more flexibility on the statistical design of the chart, especially when we are interested in determining the values for its design parameters, in order to achieve the desired IC performance.

Successive values of the statistic $Q_t^{(s)}$, $t = 1, 2, \dots$ are plotted on a chart with an upper control limit $UCL_s = u'/s$, that is, the process is considered as being IC unless an OoC signal is triggered, i.e. when $Q_t^{(s)} \geq UCL_s$. Note also that in the case of 1-EWMA chart, the $UCL_1 = u'/1 = u'$. The statistical design of the s -EWMA chart requires the determination of three design parameters (s, λ, UCL_s). We follow the steps below in order to design an upper one-sided s -EWMA chart with the desired IC performance (for more details, see also Weiß [21]).

- Step 1 We choose the IC values of the process parameters n, π_0, ρ_0 (for BAR(1) process) or n, π_0, ρ_0, ϕ_0 for BBAR(1) process) and the desired IC ARL_0 value. Furthermore, we choose the value of $s \in \{2, 3, 4, \dots\}$.
- Step 2 For the given s value, we start from the optimal pair of values (λ_1, UCL_1) , for the 1-EWMA chart, say (λ^*, UCL_1^*) , and modify u' such that the pair $(\lambda^*, u'/s)$ results in an upper one-sided s -EWMA chart with IC $zsARL$ value lower than the desired ARL_0 value. The upper control limit is now UCL_s .
- Step 3 For the value UCL_s obtained in Step 2, we search for a new λ value, say λ_s with $\lambda_s \neq \lambda^*$, such that the s -EWMA chart with (λ_s, UCL_s) has an IC $zsARL$ value as close as possible to the desired ARL_0 value. Hence, it turns out the triple (s, λ_s, UCL_s) .

The design procedure continues, by applying the above steps for the remaining pairs of values (λ, UCL_1) that emerged during the study of the 1-EWMA control chart. Therefore, for the statistical design of the s -EWMA chart, the optimal triple (s, λ_s, UCL_s) that emerged, for different values of the s , are such that the IC $zsARL$ is as close as possible to the desired ARL_0 value and it attains the minimum possible OoC $ssARL$ value for a pre-defined OoC value $\mu_{1,X} > \mu_{0,X}$. Again, we applied the same criterion as in the case of the statistical design of the 1-EWMA in order to pick up the designs (s, λ_s, UCL_s) that have an IC performance *as close as possible* to the desired ARL_0 value. The statistical design for the lower one-sided s -EWMA chart is similar to the upper one-sided case, after some direct but necessary modifications. The details are left to the readers.

Remark 3.1: Generally speaking, parameter s is actually another chart parameter, except for the λ and the control limit. According to Weiß [21], in order to avoid the oversmoothing of the EWMA values, which affects severely the ability of the chart in the detection of an OoC situation, s values that are larger than 1 are recommended. Weiß [21] considered $s \in \{2, 4\}$ as a good trade-off between oversmoothing and time-consuming computations (for large s and small λ values). Therefore, in this work, we determine the optimal design (values λ_s and UCL_s or LCL_s) for $s = 2$ and for $s = 4$ and then, between them, we picked up the one that optimizes the detection ability of the chart, at the given shift in $\mu_{0,X}$.

4. Numerical study

In this section, we present the results of an extensive numerical study on the performance of s -EWMA charts in the monitoring of BAR(1) and BBAR(1) processes. For the IC design parameters, we assume that $\mu_0 \in \{5, 10\}$, $\rho_0 \in \{0.25, 0.50, 0.75\}$ and $\phi_0 \in \{0, 0.025, 0.05\}$. With $\phi_0 = 0$, we denote the case of BAR(1) process. We also consider two different sample sizes $n \in \{15, 30\}$. In addition, when the process is OoC, we assume that $\rho_1 = \rho_0$ and $\phi_1 = \phi_0$ and the presence of assignable causes of variation affects only the parameter π_0 , which changes from π_0 to $\pi_1 = \delta \cdot \pi_0$, $\delta > 1$, with $\delta \in \{1.2, 1.4\}$ (upward shifts) and $\delta \in \{0.8, 0.6\}$ (downward shifts), i.e. we want to detect 20% and 40% changes in the IC mean level of the process. These are the shifts of interest and we want to obtain the optimal charts for their detection. The desired ARL_0 value is 370.4.

Tables 1 and 2 consist of the optimal statistical design of the upper one-sided (Table 1) and the lower one-sided (Table 2) s -EWMA chart. The optimal values of the design parameters (s, λ_s, UCL_s) for the s -EWMA chart are given in the respective columns. Note also that UCL_s is given in the form u'/s while $LCL_s = l'/s$.

The IC parameter values of the process are given in the columns entitled ' ϕ_0 ', ' μ_0 ', ' n ' and ' ρ_0 ' while the IC performance for each chart is given in the column ' $zsARL$ ' and the OoC performance ($ssARL$ values) for $\mu_{1,X} = \delta\mu_{0,X}$ ($\delta \in \{1.2, 1.4\}$ or $\delta \in \{0.8, 0.6\}$) is given in the column ' $ssARL$ '.

Tables 1 and 2 reveal that for $s > 1$ (i.e. for $s = 2$ or $s = 4$) the s -EWMA chart has (in the most of the considered IC scenarios) an IC $zsARL$ value close to the desired ARL_0 value. As already said, because of the discrete nature of the considered models, it was impossible to achieve exactly the desired ARL_0 value. However, the difference in the obtained IC $zsARL$ values, in almost all cases in Tables 1 and 2 is at most 5%. We observe that in the most of the cases, optimal values for the set of chart's parameters are for larger than 1 values of the rounding parameter s (e.g. $s = 2$ or $s = 4$). Also, these values give an IC $zsARL$ closer to the desired ARL_0 value. This is an indication of an increased flexibility as well as of an optimized detection ability for the s -EWMA chart. From a practical point of view and regarding the determination of the chart's parameters λ_s and UCL_s , for a given s , as the value of UCL_s decreases, the value of λ_s decreases, which is usually desirable, so that the EWMA charts to be more efficient than Shewhart charts. In a similar manner, for the lower one-sided charts, as LCL_s increases, λ_s decreases, in order to achieve the desired IC performance.

Furthermore, when extra-binomial variation is present on the process, larger values for s and smaller values for λ_s are needed, in order to achieve the desired IC performance. Otherwise, if the extra-binomial variation is not taken into account and proceed with the values (λ_s, UCL_s) (or (λ_s, LCL_s)) obtained under the BAR(1) model (i.e. as if no extra-binomial is present), the resulting chart will demonstrate an increased FAR. This is, for example, the case of upper one-sided s -EWMA charts for $(\phi_0, \mu_0, n, \rho_0) = (0, 10, 15, 0.5)$ and $(0.05, 10, 15, 0.5)$, where the UCL_s has to be increased from 54/4 (under the BAR(1) model) to 57/4 (under the BBAR(1) model) in order to preserve the IC performance (as close as possible) at the desired levels. Note also that a proper adjustment of the λ_s is necessary while the detection ability of the chart is clearly affected by the presence of extra-binomial variation; under the BBAR(1) model, the $ssARL$ at $\delta = 1.2$ equals 30.64 while under the BAR(1) equals 19.05. Both models have the same marginal mean but BBAR(1) is overdispersed.

Table 1. Design table for the upper one-sided s-EWMA charts.

ϕ_0	μ_0	n	ρ_0	δ	s	λ_s	u'	UCL_s	$zsARL$	$ssARL$	ϕ_0	μ_0	n	ρ_0	δ	s	λ_s	u'	UCL_s	$zsARL$	$ssARL$	ϕ_0	μ_0	n	ρ_0	δ	s	λ_s	u'	UCL_s	$zsARL$	$ssARL$	
0	5	15	0.25	1.2	4	0.15	27	27/4	348.08	30.06	0.025	5	15	0.25	1.2	4	0.61	37	37/4	361.06	65.50	0.05	5	15	0.25	1.2	4	0.53	37	37/4	398.44	71.91	
				1.4	4	0.15	27	27/4	348.08	10.44					1.4	4	0.53	37	37/4	398.44	21.34												
			0.50	1.2	4	0.11	27	27/4	345.24	41.55				0.50	1.2	1	0.60	10	10	359.41	79.24				0.50	1.2	2	0.70	21	21/2	366.92	95.37	
				1.4	4	0.11	27	27/4	345.24	15.62					1.4	2	0.70	21	21/2	366.92	34.03												
			0.75	1.2	1	0.25	9	9	371.31	88.76				0.75	1.2	4	0.41	37	37/4	385.18	103.64				0.75	1.2	4	0.62	41	41/4	371.70	119.21	
				1.4	1	0.25	9	9	371.31	34.03					1.4	4	0.62	41	41/4	371.70	50.74												
	30	0.25	15	0.25	1.2	4	0.33	33	33/4	352.94	48.55	30	0.25	15	0.25	1.2	4	0.25	33	33/4	365.14	60.07	30	0.25	15	0.25	1.2	4	0.54	45	45/4	372.73	108.10
					1.4	4	0.33	33	33/4	352.94	15.09					1.4	4	0.54	45	45/4	372.73	41.04											
				0.50	1.2	1	0.12	7	7	379.56	67.34				0.50	1.2	2	0.31	19	19/2	372.13	88.46				0.50	1.2	4	0.62	49	49/4	379.84	132.89
					1.4	1	0.12	7	7	379.56	25.93					1.4	2	0.31	19	19/2	372.13	32.94											
				0.75	1.2	2	0.08	14	14/2	374.46	81.85				0.75	1.2	4	0.59	45	45/4	371.75	130.84				0.75	1.2	4	0.58	49	49/4	369.66	150.91
					1.4	2	0.08	14	14/2	374.46	35.04					1.4	4	0.58	49	49/4	369.66	73.96											
10	15	15	0.25	1.2	4	0.56	53	53/4	375.92	13.70	10	15	15	0.25	1.2	4	0.80	57	57/4	366.07	25.12	10	15	0.25	1.2	4	0.13	47	47/4	373.92	11.97		
				1.4	4	0.56	53	53/4	375.92	3.03					1.4	4	0.80	57	57/4	366.07	4.13												
			0.50	1.2	4	0.51	54	54/4	352.91	19.05				0.50	1.2	4	0.65	56	56/4	365.26	25.26			0.50	1.2	4	0.67	57	57/4	349.83	30.64		
				1.4	4	0.51	54	54/4	352.91	4.63					1.4	4	0.65	56	56/4	365.26	5.43												
			0.75	1.2	2	0.37	27	27/2	390.62	32.77				0.75	1.2	4	0.41	55	55/4	372.72	37.02			0.75	1.2	4	0.59	57	57/4	373.97	47.69		
				1.4	2	0.37	27	27/2	390.62	9.16					1.4	4	0.41	55	55/4	372.72	10.00												
	30	0.25	15	0.25	1.2	2	0.26	27	27/2	353.54	20.01	30	0.25	15	0.25	1.2	4	0.58	65	65/4	370.73	41.85	30	0.25	15	0.25	1.2	4	0.74	73	73/4	360.95	61.52
					1.4	2	0.26	27	27/2	353.54	6.34					1.4	4	0.58	65	65/4	370.73	10.72											
				0.50	1.2	4	0.48	61	61/4	364.46	38.90				0.50	1.2	4	0.53	66	66/4	365.93	53.27				0.50	1.2	4	0.16	56	56/4	364.22	42.69
					1.4	4	0.48	61	61/4	364.46	10.99					1.4	4	0.53	66	66/4	365.93	15.56											
				0.75	1.2	4	0.05	47	47/4	349.36	39.03				0.75	1.2	4	0.60	69	69/4	373.90	78.41				0.75	1.2	4	0.60	73	73/4	374.48	93.23
					1.4	4	0.05	47	47/4	349.36	17.14					1.4	4	0.60	69	69/4	373.90	27.44											

Table 2. Design table for the lower one-sided s -EWMA charts.

ϕ_0	μ_0	n	ρ_0	s	λ_s	l'	LCL_s	$zsARL$	$ssARL$	
									$\delta = 0.8$	$\delta = 0.6$
0	5	15	0.25	4	0.52	7	7/4	366.63	60.64	14.24
			0.50	2	0.43	3	3/2	388.09	99.30	24.79
			0.75	4	0.25	6	6/4	361.05	256.22	60.87
		30	0.25	4	0.53	6	6/4	384.16	74.98	17.65
			0.50	4	0.74	3	3/4	379.04	106.21	34.61
			0.75	2	0.78	1	1/2	372.40	131.39	25.84
	10	15	0.25	4	0.53	23	23/4	367.67	32.08	5.62
			0.50	4	0.65	22	22/4	352.29	26.22	6.64
			0.75	4	0.67	21	21/4	372.61	41.99	12.39
		30	0.25	1	0.93	3	3	359.11	46.42	8.84
			0.50	4	0.61	18	18/4	370.63	47.72	11.24
			0.75	4	0.38	17	17/4	359.87	129.67	28.53
0.025	5	15	0.25	4	0.86	2	2/4	358.43	83.96	24.73
			0.50	1	0.85	0	0	352.30	94.62	31.30
			0.75	4	0.36	5	5/4	386.76	122.62	41.93
		30	0.25	4	0.74	2	2/4	373.14	90.76	26.53
			0.50	4	0.37	5	5/4	384.39	111.56	31.94
			0.75	4	0.52	2	2/4	359.15	130.90	52.60
	10	15	0.25	4	0.63	22	22/4	369.72	23.15	5.16
			0.50	1	0.90	4	4	353.08	32.00	8.16
			0.75	1	0.50	5	5	372.04	70.80	17.88
		30	0.25	2	0.74	7	7/2	363.02	47.85	10.64
			0.50	4	0.69	14	14/4	357.92	57.50	14.83
			0.75	2	0.54	7	7/2	363.54	80.21	24.99
0.050	5	15	0.25	2	0.76	1	1/2	355.15	79.91	23.25
			0.50	4	0.74	2	2/4	370.84	96.32	31.25
			0.75	4	0.61	3	3/4	355.01	121.13	48.20
		30	0.25	4	0.48	4	4/4	363.11	96.54	27.21
			0.50	4	0.51	2	2/4	364.58	109.52	36.61
			0.75	4	0.38	2	2/4	362.52	138.33	56.57
	10	15	0.25	4	0.80	18	18/4	370.33	28.94	6.31
			0.50	4	0.75	18	18/4	368.05	36.39	9.17
			0.75	4	0.66	18	18/4	368.82	53.45	16.16
		30	0.25	1	0.64	3	3	378.90	40.22	12.35
			0.50	4	0.53	14	14/4	365.52	66.35	17.40
			0.75	2	0.75	5	5/2	366.02	91.16	30.87

For the lower one-sided charts, this can be noticed for $(\phi_0, \mu_0, n, \rho_0) = (0, 10, 30, 0.50)$ and $(0.025, 10, 30, 0.50)$, where the LCL_s has to be decreased from $18/4$ (under the BAR(1) model) to $14/4$ (under the BBAR(1) model) in order to preserve the IC performance (as close as possible) at the desired levels.

Tables 3 and 4 give the results of a comparative study between the one-sided s -EWMA, CUSUM and Shewhart charts, for both the BAR(1) and BBAR(1) processes. For each shift $\delta \in \{1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 2.0\}$ (upward shifts) and $\delta \in \{1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.2\}$ (downward shifts), the ARL profiles are given for the respective charts. Also, the IC parameter values of the process are given in the columns ' ϕ_0 ', ' $\mu_{0,X}$ ', ' n ' and ' ρ_0 '. It is worth mentioning that the shifts occur only in $\mu_{0,X}$ and it is $\mu_{1,X} = \delta \cdot \mu_{0,X} = n \cdot (\delta \cdot \pi)$, i.e. only π is affected by the presence of assignable causes. Once again, because of the discrete nature of the BAR(1) and BBAR(1) models, it was impossible to have exactly the same $zsARL$ value for all the schemes under comparison. However, since the IC $zsARL$ value between the different schemes varies, the interpretation should be made with caution.

Table 3. ARL comparison, upper one-sided Shewhart, s-EWMA and CUSUM charts.

ϕ_0	μ_0	n	ρ_0	δ	Shewhart	1-EWMA	2-EWMA	4-EWMA	CUSUM			
0	5	30	0.25	1.0	502.72	503.62	492.27	489.95	490.07			
				1.1	223.27	204.62	135.39	134.48	159.42			
				1.2	110.34	95.09	55.90	51.95	63.85			
				1.3	59.68	49.60	30.95	25.86	30.89			
				1.4	34.85	28.57	20.66	15.51	17.61			
				1.5	21.74	17.91	15.46	10.59	11.45			
				1.7	9.92	8.61	10.42	6.29	6.34			
				2.0	4.13	4.09	7.20	3.93	3.78			
				UCL				12	11	15/2	31/4	
				λ				1	0.78	0.18	0.24	
h								8				
k								7				
ϕ_0	μ_0	n	ρ_0	δ	Shewhart	1-EWMA	2-EWMA	4-EWMA	CUSUM			
0.025	5	15	0.75	1.0	393.26	320.84	379.95	385.18	390.13			
				1.1	208.46	164.18	190.31	185.98	154.32			
				1.2	119.57	95.28	107.32	103.64	77.85			
				1.3	73.69	58.44	65.46	61.43	46.76			
				1.4	48.39	37.75	41.30	41.16	31.99			
				1.5	33.60	26.68	28.96	28.20	23.99			
				1.7	18.60	15.41	16.46	16.36	16.01			
				2.0	9.88	8.69	9.02	9.15	10.98			
				UCL				11	10	19/2	37/4	
				λ				1	0.5	0.46	0.41	
h								30				
k								6				
ϕ_0	μ_0	n	ρ_0	δ	Shewhart	1-EWMA	2-EWMA	4-EWMA	CUSUM			
0.05	5	30	0.50	1.0	385.96	369.19	348.06	378.34	385.29			
				1.1	247.82	217.53	202.43	181.05	150.51			
				1.2	163.98	135.21	124.80	99.72	74.55			
				1.3	111.71	87.98	80.83	60.34	44.10			
				1.4	78.25	59.71	54.75	39.53	29.82			
				1.5	56.30	42.11	38.64	27.71	22.13			
				1.7	31.42	23.22	21.44	15.75	14.48			
				2.0	15.49	11.75	11.01	9.02	9.63			
				UCL				15	13	25/2	37/4	
				λ				1	0.69	0.65	0.25	
h								34				
k								6				
ϕ_0	μ_0	n	ρ_0	δ	Shewhart	1-EWMA	2-EWMA	4-EWMA	CUSUM			
0.05	10	30	0.75	1.0	345.27	367.46	351.50	360.94	362.10			
				1.1	175.59	181.17	167.58	157.92	134.61			
				1.2	96.96	96.88	89.34	79.83	65.11			
				1.3	57.95	55.53	52.45	46.79	38.27			
				1.4	37.23	33.71	33.51	30.08	25.96			
				1.5	25.51	21.41	23.02	20.78	19.44			
				1.7	14.04	9.49	12.90	12.20	13.05			
				2.0	7.60	3.26	7.24	7.39	9.07			
				UCL				20	20	37/2	66/4	
				λ				1	0.93	0.63	0.30	
h								44				
k								12				

For the one-sided CUSUM chart (see Rakitzis *et al.* [15]), the plotted statistic for the upper one-sided chart is given by

$$C_0^+ = 0$$

$$C_t^+ = \max(0, C_{t-1}^+ + X_t - k^+), \quad t \geq 1.$$

Table 4. ARL comparison, lower one-sided Shewhart, s-EWMA and CUSUM charts.

ϕ_0	μ_0	n	ρ_0	δ	Shewhart	1-EWMA	2-EWMA	4-EWMA	CUSUM			
0	10	30	0.5	1	332.96	352.21	353.77	370.63	339.04			
				0.9	124.73	120.79	127.00	131.67	78.11			
				0.8	53.59	45.59	46.46	47.72	27.16			
				0.7	24.86	20.65	20.72	21.13	13.45			
				0.6	14.27	11.09	11.06	11.24	8.50			
				0.5	8.00	6.88	6.90	6.97	6.22			
				0.4	5.32	4.80	4.83	4.88	4.97			
				0.2	2.97	2.97	3.04	3.06	3.67			
				LCL				3	4	9/2	18/4	
				λ				1	0.65	0.6	0.61	
h								14				
k								8				
ϕ_0	μ_0	n	ρ_0	δ	Shewhart	1-EWMA	2-EWMA	4-EWMA	CUSUM			
0.025	10	30	0.25	1	378.38	392.10	363.02	373.38	388.22			
				0.9	150.30	132.53	126.76	127.26	56.47			
				0.8	64.34	50.04	47.85	48.41	20.13			
				0.7	29.85	21.84	20.95	21.26	11.41			
				0.6	15.09	11.02	10.64	10.79	7.93			
				0.5	8.36	6.37	6.21	6.27	6.12			
				0.4	5.07	4.15	4.08	4.10	5.02			
				0.2	2.40	2.32	2.30	2.30	3.76			
				LCL				2	3	7/2	14/4	
				λ				1	0.78	0.74	0.76	
h								24				
k								9				
ϕ_0	μ_0	n	ρ_0	δ	Shewhart	1-EWMA	2-EWMA	4-EWMA	CUSUM			
0.05	10	15	0.25	1	507.28	456.25	500.75	502.23	460.43			
				0.9	143.14	107.49	113.72	115.49	43.23			
				0.8	48.68	33.32	30.83	33.43	13.98			
				0.7	19.79	13.48	12.17	13.25	7.88			
				0.6	9.51	6.48	6.38	6.77	5.53			
				0.5	5.32	4.16	4.08	4.18	4.31			
				0.4	3.38	2.89	2.98	2.95	3.58			
				0.2	1.82	1.80	2.01	1.90	2.74			
				LCL				3	4	10/2	19/4	
				λ				1	0.79	0.59	0.69	
h								15				
k								9				

The chart gives an OoC signal when for the first time $C_t^+ > h^+$. Parameter k^+ is known as reference value and control limit h^+ is the decision interval. Rakitzis *et al.* [15] suggested for $k^+ = \lfloor \mu_{X,0} \rfloor + 1$ or $\lfloor \mu_{X,0} \rfloor + 2$ in order to assure that C_t^+ values are integers and the performance of the CUSUM chart can be evaluated exactly by means of the Markov chain method. Also, these values for k^+ guarantee an optimized detection ability for the chart. In a similar manner, the plotted statistic for the lower one-sided CUSUM chart is

$$C_0^- = 0$$

$$C_t^- = \max(0, C_{t-1}^- - X_t + k^-), \quad t \geq 1$$

and the chart gives an OoC signal when for the first time $C_t^- > h^-$. The design parameters k^-, h^- are defined analogously; the suggested k^- values are $\lceil \mu_{X,0} \rceil - 1$ or $\lceil \mu_{X,0} \rceil - 2$.

From the ARL profiles of the charts in Tables 3 and 4, it is clear that EWMA-type charts are superior to the Shewhart-type charts. The superiority is obvious in cases where the IC $zsARL$ of the Shewhart-type chart is lower than the respective value for the s-EWMA chart

but the $ssARL$ of the latter, at a given shift, is lower than the one corresponds to the Shewhart chart. This means that the Shewhart chart signals, on average, more false alarms (which is not desirable) than the s -EWMA chart whereas the s -EWMA chart detects more quickly the given shift. Also, this can be verified by comparing directly the respective entries in Tables 1–2 with Tables 1 and 3 in Rakitzis *et al.* [15]. Also, either in the upper or the lower one-sided case, the performance of the respective EWMA control charts becomes better as the rounding parameter s increases. The comparison of the ARL profiles between the s -EWMA charts and the CUSUM charts for BAR(1) and BBAR(1) reveals that CUSUM charts are very powerful charts in the detection of small and moderate shifts in the mean level of the process. Only for specific IC scenarios, the performance of the proposed s -EWMA charts is better or at least comparable with the performance of the corresponding one-sided CUSUM charts; this is mostly happens for moderate to large shifts in $\mu_{0,X}$, i.e. for $\delta > 1.5$ (upward shifts) or for $\delta \leq 0.5$ (downward shifts).

5. Application

In this section, we present an application of the proposed s -EWMA charts to real data for monitoring BAR(1) and BBAR(1) processes. These data refer to the regional spread of an infection in Germany within a year. Specifically, we have the weekly number X_t of regions in Germany, with a new case of hantavirus infection in 2011, for $T = 52$ weeks. The number of regions is $n = 38$. More details on this dataset can be found in Weiß and Pollett [26] and Ristić *et al.* [16].

The time series plot in Figure 1 shows that the values in the sample are between 0 and 11. The sample mean is 4.173 and the sample variance is 7.793. Furthermore, the binomial index of dispersion is $I_d = 2.098 > 1$, which indicates extra-binomial variation. In addition, it is reasonable to assume that the probability of the occurrence of a new infection may not be the same in all regions as they usually differ (e.g. according to various socio-economic criteria). The empirical first-order autocorrelation function equals $\rho(1) = 0.634$. Here, for the sake of illustration, we follow the work of Weiß and Pollett [26] who considered as an appropriate model for this dataset a first-order autoregressive model. However, it should be mentioned that Ristić *et al.* [16] found that a model of order 3 (i.e. a BAR(3)) has a better fit on these data, than the BAR(1). Also, Ristić *et al.* [16] did not consider a higher order BBAR model. It goes without saying that it is of interest to analyse further this dataset by fitting also a higher-order Binomial or Beta-binomial models and then propose suitable methods for process monitoring. However, since the intention is to show how the proposed control charts can be applied in practice, we will not consider more complex models and they are left for future research.

Next, we fit both the BAR(1) and BBAR(1) models to the data and we estimate (via the maximum likelihood method) the unknown parameters for both models. Also, in order to choose between the two models the one that fits better to the data, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are used, which are given by

$$AIC = -2l_{\max} + \kappa \cdot 2, \quad BIC = -2l_{\max} + \kappa \cdot \log(T).$$

The l_{\max} is the maximum value for the log-likelihood function, κ is the number of model parameters ($\kappa = 2$ for the BAR(1) model and $\kappa = 3$ for the BBAR(1) model) and $T = 52$.

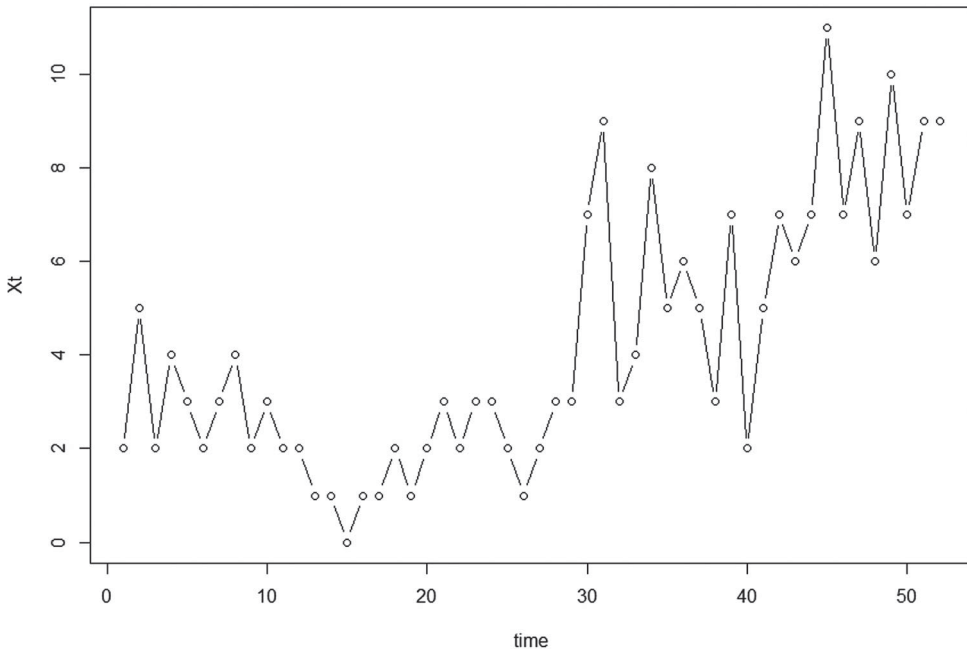


Figure 1. Weekly number of districts with new cases of Hantavirus.

Table 5. Fitting the BAR(1) and BBAR(1) models in the Hantavirus data.

Model	$\hat{\pi}$	$\hat{\rho}$	$\hat{\phi}$	AIC	BIC
BAR(1)	0.111 (0.013)	0.529 (0.070)		227.029	230.931
BBAR(1)	0.112 (0.179)	0.564 (0.074)	0.029 (0.016)	217.929	223.782

The standard errors of the estimates are given in the parentheses. The results are given in Table 5. The estimates for the parameters of the BAR(1) model coincide with the ones listed in the first row of Table 3 in Ristić *et al.* [16]; the differences are due to rounding.

Clearly, the BBAR(1) model has a better fit than the BAR(1) model since it attains the minimum value for both AIC and BIC. Therefore, we assume that the process $\{X_t\}_{t \in \mathbb{N}}$ is a BBAR(1) process where its true parameter values are equal to the respective estimates in Table 5. By following the steps that are described in Section 3, we develop upper and lower one-sided Shewhart, CUSUM and s -EWMA charts, $s = 1, 2, 4$, for the statistical monitoring of the process. We choose, for illustrative purposes, $ARL_0 = 370.4$ and the values of the design parameters for each chart are such that they have their IC $zsARL$ value as close as possible to the desired ARL_0 value. Next, we present our findings, by starting from the upper-sided case. We consider the upper one-sided BBAR(1) Shewhart chart with $UCL = 13$ and $zsARL = 324.52$, the upper-sided CUSUM chart with $k^+ = 5$, $h^+ = 35$ and $zsARL = 378.75$, the 1-EWMA chart with $(\lambda_1, UCL_1) = (0.43, 10)$ and $zsARL = 352.85$, the 2-EWMA chart with $(\lambda_2, UCL_2) = (0.38, 19/2)$ and $zsARL = 354.62$ and the 4-EWMA chart with $(\lambda_4, UCL_4) = (0.36, 37/4)$ and $zsARL = 370.60$. The upper one-sided charts for the Hantavirus data are given in Figures 2 and 3.

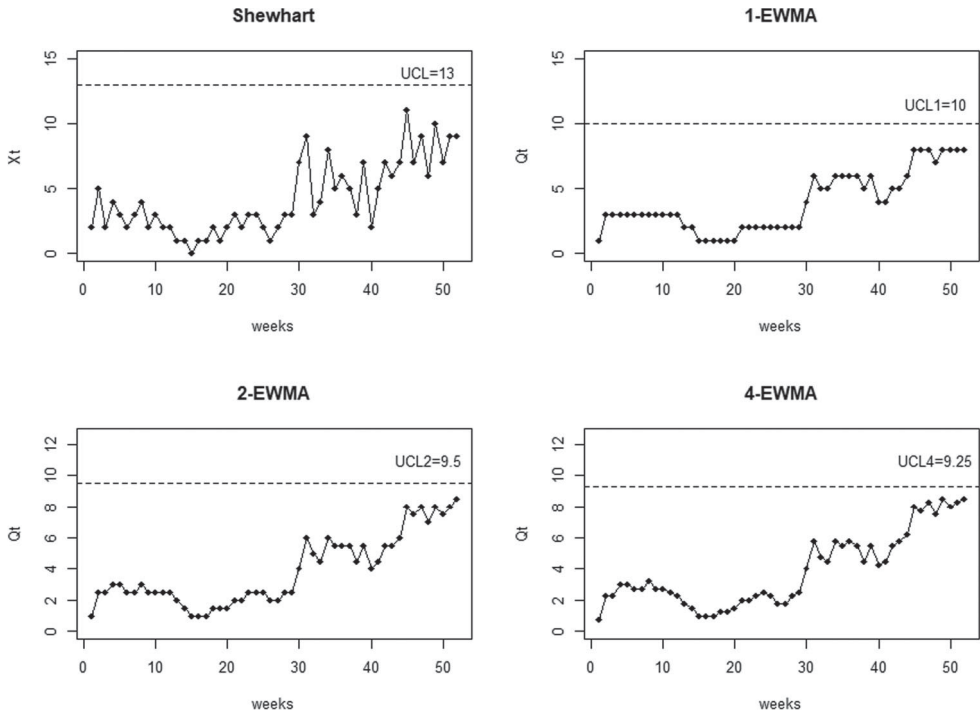


Figure 2. Upper one-sided Shewhart and s -EWMA charts, $s = 1, 2, 4$ for the Hantavirus data.

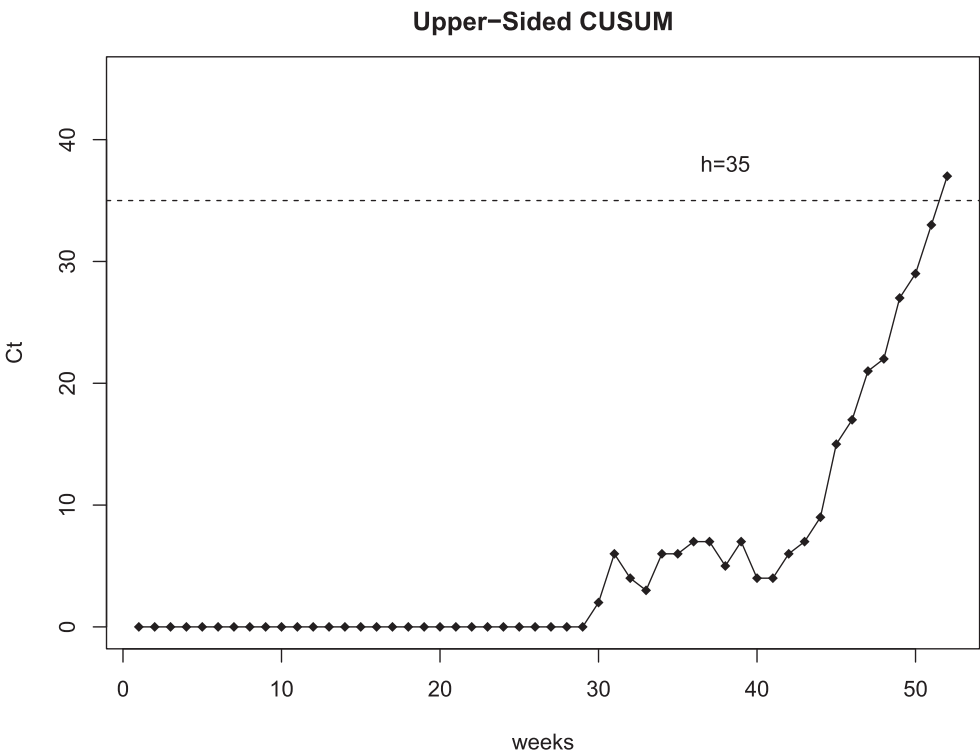


Figure 3. Upper one-sided CUSUM chart for the Hantavirus data.

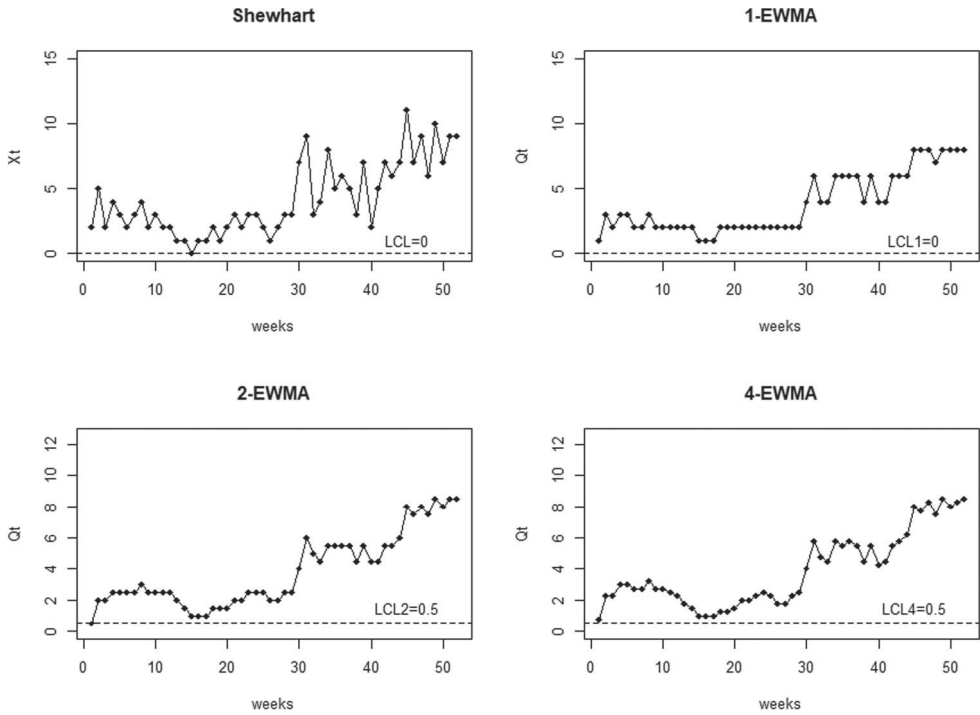


Figure 4. Lower one-sided Shewhart and s -EWMA charts, $s = 1, 2, 4$ for the Hantavirus data.

Lower-Sided CUSUM

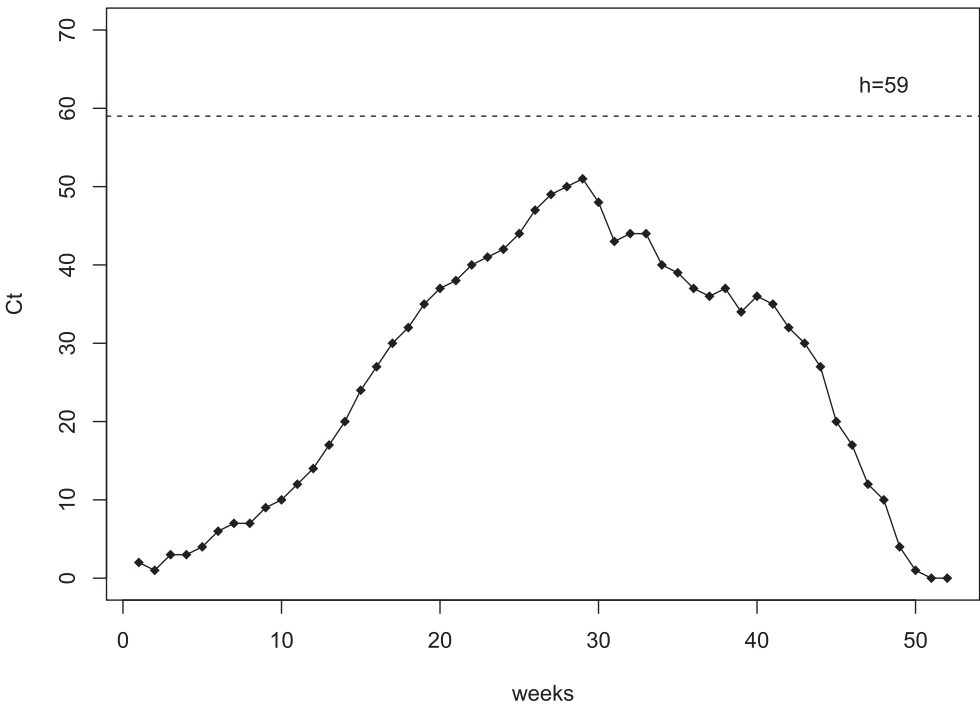


Figure 5. Lower one-sided CUSUM chart for the Hantavirus data.

We observe that in the last weeks of the year, there is a strong upward movement which begins shortly after the 30th week, which is followed by a clear indication of an upward shift from the 41st week onwards. This shift is perceived by all schemes; the CUSUM gives an OoC signal at the 52nd week.

We continue now with the lower one-sided case, where we consider the lower one-sided BBAR(1) Shewhart chart with $LCL = 0$ and $zsARL = 44.28$, the lower one-sided CUSUM chart with $k^- = 4$, $h^- = 59$ and $zsARL = 366.34$, the 1-EWMA chart with $(\lambda_1, LCL_1) = (0.5, 0)$ and $zsARL = 173.58$, the 2-EWMA chart with $(\lambda_2, LCL_2) = (0.35, 1)$ and $zsARL = 375.62$, and the 4-EWMA chart with $(\lambda_4, LCL_4) = (0.36, 2)$ and $zsARL = 366.28$. It should be mentioned that both the lower one-sided Shewhart and 1-EWMA charts have IC performance that it is not very close to the desired one. The lower one-sided charts for the Hantavirus data are given in Figures 4 and 5.

We notice that the lower one-sided Shewhart chart gives an OoC signal ($X_{15} = 0$), but it should be interpreted with caution because of its low IC $zsARL$ value. In cases like this, i.e. when a chart has a low IC $zsARL$ value, we suggest to use a more efficient scheme in order to detect decreases in the mean level of the process. On the contrary the lower one-sided CUSUM, 2-EWMA and 4-EWMA control charts give IC ARL_0 close to the desired value. Also we observe that in the first 29 weeks of the year there is not a clear indication of regional spread of hantavirus infection; the range of values is between 0 and 5 during this period. The low counts during this period cause a disorder that it is captured by all schemes.

6. Conclusion and discussion

In this work, we developed and studied one-sided s -EWMA control charts that are suitable for the detection of upward and downward shifts in the mean level of Binomial AR(1) and Beta-Binomial AR(1) processes. Both models are used frequently in practice in order to describe correlated data with a finite range. The BAR(1) model is a good choice to model count data processes with a first-order AR dependence structure and a binomial marginal distribution. However, when there are indications of extra-binomial variation and/or inhomogeneity among the n items of the sample, the BBAR(1) is a more appropriate choice. The results of an extensive numerical study regarding the statistical design and the performance of the proposed s -EWMA charts revealed that between the Shewhart and the s -EWMA charts, the latter perform better than the former; under certain circumstances, their performance can be comparable to the performance of CUSUM-type charts.

Moreover, the numerical study revealed that the extra-binomial variation affects the detection ability of the s -EWMA chart, as well as its IC performance. For $\phi_0 > 0$, the control limit obtained under the BAR(1) model needs a proper adjustment, in order to keep the FAR at the desired level. Also, at a given shift δ , the $ssARL$ is larger under the BBAR(1) model than the corresponding value under the BAR(1) model, for control charts with comparable IC performance.

Finally, the practical application of the proposed schemes was illustrated via a real-data example. For all calculations, the R statistical software R Core Team [14] was used and the programs are available from the authors upon request.

Topics for future research consist of the development and study of other types of control charts, such as the mixture cumulative count [9], in order to compare them with

the existing ones. Also, an extension to this study is the development of simultaneous control charting procedures that can provide, except for a quick and accurate detection ability, additional information about which of the process parameters has changed; one of the main assumptions is that the presence of assignable causes affects only the mean level of the process. Finally, as already said, the discrete nature of the considered processes does not allow to develop an s -EWMA chart with the desired IC performance. In order to overcome this limitation, the methodology developed by Paulino *et al.* [12], based on the theory of uniformly most powerful unbiased tests (UMPU) and the use of randomization probabilities, is expected to provide designs for all the charts under comparison, with an in-control ARL value equal to a pre-specified one. However, even though the case of one-sided Shewhart chart is feasible, some technical difficulties need to be tackled properly for the one-sided CUSUM and EWMA charts. Certainly, this is a topic worthy of future research.

Acknowledgments

The authors would like to express their gratitude to the Associate Editor and the two anonymous reviewers for their constructive comments which improved the content and the presentation of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- [1] M.A. Al-Osh and A.A. Alzaid, *First-order integer-valued autoregressive (INAR(1)) process*, J. Time Ser. Anal. 8 (1987), pp. 261–275.
- [2] D. Brook and D.A. Evans, *An approach to the probability distribution of CUSUM run length*, Biometrika 59 (1972), pp. 539–549.
- [3] P. Castagliola, K.P. Tran, G. Celano, A.C. Rakitzis, and P.E. Maravelakis, *An EWMA-type sign chart with exact run length properties*, J. Quality Technol. 51 (2019), pp. 51–63.
- [4] T.C. Chang and F.F. Gan, *Cumulative sum charts for high yield processes*, Stat. Sin. 11 (2001), pp. 791–805.
- [5] F.F. Gan, *Monitoring observations generated from a binomial distribution using modified exponential weighted moving average control chart*, J. Stat. Comput. Simul. 37 (1990), pp. 45–60.
- [6] F.F. Gan, *An optimal design of CUSUM control charts for binomial counts*, J. Appl. Stat. 20 (1993), pp. 445–460.
- [7] S. Knoth, *The art of evaluating monitoring schemes – How to measure the performance of control charts?* In H.-J. Lenz and P.-T. Wilrich, editors, *Frontiers in Statistical Quality Control* 8, pages 74–99. Heidelberg: Physica-Verlag, 2006.
- [8] J.M. Lucas and R.B. Crosier, *Fast initial response for CUSUM quality control schemes*, Technometrics 24 (1982), pp. 199–205.
- [9] M.Y. Majeed, M. Aslam, and M. Riaz, *Mixture cumulative count control chart for mixture geometric process characteristics*, Qual. Quant. 47 (2013), pp. 2289–2307.
- [10] E. McKenzie, *Some simple models for discrete variate time series*, Water Resour. Bull. 21 (1985), pp. 645–650.
- [11] D.C. Montgomery, *Introduction to Statistical Quality Control*, 6th ed. John Wiley & Sons, Inc., New York, 2009.
- [12] S. Paulino, M.C. Morais, and S. Knoth, *On ARL-unbiased c -charts for INAR(1) Poisson counts*, Statistical Papers 60 (2019), pp. 1021–1038.

- [13] S. Psarakis and G.E.A. Papaleonida, *SPC procedures for monitoring autocorrelated processes*, Qual. Technol. Quant. Manag. 4 (2007), pp. 501–540.
- [14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. Available at <http://www.R-project.org>
- [15] A.C. Rakitzis, C.H. Weiß, and P. Castagliola, *Control charts for monitoring correlated counts with a finite range*, Appl. Stoch. Models. Bus. Ind. 33 (2017), pp. 733–749.
- [16] M.M. Ristić, C.H. Weiß, and A.D. Janjić, *A binomial integer-valued ARCH model*, Int. J. Biostat. 12 (2016)
- [17] S.W. Roberts, *Control chart tests based on geometric moving averages*, Technometrics 1 (1959), pp. 239–250.
- [18] F.W. Steutel and K. van Harn, *Discrete analogues of self-decomposability and stability*, Ann. Probab. 7 (1979), pp. 893–899.
- [19] C.H. Weiß, *EWMA monitoring of correlated processes of Poisson counts*, Qual. Technol. Quant. Manag. 6 (2009a), pp. 137–153.
- [20] C.H. Weiß, *Monitoring correlated processes with binomial marginals*, J. Appl. Stat. 36 (2009b), pp. 399–414.
- [21] C.H. Weiß, *Detecting mean increases in Poisson INAR(1) processes with EWMA control charts*, J. of Appl. Stat. 38 (2011b), pp. 383–398.
- [22] C.H. Weiß, *The Markov chain approach for performance evaluation of control charts – a tutorial*. In Samuel P. Werther, editor, *Process Control: Problems, Techniques and Applications*, pages 205–228. Nova Science Publishers, Inc., 2011a
- [23] C.H. Weiß, *SPC methods for time-dependent processes of counts—a literature review*, Cogent Mathematics 2 (2015), pp. 1111116.
- [24] C.H. Weiß and H.Y. Kim, *Parameter estimation for binomial AR(1) models with applications in finance and industry*, Stat. Papers 54 (2013), pp. 563–590.
- [25] C.H. Weiß and H.-Y. Kim, *Diagnosing and modeling extra-binomial variation for time-dependent counts*, Appl. Stoch. Models. Bus. Ind. 30 (2014), pp. 588–608.
- [26] C.H. Weiß and P.K. Pollett, *Binomial autoregressive processes with density dependent thinning*, J. Time Ser. Anal. 35 (2014), pp. 115–132.
- [27] C.H. Weiß and M.C. Testik, *Detection of abrupt changes in count data time series: Cumulative sum derivations for INARCH(1) models*, J. Quality Technol. 44 (2012), pp. 249–264.
- [28] Z. Wu, J. Jiao, and Y. Liu, *A binomial CUSUM chart for detecting large shifts in fraction nonconforming*, J. Appl. Stat. 35 (2008), pp. 1267–1276.
- [29] A.B. Yeh, R.N. Mcgrath, M.A. Sembower, and Q. Shen, *EWMA control charts for monitoring high-yield processes based on non-transformed observations*, Int. J. Product Res. 46 (2008), pp. 5679–5699.
- [30] P. Yontay, C.H. Weiß, M.C. Testik, and Z.P. Bayindir, *A two-sided cumulative sum chart for first-order integer-valued autoregressive processes of Poisson counts*, Qual. Reliab. Eng. Int. 29 (2013), pp. 33–42.