CUSUM control charts for the Coefficient of Variation with Measurement Errors

Phuong Hanh Tran ∗ Cédric Heuchenne**

∗ HEC - Management School, University of Liège, Liège, Belgium (e-mail: tran@student.uliege.be).
** HEC - Management School, University of Liège, Liège, Belgium (e-mail: C.Heuchenne@ulg.ac.be)

Abstract: Due to several applications in applied statistics, there is an increasing attention to the coefficient of variation (CV) in quality control. In this paper, we propose investigating the effect of measurement errors on the performance of one-sided cumulative sum (CUSUM) control charts monitoring the CV. According to the simulated results, the precision and accuracy errors have considerable influence on the performance of the CUSUM control charts. It is also shown that increasing the number of times measuring an item in the linear covariate error model insignificantly improves the proposed chart performance.

Keywords: SPC; Measurement Errors; CUSUM Control charts; Coefficient of Variation; Markov Chain

1. INTRODUCTION

It is true that control charts play an important role in quantity control. Since being invented by Shewhart, control charts have been widely applied in many fields of industry. Beside the advantage of being easy to design and interpret, the first versions of Shewhart charts, however, has disadvantage that they are only sensitive to the considerable process shifts. That is to say, they are inefficient or need very long time to detect small or moderate process shifts. This weakness of these charts is a motivation for many researchers to either develop new control charts or propose adaptive strategies. The advanced-type control charts can be named are the supplementary Run Rules chart, the cumulative sum (CUSUM) chart and the exponentially weighted moving average (EWMA) chart; while the adaptive charts include the variable sample size (VSS) chart and variable sampling interval (VSI) chart. Among these new approaches, CUMSUM-type control charts are very good alternatives for the Shewhart chart: they are proven to be highly sensitive in detecting small or moderate shifts (Montgomery (2013)).

The earliest applications of using control charts are for monitoring the mean and the variance/standard deviation of a process. These characteristics, however, do not need to be constant or to be independent from others. In fact, the variance of many processes is a function of its mean while the mean itself varies from time to time. In such processes, coefficient of variation (CV) should be monitored instead of other characters like mean or variance. Examples of using CV charts can be seen in various fields such as textile industry, chemical and biological quality control, see Castagliola et al. (2011). These widely applications has attracted many researches on monitoring CV, see, for instance, Calzada and Scarlano (2013), Castagliola et al. Amdouni et al. (2015); Castagliola et al. (2013a,b, 2015a, 2011, 2015b) and Tran and Tran (2016) for more detail.

In most of control charts monitoring the CV cited above, it is important to consider that no measurement error in measuring quality characteristic was assumed. This assumption, however, may not be true in practice because in many industrial situations, there often exist measurement errors affecting significantly the performance of control charts. Ignoring the presence of measurement error may result in the misunderstanding about the statistical properties of control charts. In the literature on control charts, the effect of measurement errors has been already considered by a number of authors, for instance Linna et al. (2001); Costa and Castagliola (2011); Maravelakis (2012); Hu et al. (2015); Noorossana and Zereshkaz (2015), Tran et al. Tran et al. (2016a,b). Nevertheless, there is still very few papers studying the measurement errors impact on the efficiency of control charts monitoring the CV. The first study is perhaps conducted by Yeong et al. (2017) using the Shewhart CV chart. In this study, the authors used an assumption that the ratios \( \sigma_M/\sigma \) and \( A/\mu \) (described in Section 2) does not change from in-control process to out-of-control process. However, this assumption can hardly be implemented in practice. As a consequence, unreasonable results can be seen in Yeong et al. (2017) as the performance of the upward Shewhart CV control chart increases proportionally to the increase of measurement error (see Table 3 in Yeong et al. (2017)). Very recently, Tran et al. (2018) took into account the changes of these ratios when studied the performance of two CV control charts considering the occurrence of measurement errors.

The purpose of this study is to investigate the performance of one-sided CUSUM control charts monitoring CV introduced by Tran and Tran (2016) under an assumption of existing measurement errors. It is desirable to design a CUSUM-type chart for monitoring the CV because this
kind of chart generally leads to more efficient statistical
derformance than others, see Tran and Tran (2016) and
Hawkins and Wu (2014). The linear covariate error model
similar to that in Tran et al. (2018) is also applied.

The rest of the paper is organized as follows: in Section
2, we introduce the linear covariate error model for the
coefficient of variation as well as the distribution of new co-
efficient of variation in the presence of measurement errors.
The formulae of control limits and the implementation of
the two one-sided CUSUM control charts are presented
in Section 3. Section 4 is devoted to analyzing the effect
of measurement errors on the charts performance. Some
suggestions and remarks are given to conclude in Section
5.

2. LINEAR COVARIATE ERROR MODEL FOR THE
COEFFICIENT OF VARIATION

We briefly recall in this Section the linear covariate error
model for the CV suggested in Tran et al. (2018). Assume
that one wants to measure the variable of interest \( X \). A
sample of size \( n \), \( \{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\} \), is taken, in which \( i \)
stands for consecutive times of measuring, \( i = 1, 2, \ldots; X_{i,j} \)
are supposed to be independent identically distributed
\((i.i.d)\) from normal distribution, namely \( X_{i,j} \sim N(\mu_0 + 
a \sigma_0, b \sigma_0) \). The parameters \( a \) and \( b \) represent the
mean shift and standard deviation shift of the process. If \( a \neq 0 \)
or \( b \neq 1 \), the process has been shifted; on the contrary,
the process is in-control. As mentioned before, a problem
facing in measuring quality characteristic of interest
is measurement error since one cannot measure exactly true
values \( \{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\} \) of \( X \) but approximated obser-
vation values. A classical way to deal with the problem is
to repeat measuring the same item for a number of times
and treat the mean of these repeatedly measured values as
the best approximation for exact value. That means the
observation now is of the form \( \{X_{i,j,1}, X_{i,j,2}, \ldots, X_{i,j,m}\}, \)
m \( \geq 1 \), where \( X_{i,j,k} \) is the \( k \)th measurement of the item \( j \)
the \( i \) sampling; the symbol **"** is to imply the actually
observed values. Limna and Woodall (2001) proposed to
use a linearly covariate error model by the the form
\[ X_{i,j,k} = A + B X_{i,j} + \epsilon_{i,j,k}. \]

In this model, \( A \) and \( B \) are constants; \( \epsilon_{i,j,k} \) is a normal
\((0, \sigma_M)\) random error representing the measurement in-
accuracy, which is independent of \( X_{i,j} \). The constants \( A \)
and \( B \) are well-known estimated for linear covariate error
model from phase I data.

Let \( \bar{X}_{i,j} \) be the mean of \( m \) observed quantities of the same
item \( j \) at the \( i \)th sampling, then
\[ \bar{X}_{i,j} = \frac{1}{m} \sum_{k=1}^{m} X_{i,j,k} = A + B X_{i,j} + \frac{1}{m} \sum_{k=1}^{m} \epsilon_{i,j,k}. \]

The distribution of \( \bar{X}_{i,j} \) can be obtained according to
properties of normal distribution
\[ \bar{X}_{i,j} \sim N(\mu, \sigma^2) = N(A + B(\mu_0 + a \sigma_0), B^2 a^2 \sigma_0^2 + \frac{\sigma_M^2}{m}). \]
The coefficient of variation of the measured quantity \( \bar{X}_{i,j} \)
is therefore
\[ \gamma^* = \frac{\sigma^*}{\mu^*} = \frac{\sqrt{B^2 a^2 \sigma_0^2 + \frac{\sigma_M^2}{m}}}{A + B(\mu_0 + a \sigma_0)} = \frac{\sqrt{B^2 a^2 + \frac{\sigma^2}{m}}}{\theta + B(1 + a^2)} \times \gamma_0. \]

where \( \eta = \frac{\sigma_\epsilon}{\sigma_0}, \gamma_0 = \frac{\sigma_\epsilon}{\mu_0}, \theta = \frac{\mu_0}{\mu_0} \). By these denoting, \( \gamma_0 \) is in-control
value of CV, \( \eta \) is the precision error ratio and \( \theta \) is
accuracy error. The sample coefficient of variation \( \gamma_i \)
is defined as
\[ \gamma_i = \frac{S_i}{\bar{X}_i} \]
in which
\[ \bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{i,j} \text{ and } S_i^* = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (X_{i,j} - \bar{X}_i)^2} \]
are the sample mean and the sample standard deviation
of \( X_{i,j} \). In the case of out-of-control condition,
the CV is defined by \( \gamma_i = \frac{b \sigma_0}{\mu_0} \). Let \( \tau \) denote
the shift size, i.e. \( \gamma_i = \tau \gamma_0 \), then it is easy to show that
\( b/\tau = 1 + a \gamma_0 \). By substituting \( 1 + a \gamma_0 \) with \( b/\tau \) in \( \gamma_i \),
the out-of-control CV \( \gamma_i \) can be rewritten as
\[ \gamma_i = \frac{\sigma^*}{\mu^*} = \frac{\sqrt{B^2 b^2 + \frac{\sigma^2}{m}}}{\theta + \frac{b}{\tau}} \times \gamma_0. \]

It is important to consider that the relation between the
out-of-control value \( \gamma_i \) with the in-control value \( \gamma_0 \) of the
CV in \( \gamma_i \) is different from that between the relation between
the out-of-control value \( \gamma_i \) with the in-control value \( \gamma_0 \) in
Yeong et al. (2017). It does not need to assume in \( \gamma_i \) that
the ratios \( \eta = \frac{\sigma_M}{\sigma_0} \) and \( \theta = \frac{A}{\mu} \) are constant; they are
free to change from in-control to out-of-control conditions.

The distribution of the sample coefficient of variation of
normal variables has been studied by many authors, for
example Iglewicz and Myers (1970); Reh and Scheffler
(1996); Vangel (1996). In this paper, we adopt the ap-
proximation formula for the cumulative distribution func-
tion \((c.d.f.)\) of CV suggested by Castagliola et al. (2011).
Changing the role of \( X_i \) in Castagliola et al. (2011) by \( \bar{X}_{i,j} \)
results in the following approximation of the \( c.d.f. \) and the
inverse distribution function \((i.d.f.)\) of \( \gamma^* \).

\[ F_{\gamma^*}(x|n, \gamma^*) = 1 - F_i \left( \sqrt[n]{x}, n - 1, \sqrt[n]{\gamma^*} \right) \]
and
\[ F_{\gamma^*}^{-1}(p|n, \gamma^*) = F_i^{-1} \left( 1 - p, n - 1, \sqrt[n]{\gamma^*} \right) \]
where \( F_i(.) \) is the \( c.d.f. \) and the \( i.d.f. \) of the noncentral
T distribution. Castagliola et al. (2011) also proposed to
monitor the coefficient of variation square using CUSUM
because it is more efficient than monitoring the coefficient
of variation itself. The \( c.d.f. \) and \( i.d.f. \) of \( \gamma^2 \) is therefore
needed. It is shown by Castagliola et al. (2011) that

\[ F_{\gamma^2}(x|n, \gamma^*) = 1 - F_F \left( \sqrt[n]{x}, n - 1, 1, \frac{n}{\gamma^2} \right) \]
and
\[ F_{\gamma^2}^{-1}(p|n, \gamma^*) = F_F^{-1} \left( 1 - p, 1, n - 1, \frac{n}{\gamma^2} \right) \]
where \( F_F(.) \) is the \( c.d.f. \) and \( i.d.f. \) of the noncentral \( F \)
distribution; \( \gamma^2 \) is computed from \( \gamma^* \).
3. IMPLEMENTATION OF THE CUSUM-$\gamma^2$ CONTROL CHARTS WITH MEASUREMENT ERRORS

For a number of reasons, instead of investigating one two-sided chart or monitoring coefficient of variation itself, the two separate one-sided CUSUM control charts for monitoring coefficient of variation squared are explored in this paper. More detail of these reasons are discussed in Castagliola et al. (2011); Tran and Tran (2016). Denote $\mu_0(\gamma^2)$ and $\sigma_0(\gamma^2)$ the mean and standard deviation of the sample $\gamma^2$ when the process is in-control. The two one-sided CUSUM-CV charts in the presence of measurement errors are defined as follows.

- An upward CUSUM chart (denoted by “upward CUSUM-$\gamma^2$”) for detecting an increase in coefficient of variation,

$$C_i^+ = \max\{0,C_{i-1}^+ + (\hat{\gamma}_i^2 - \mu_0(\gamma^2) - K^+)\} \quad (8)$$

where $K^+ = k^+\sigma_0(\gamma^2)$, the initial values $C_0^+ = 0$ and corresponding upper control limit $H_{CUSUM-\gamma^2}^+ = h^+\mu_0(\gamma^2) > 0$.

- A downward CUSUM chart (denoted as “downward CUSUM-$\gamma^2$”) for detecting a decrease in coefficient of variation,

$$C_i^- = \max\{0,C_{i-1}^- - (\hat{\gamma}_i^2 - \mu_0(\gamma^2) - K^-)\} \quad (9)$$

where $K^- = k^-\sigma_0(\gamma^2)$, the initial values $C_0^- = 0$ and corresponding lower control limit $H_{CUSUM-\gamma^2}^- = h^-\mu_0(\gamma^2) > 0$. The parameter couples $(k^+,h^+)$ and $(k^-,h^-)$ play the role of the upward and downward CUSUM-$\gamma^2$ chart coefficients, respectively.

According to the above designing, the calculation of $\mu_0(\gamma^2)$ and $\sigma_0(\gamma^2)$ is needed. However, there is no closed form for both of them. We then resort to the following accurate approximations provided by Breunig (2001):

$$\mu_0(\gamma^2) = \gamma_0^2 \left(1 - \frac{3\gamma_0^2}{n}\right), \quad \text{and} \quad (10)$$

$$\sigma_0(\gamma^2) = \sqrt{\gamma_0^4 \left(\frac{1}{n} + \gamma_0^2 \left(\frac{1}{n} + \frac{\gamma_0^2}{n}\right)\right) - (\mu_0(\gamma^2) - \gamma^2)} \quad (11)$$

The in-control value $\gamma_0^2$ in (10) and (11) is computed from (1) with $a = 0$ and $b = 1$. After calculating the control chart parameters $(k^+,h^+)$ and $(k^-,h^-)$, the CUSUM-$\gamma^2$ charts are defined and the next step is to evaluate the performance of the charts via the measure of zero-state ARL from an in-control value $\gamma_0^2$ to out-of-control value $\gamma_1^2$ for specific shifts $a$ and $b$. The investigated statistical measure of the performance is the zero-state ARL (Average Run Length), defined as the average number of samples before a control chart signals an “out-of-control” condition or issues a false alarm. In this paper, ARL is calculated by using a Markov-chain approximation (Brook and Evans (1972)). We firstly divide the control interval of upward (downward) chart into $N$ sub-intervals in which the first sub-interval is $\delta = \frac{H^+}{2N}$ (or $\delta = \frac{H^-}{2N}$) in width and the others are $2\delta$ in width. Figure 1 demonstrates this subdivision with upward chart.

Fig. 1. Control limit interval of upward chart divided into $N - 1$ subintervals of width $2\delta$ and first interval of with $\delta$.

In this figure, each sub-interval $(H_j - \delta, H_j + \delta)$ represents a transient state of Markov chain, where $H_j$ is the midpoint of sub-interval $j, j = 0, \ldots, N - 1$; the states 0 to $N - 1$ are in-control states while the state $N$ is out-of-control. If the statistic $C_{i}^+ (C_{i}^-)$ $(H_j - \delta, H_j + \delta)$, the Markov chain is in the transient state $j$ for sample $i$; if not, the chain reaches absorbing state. $N$ is chosen sufficiently large so that $H_j$ can be considered as an approximately representative value of the state $j$ ($N$ is set to be 200 in this paper). In this subdivision, the zero state has half size of the others, leading to better Markov chain approximation with the same width of sub-intervals as in Castagliola et al. (2011).

The transition probability matrix $P$ of the discrete-time Markov chain is

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,N-1} & r_0 \\ P_{1,0} & P_{1,1} & \cdots & P_{1,N-1} & r_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ P_{N-1,0} & P_{N-1,1} & \cdots & P_{N-1,N-1} & r_{N-1} \end{pmatrix},$$

where $Q$ is the $(N,N)$ matrix of transient probabilities, $0 = (0,0,\ldots,0)^T$ and $r$ is N-vector satisfying $r = (1 - Q1)$ (i.e., row probabilities sum to 1) with 1 $(1,1,\ldots,1)^T$. Secondly, we calculate the elements $Q_{i,j}$ of the matrix $Q$ by the following formulations:

- for the upward chart,

$$Q_{i,0} = F_{\gamma^2} \left(\mu_0(\gamma^2) - H_i + K^+ + \delta| n, \gamma_1^i \right),$$

$$Q_{i,j} = F_{\gamma^2} \left(\mu_0(\gamma^2) + H_j - H_i + \delta + K^+| n, \gamma_1^i \right) - F_{\gamma^2} \left(\mu_0(\gamma^2) + H_j - H_i - \delta + K^+| n, \gamma_1^i \right).$$

- for the downward chart,

$$Q_{i,0} = 1 - F_{\gamma^2} \left(\mu_0(\gamma^2) + H_i - K^- - \delta| n, \gamma_1^i \right),$$

$$Q_{i,j} = F_{\gamma^2} \left(\mu_0(\gamma^2) + H_j - H_i - \delta + K^-| n, \gamma_1^i \right) - F_{\gamma^2} \left(\mu_0(\gamma^2) + H_j - H_i + \delta + K^-| n, \gamma_1^i \right),$$

where $F_{\gamma^2} \left(\cdot\right)$ is the c.d.f. of $\gamma^2$ in (6).
Next, the two one-sided CUSUM-γ² control charts can be optimally designed in terms of ARL. This procedure, however, is only performed for a specific shift size. In practice, it seems to be rather difficult for quality practitioner to predict the shift size τ because (1) without related historical data, they have no information about the entity of next shift size, and (2) the shift size is unstable: it varies from stochastic model to stochastic model. An alternative method to overcome this situation is to evaluate the statistical performance of the chart through EARL (expected average run length)

\[ \text{EARL} = \int_{\Omega} ARL \times f_\tau(\tau) \text{d}\tau, \]

where Ω is support of shift size τ and \( f_\tau(\tau) \) is density function of τ. The design procedure of our charts now is implemented by finding out the couple \((k^+, h^+)\) and \((k^−, h^−)\) that minimize the out-of-control ARL for a given in-control ARL₀. That is to say, we will look for the couples \((k^+, h^+)\) and \((k^−, h^−)\) satisfying:

- for downward chart,

\[
\begin{cases}
(k^{*−}, h^{*−}) = \arg\min_{(k^−, h^−)} \text{EARL}(n, m, B, \eta, \gamma_0^2, \tau, k^−, h^−), \\
\text{ARL}(n, m, B, \eta, \gamma_0^2, \tau = 1, k^−, h^−) = ARL₀;
\end{cases}
\]

- for upward chart,

\[
\begin{cases}
(k^{*+}, h^{*+}) = \arg\min_{(k^+, h^+)} \text{EARL}(n, m, B, \eta, \gamma_0^2, \tau, k^+, h^+), \\
\text{ARL}(n, m, B, \eta, \gamma_0^2, \tau = 1, k^+, h^+) = ARL₀;
\end{cases}
\]

In the case there is no information about τ, one can choose an uniform distribution for τ on fixed support \( \Omega = [a_0, b_0] \), i.e. \( f_\tau(\tau) = \frac{1}{b_0-a_0} \) over the possibly guessed interval \([a_0, b_0]\) of τ. In the following section, we consider two different ranges of shifts: \( \Omega_D = [0.5, 1] \) and \( \Omega_F = (1, 2] \), corresponding to decreasing and increasing shift sizes, respectively. It has the advantage that one does not need to guess the true value for τ.

4. THE EFFECT OF MEASUREMENT ERRORS ON THE CUSUM-γ² CONTROL CHARTS

We explore now the performance of two proposed one-sided CUSUM-γ² control charts in the presence of measurement errors. The value for the values of in-control ARL (ARL₀) is set at 370.4. Without loss of generality, the shift of variance of true value \( X_{i,j} \) is assumed to be unit in the remaining of this section, i.e. \( b = 1 \). Given the values of \( m, n, a, B, \eta, \theta \) and \( \gamma_0 \), the optimal couple \((k^{*+}, h^{*+})\) in equation (14) for downward chart and \((k^{*−}, h^{*−})\) in equation (15) for upward chart are found by using simultaneously a non-linear equation solver joint to an optimization algorithm developed in Scicoslab software.

More specifically, for fixed values of \( m, n, a, B, \eta, \theta \) and \( \gamma_0 \) we explore the optimal combinations \((k^+, h^+)\) or \((k^−, h^−)\) such that \( \text{ARL}(n, m, B, \eta, \gamma_0^2, \tau = 1, k^+, h^+) = ARL₀ \) or \( \text{ARL}(n, m, B, \eta, \gamma_0^2, \tau = 1, k^−, h^−) = ARL₀ \) (using the non-linear equation solver) minimizing EARL\((k^+, h^+)\) or EARL\((k^−, h^−)\) (using the optimizer). The effect of measurement errors on the overall performance of the CUSUM-γ² control charts are left to the Tables 1-4. A number of following conclusions can be drawn from these obtained results.
Tables 1–4 present the performance of the two proposed charts based on the global variation of the shift size $\tau$. The effect of measurement errors in this case, measured by $EARL$, are in general consistent with those discussed in the previous items. The values of $EARL$ sharply increase as $\theta$ increases (fixed others). For example, $EARL = 18.4$ for $\theta = 0$ while $EARL = 19.3$ for $\theta = 0.05$ ($n = 10, \gamma_0 = 0.2, B = 1, m = 1$) (Table 2). This tendency is also true for the impact of $\eta$ (Table 1) on $EARL$, but its effects are slight less than those of $\theta$. The contribution of $B$ (Table 3) and $m$ (Table 4) to the performance of charts is the opposite of the other ones: the increase of $B$ leads to better performance of the charts. From these tables, it is also considered that the values of $EARL$ corresponding to the upward chart are always smaller than those corresponding to the downward chart, no matter the value of parameters is. This can be explained by the skewed distribution of $CV$.

5. CONCLUDING REMARKS

We have investigated in this paper the effects of measurement errors on the performance of the CUSUM-$\gamma^2$ control charts. We have found that the presence of measurement errors obviously have influence on the effectiveness of the proposed charts. That is shown by the variation of $EARL$ following the fluctuating of parameters of measurement error model. More specifically, the larger the value of precision error or accuracy error is, the slower the CUSUM-$\gamma^2$ chart is in detecting the out-of-control process state. Among them, the effect of accuracy error is much stronger than that of precision error. On the contrary, the increase of slope coefficient under linear covariate error model is not weakened the performance of the charts; it even improve slightly the effectiveness on the charts. Moreover, it turns out that the efficiency of increase the number of multiple measurement per item is not really impressed for CUSUM-$\gamma^2$ chart. The $EARL$ was insignificantly reduced as we heightened the value of multiple measurement per item. This might be useful information for quantity practitioners to develop suitable strategies for improving the performance of CUSUM-$\gamma^2$ charts: They should focus more on parameters like the sample size rather than repeatedly measuring an item.

REFERENCES


