

Nonconforming Control Charts for Zero-Inflated Poisson distribution

N. Katemee, T. Mayureesawan

Abstract—This paper developed the c -Chart based on a Zero-Inflated Poisson (ZIP) processes that approximated by a geometric distribution with parameter p . The p estimated that fit for ZIP distribution used in calculated the mean, median, and variance of geometric distribution for constructed the c -Chart by three difference methods. For c_g -Chart, developed c -Chart by used the mean and variance of the geometric distribution constructed control limits. For c_{mg} -Chart, the mean used for constructed the control limits. The c_{me} -Chart, developed control limits of c -Chart from median and variance values of geometric distribution. The performance of charts considered from the Average Run Length and Average Coverage Probability. We found that for an in-control process, the c_g -Chart is superior for low level of mean at all level of proportion zero. For an out-of-control process, the c_{mg} -Chart and c_{me} -Chart are the best for mean = 2, 3 and 4 at all level of parameter.

Keywords—average coverage probability, average run length, geometric distribution, zero-inflated poisson distribution

I. INTRODUCTION

THE classical Shewhart control chart of nonconformities (c -Chart) is used to monitor the number of nonconformities per unit of product base on Poisson distribution when the sample sizes are constant. In situation, the sampling of products takes place as a set of repeated samplings, with each sampling finding either zero nonconformities or a nonzero number of nonconformities. If in some sampling processes an excess number of zeros might be observed, then distribution is called a “Zero-Inflated Poisson (ZIP)”. In this case, the estimation of the sample mean often underestimates of the Poisson distribution. When the estimated variance is greater than the mean, this is so called “Over Dispersion”, and the estimated limits in the c -Chart are improperly narrow [7]. Cohen [1] studied the maximum likelihood estimator (MLE) $\hat{\lambda}$ of a ZIP model because he found that the estimated value for the mean by using the MLE value is closer to the actual value. This study has been used in many applications [3], [8]. Xie et al. [7] developed a c -Chart for the ZIP model that they called the c_{ZIP} -Chart. They studied the efficiency of the c_{ZIP} -Chart for monitoring upward shifts of the mean value of number of nonconformities in a process. Sim and Lim [2] illustrate a charting method called a c_J -Chart in which they used a one-sided Jeffreys prior interval [9] to detect upward shifts. They compared this c_J -Chart with a usual c -Chart and a c_{ZIP} -Chart.

N. Katemee is with the Applied Statistics Department, Applied Science Faculty, King Mongkut's University of Technology North Bangkok, 10800, Thailand (phone: 668-1-707-1513 ; e-mail: narun_ess@hotmail.com).

T. Mayureesawan is with the Applied Statistics Department, Applied Science Faculty, King Mongkut's University of Technology North Bangkok, 10800, Thailand (e-mail: tidadeaw@yahoo.com).

They showed that the c_J -Chart was appropriate for processes when the mean was in-control. On the other hand, if the process mean was in an out-of-control situation, then the c -Chart performed better than the other charts. However, they found that the c -Chart yields poor coverage probability. Peerajit and Mayureesawan [10] extended the research ideas of [2] by supplying both the proportion of zero nonconformities and the mean shift in a production process. The results obtained for the performance of the c -Chart, c_{ZIP} -Chart and c_J -Chart were in agreement with those of Sim and Lim.

The results of the research mentioned, the authors have either studied the performance of the charts for either the average run length (ARL) or the average coverage probability (ACP) but not both, or if they studied both they found that charts might perform well for one measure but poorly for the other measure.

The aims of the present study are to develop modified the control limits of a c -Chart for the ZIP distribution that perform satisfactorily for a range of parameters of the ZIP distribution and to compare the performance of these charts with the charts mentioned. The outline of the paper is as follows. We first develop an approximation for the distribution of the ZIP distribution as a geometric distribution with parameter $p(p_g)$, and we examine how the value of p_g varies as the parameter of the ZIP distribution is changed. We then use the p_g estimated for calculated the mean, median, and variance of geometric distribution for constructed the c -Chart by three different procedures to develop three control charts, which we call c_g -Chart, c_{mg} -Chart and c_{me} -Chart. The performance of these developed control charts is then compared with the performance of c -Chart, c_{ZIP} -Chart and c_J -Chart.

II. MATERIALS AND METHODS

Zero-Inflated Poisson (ZIP)

The probability mass function is given by [8]:

$$P(Y=y) = \begin{cases} \omega + (1-\omega) \exp -\lambda & , (y=0) \\ \frac{(1-\omega) \exp -\lambda \lambda^y}{y!} & , (y>0), \end{cases} \quad (1)$$

where Y = the random variables of nonconformities in a sample unit,

= the mean of nonconformities in a sample unit,

= is a measure of the extra proportion of zero

nonconformity in a sample unit, and

$$E(Y) = (1-\omega) \mu \text{ and } V(Y) = \mu + \left(\frac{\omega}{1-\omega} \right) \mu^2. \quad (2)$$

Note: $\omega = 0$ is the Poisson distribution.

The Geometric distribution

The probability function is given by [6]:

$$P(Y = k) = (1-p)^k p, \quad k = 0, 1, 2, \dots, \quad (3)$$

where Y = the random variables of the number of failures until the first success to occur,

$p = p_g$ = the probability of success on each trial, and

$$E(Y) = \frac{1-p_g}{p_g} \text{ and } V(Y) = \frac{1-p_g}{p_g^2}. \quad (4)$$

The Shewhart control chart of nonconformities (c -Chart)

The control limits are given by [4]:

$$\begin{aligned} UCL &= c + 3\sqrt{c} \\ CL &= c \\ LCL &= c - 3\sqrt{c} \end{aligned} \quad (5)$$

c is assumed to be the mean number of nonconformities if the mean of the probability distribution is known, otherwise c is estimated as the mean of the number of nonconformities in a sample of observed product units (\bar{c}).

The control chart of nonconformities with ZIP model (c_{ZIP} -Chart)

In 1991, Cohen [1] developed a ZIP model for a Poisson probability function $g(y, \lambda)$, $y = 0, 1, 2, \dots$ given by:

$$P(Y = y) = \omega I_{(y,0)} + (1-\omega)g(y, \lambda), \quad y = 0, 1, 2, \dots \quad (6)$$

where Y = the random variables of nonconformities in a product process, and $I_{(y,0)} = 1$ if $y = 0$ and $I_{(y,0)} = 0$ if $y \neq 0$.

The maximum likelihood estimate (MLE) of parameter λ in the ZIP model of Cohen is given by:

$$\hat{\lambda} = \bar{y}^+ \left[1 - e^{-\hat{\lambda}} \right], \quad (7)$$

where \bar{y}^+ = the mean of the number of nonconformities in product units that have a nonzero number of nonconformities.

The $\hat{\lambda}$ are then used in the control limits for the c_{ZIP} -Chart [7] as follows:

$$\begin{aligned} UCL &= \hat{\lambda} + 3\sqrt{\hat{\lambda}} \\ CL &= \hat{\lambda} \\ LCL &= \hat{\lambda} - 3\sqrt{\hat{\lambda}} \end{aligned} \quad (8)$$

The control chart of nonconformities with Jeffreys Prior Interval method (c_J -Chart)

The one-sided Jeffreys prior interval is given by [9]:

$$CI_J^\lambda(y) = [G(\alpha; y + 0.5, 1), \infty], \quad (9)$$

where y = the number of nonconformities for a Poisson distribution,

λ = the parameter estimate $\hat{\lambda}$ for the ZIP model from (7).

If $y = 0$ the confidence interval is $[0, \infty)$, if $y \neq 0$ the confidence interval is $[G(\alpha; y + 0.5, 1), \infty)$, where $G(\alpha; a, b)$ is the 100 α th percentile of a Gamma distribution with shape parameter $a = y + 0.5$ and scale parameter $b = 1$,

For the control chart of nonconformities with Jeffreys Prior Interval method, i.e., the c_J -Chart, the control limit is given by [2]:

$$UCL = \max[y \mid \lambda > G(\alpha; y + 0.5, 1)]. \quad (10)$$

Development of the c -Charts for Zero-Inflated Poisson (ZIP) processes

For a given ZIP distribution and k is geometric distribution, we first obtain an approximate geometric distribution with parameter p (p_g) by using the Kolmogorov-Smirnov test [5].

The charts for a ZIP processes are then defined as follows.

1. c_g -Chart is a modified control limits of the c -Chart obtained by using the \hat{p}_g that fit for ZIP distribution for calculated the mean and variance of geometric distribution for constructed the control limits of a one-sided c -Chart. Therefore the control limit of c_g -Chart is given by:

$$UCL = E(Y) + 3\sqrt{V(Y)} \quad (11)$$

$$LCL = 0,$$

where

$$\hat{p}_g = \left(1 + \frac{1}{n} \sum_{i=1}^n k_i \right)^{-1}, \quad E(Y) = \frac{1 - \hat{p}_g}{\hat{p}_g} \text{ and } V(Y) = \frac{1 - \hat{p}_g}{\hat{p}_g^2}. \quad (12)$$

2. c_{mg} -Chart is a modified control limits of the c -Chart obtained by using the \hat{p}_g that fit for ZIP distribution for calculated the mean of geometric distribution for constructed the control limits of a one-sided c -Chart. Therefore the control limit of c_{mg} -Chart is given by:

$$UCL = E(Y) + 3\sqrt{E(Y)} \quad (13)$$

$$LCL = 0.$$

3. c_{me} -Chart is a modified control limits of the c -Chart obtained by using the \hat{p}_g that fit for ZIP distribution for calculated the median (M) and variance of geometric distribution for developed control limits of a one-sided c -Chart. Therefore the control limit of c_{me} -Chart is given by:

$$UCL = M + 3\sqrt{V(Y)} \quad (14)$$

$$LCL = 0,$$

where $Median = \left[\frac{-1}{\log_2(1 - \hat{p}_g)} \right] - 1$.

III. SIMULATION RESULTS

In this section we have shown the results of tests of the charts by a simulation study. For the simulations, we assume the following ranges of parameter values. The means for the in-control process are: $(\mu_0) = 1.0(1.0)4.0$. The means for the out-of-control process are: $(\mu_1 = \mu_0 + \rho)$ where the mean shifts are: $(\rho) = 0.00, 0.40, 0.80$ and 1.20 . The proportions of zero nonconformity are: $(\omega) = 0.30(0.10)0.90$. Finally, the value for the over-dispersion $(\phi) = 1$.

The evaluation of the performance of the control charts was conducted as follows:

1. The R program was used to simulate the number of nonconforming items for a ZIP distribution with values for the parameters (n, μ_0, ϕ, ω) chosen from the set of values given above.

- The value of the parameter p_g which gives a best fit between the ZIP distribution from step 1 and geometric distribution.
- The Kolmogorov-Smirnov test was used to test the hypothesis that a geometric distribution with the p_g value from step 2 could give a reasonable fit to the distribution of data obtained in step 1. Based on simulations with 20,000 replications, the results of the test showed that the hypothesis was satisfied for at least 95% of the replications. For the p_g fit values with a ZIP distribution, we used the number of failures until the first success to occur of a geometric distribution for calculated the \hat{p}_g from (12).
- Based on 100,000 replications, the averaged control limits were calculated for the c -Chart, c_{ZIP} -Chart and c_J -Chart. For the c_g -Chart, c_{mg} -Chart and c_{me} -Chart the values for the average \hat{p}_g in step 3 were then used for calculating the mean, median, and variance of geometric distribution for construct the control limits.
- Based on a new set of 100,000 replications, the control limits calculated in step 4 were then used to compute the ARL and the ACP for each chart.
- Steps 1 to 5 were then repeated for a new set of values for parameters (n, μ_0, ϕ, ω) .

IV. RESULTS

In this section a summary is given of some of the results that were obtained from the simulations.

Table I shows the values of \hat{p}_g for the geometric distribution that gives the best fit between the geometric and the ZIP distribution for a range of ω and μ values. It can be seen that as the values of $\mu = 1.0$, the values of \hat{p}_g for all of ω are a constant value (0.53) and when the values of $\mu = 2.0 - 4.0$, as the values of ω are increased, the values of \hat{p}_g vary depend on the μ values.

The results for the in-control case ($\rho = 0.00$) are shown in table II. Table II shows a comparison of ARL_0 and ACP values for the c -Chart (c), c_{ZIP} -Chart (c_{ZIP}), c_J -Chart (c_J), c_g -Chart (c_g), c_{mg} -Chart (c_{mg}) and c_{me} -Chart (c_{me}). A comparison of ARL_0 values for the charts is given in Fig. 1. It can be seen that when $\mu_0 = 1.0$ for all levels of ω the c_g -Chart and c_J -Chart return similar the highest ARL_0 values. Therefore, they are accepted as the preferred control chart because it detects shifts slowly. When $\mu_0 = 2.0-4.0$ and for $\omega = 0.3$ and 0.4 , the c_g -Chart returns the highest ARL_0 values. Therefore c_g -Chart is appropriate control chart. However, when $\omega = 0.5-0.9$, the c_J -Chart is appropriate control chart. Fig. 2 shows the absolute values of the differences between the ACP values and the confidence level of 0.9973, which we call the $ACP-DIFF$ value, for the preferred charts for the ARL values i.e., for the c_J -Chart and c_g -Chart. It can be seen that when $\omega = 0.3$ and 0.4 , these two charts have similar low $ACP-DIFF$ values for all values of μ_0 .

That is, these control charts all give ACP values close to the target level of 0.9973. However, for higher ω (0.5–0.9), only the c_{ZIP} -Chart and c_J -Chart give ACP values close to the required confidence level. When both ARL_0 and ACP values are considered, the c_g -Chart and c_J -Chart will be the preferred control chart when $\omega = 0.3$ and 0.4 for levels of $\mu_0 = 1.0$. However, only the c_g -Chart will be the preferred control chart for $\mu_0 = 2.0-4.0$. When all of μ_0 and for $\omega = 0.5-0.9$, only the c_J -Chart will be the preferred control chart. The process is in an out-of-control state ($\rho > 0.00$)

Results for this case are shown in Fig. 3 and 4. Fig. 3 gives a comparison of ARL_1 values for a range of values of $\mu_1 = \mu_0 + \rho$, ω and ρ . It can be seen that when $\mu_0 = 1.0$ for $\omega = 0.3-0.6$ and all of ρ , the c -Chart, c_{mg} -Chart and c_{me} -Chart, return similar low values of ARL_1 . That is, these three charts are able to detect shifts faster than the other charts. However, only the c -Chart return low values for $\omega = 0.7-0.9$. When $\mu_0 = 4.0$ (as $\mu_0 = 2.0$ and 3.0 return similar of result) for all of ω and ρ , the c -Chart, c_{mg} -Chart and c_{me} -Chart return similar low values of ARL_1 . It can be seen from Fig. 3, that all control charts detect a shift slowly for values of ω of (0.8, 0.9). Fig. 4 gives a comparison of the $ACP-DIFF$ values for the preferred charts for the ARL_1 values, that is, for the c -Chart c_{mg} -Chart and c_{me} -Chart. It can be seen that when all of ω , the c_{mg} -Chart and c_{me} -Chart returns the lowest $ACP-DIFF$ values for all values of μ_0 , ρ tested, that is, it gives the ACP value closest to the target value. When both ARL_1 and ACP values are considered, the c_{mg} -Chart and c_{me} -Chart will be the preferred control chart for $\mu_0 = 1.0, \omega = 0.3-0.6$ and all of ρ . However, no control charts are to be preferred for $\omega = 0.7-0.9$. When levels of $\mu_0 = 2.0-4.0$ for all ω and ρ , the c_{mg} -Chart and c_{me} -Chart will be the preferred control chart.

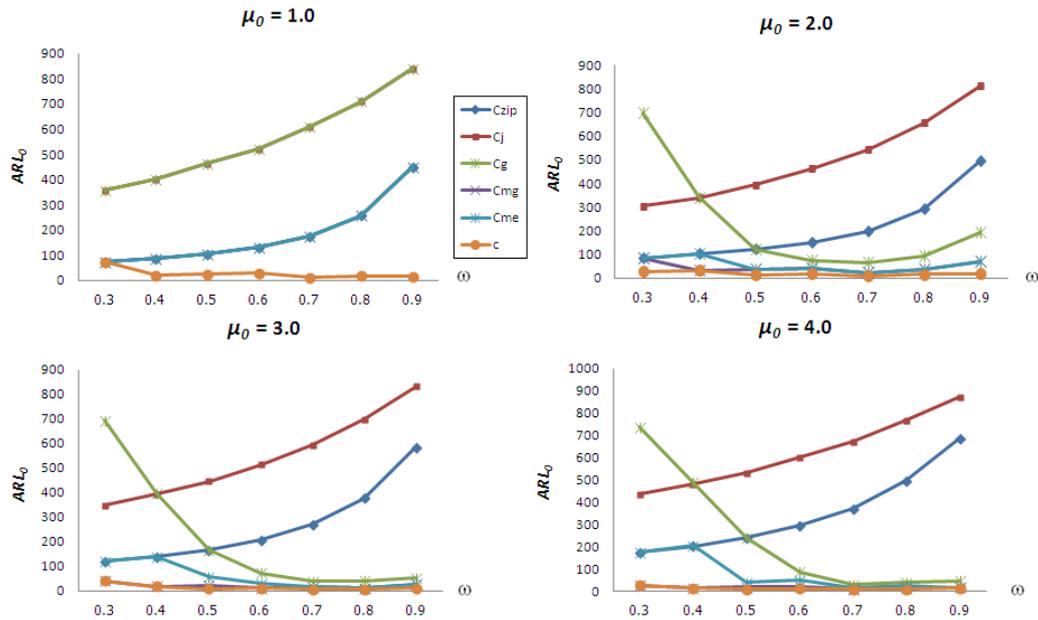
TABLE I

THE \hat{p}_g VALUES FOR THE GEOMETRIC THAT GIVE THE BEST FIT TO THE DISTRIBUTION OF THE ZIP MODEL FOR A RANGE OF μ AND ω VALUES

ω	μ			
	1.0	2.0	3.0	4.0
0.30	0.53	0.39	0.32	0.28
0.40	0.53	0.43	0.33	0.29
0.50	0.53	0.46	0.36	0.32
0.60	0.53	0.46	0.41	0.32
0.70	0.53	0.53	0.46	0.42
0.80	0.53	0.53	0.53	0.42
0.90	0.53	0.53	0.53	0.51

TABLE II
COMPARISON OF ARL_0 AND ACP VALUES OF THE C -CHART, C_{ZIP} -CHART, C_J -CHART, C_G -CHART,
 C_{MG} -CHART AND C_{ME} -CHART FOR A RANGE OF μ_0 AND ω VALUES

μ_0	ω	ARL_0						ACP					
		C	C_{ZIP}	C_J	C_G	C_{mg}	C_{me}	C	C_{ZIP}	C_J	C_G	C_{mg}	C_{me}
1.0	0.3	74.5	74.2	357.6	357.7	74.3	74.2	0.9742	0.9876	0.9978	0.9978	0.9914	0.9867
	0.4	19.8	87.1	402.2	402.3	87.2	87.1	0.9517	0.9891	0.9981	0.9981	0.9932	0.9882
	0.5	23.8	104.7	462.2	462.3	104.7	104.7	0.9618	0.9923	0.9985	0.9983	0.9944	0.9914
	0.6	30.1	130.6	521.7	521.8	130.6	130.6	0.9654	0.9954	0.9988	0.9987	0.9962	0.9932
	0.7	11.7	174.3	610.2	610.2	174.4	174.3	0.9227	0.9954	0.9987	0.9986	0.9980	0.9944
	0.8	17.9	256.8	708.5	708.6	256.8	256.8	0.9484	0.9970	0.9994	0.9993	0.9890	0.9962
	0.9	14.7	448.8	840.7	840.7	448.9	448.8	0.9515	0.9986	0.9997	0.9997	0.9690	0.9980
2.0	0.3	26.2	84.5	302.6	699.5	84.5	84.5	0.9651	0.9917	0.9977	0.9985	0.9726	0.9890
	0.4	30.8	100.3	338.7	340.9	30.2	100.3	0.9678	0.9931	0.9976	0.9970	0.9806	0.9906
	0.5	13.1	119.7	392.4	119.9	36.8	36.8	0.9307	0.9943	0.9983	0.9931	0.9557	0.9726
	0.6	16.3	150.5	463.0	72.7	40.7	40.9	0.9442	0.9953	0.9989	0.9838	0.9727	0.9800
	0.7	9.3	197.5	542.7	62.2	22.3	22.3	0.9025	0.9964	0.9990	0.9830	0.9853	0.9557
	0.8	14.4	289.7	654.9	94.5	34.1	34.1	0.9320	0.9980	0.9994	0.9894	0.9767	0.9727
	0.9	15.8	494.5	809.7	190.4	69.9	69.9	0.9391	0.9991	0.9997	0.9943	0.9509	0.9853
3.0	0.3	41.7	119.7	347.3	693.0	41.5	119.4	0.9725	0.9920	0.9972	0.9992	0.9572	0.9920
	0.4	18.7	138.2	394.6	398.7	18.7	138.2	0.9489	0.9939	0.9983	0.9982	0.9269	0.9934
	0.5	9.6	164.9	445.8	164.9	22.6	58.3	0.9173	0.9942	0.9980	0.9935	0.9443	0.9827
	0.6	12.6	205.6	513.7	73.6	12.5	29.0	0.9236	0.9956	0.9987	0.9859	0.9265	0.9664
	0.7	8.3	271.9	591.8	38.2	16.7	16.7	0.8932	0.9970	0.9989	0.9747	0.9650	0.9443
	0.8	7.6	376.9	698.0	40.8	13.0	13.0	0.8829	0.9982	0.9993	0.9761	0.9914	0.9265
	0.9	11.4	583.8	830.9	53.3	27.3	27.3	0.9201	0.9992	0.9996	0.9819	0.9932	0.9650



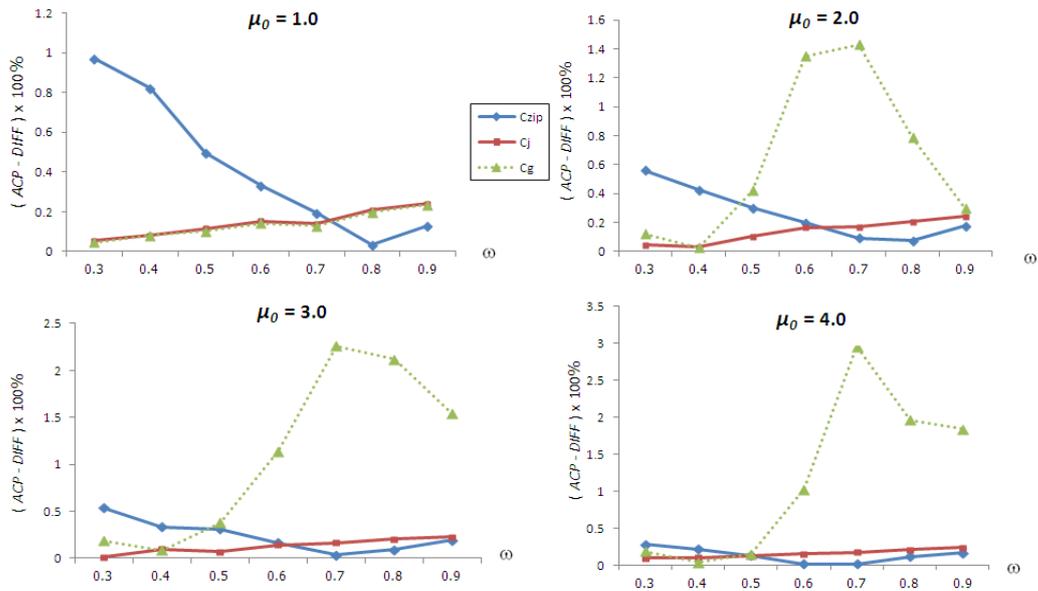


Fig. 2 Comparison of ACP-DIFF values of the c_{ZIP} -Chart, c_J -Chart and c_g -Chart for a range of μ_0 and ω values.

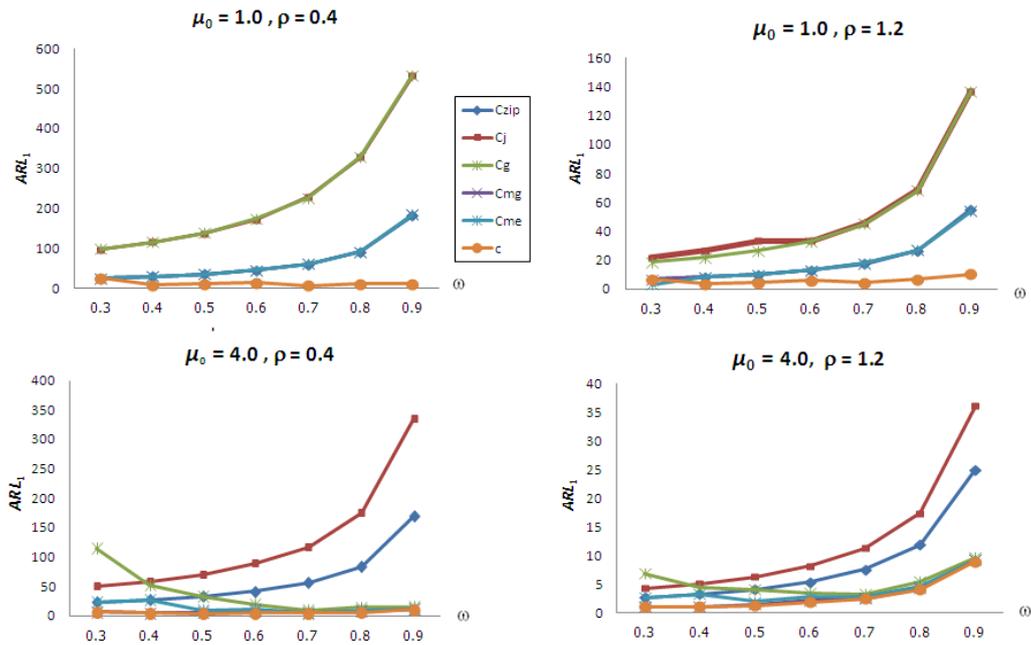


Fig. 3 Comparison of ARL_1 values of the c -Chart, c_{ZIP} -Chart, c_J -Chart, c_g -Chart, c_{mg} -Chart and c_{me} -Chart for a range of $\mu_1 = \mu_0 + \rho$, ω and ρ values

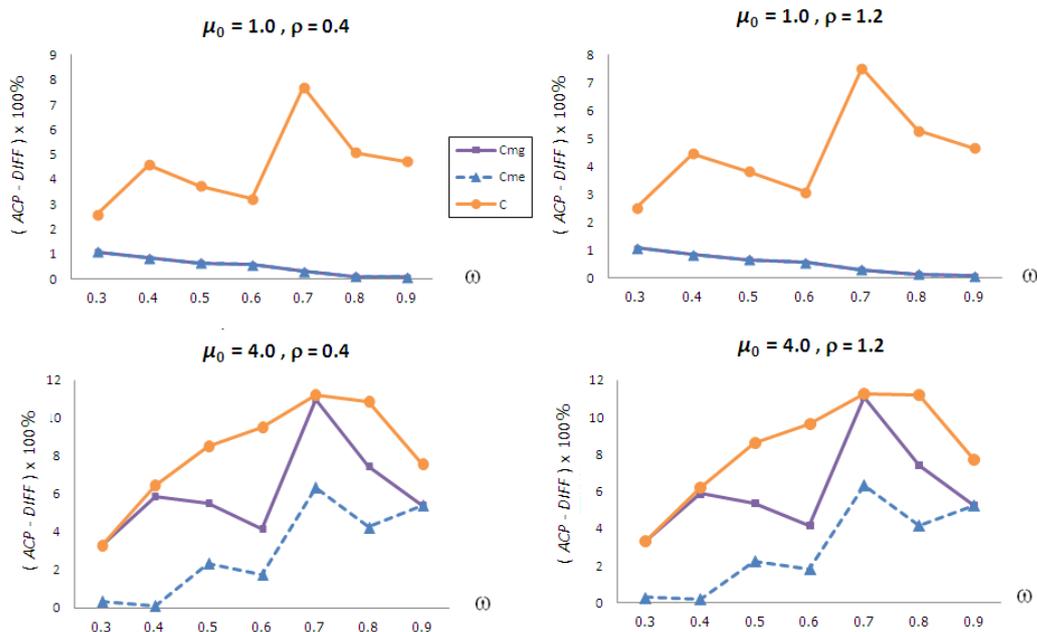


Fig. 4 Comparison of $ACP-DIFF$ values of the c -Chart and c_{mg} -Chart and c_{me} -Chart for a range of $\mu_1 = \mu_0 + \rho$, ω and ρ values

V. CONCLUSION

In this paper, three control charts have been proposed for a process with number of non-conformities from a ZIP distribution. In developing the charts, the number of non-conformities is modeled as a geometric distribution with parameter p_g , where p_g estimated gives the best fit between the geometric and ZIP distributions used in calculated the mean, median, and variance of geometric distribution for constructed the c -Chart by three difference methods. The three charts are called the c_g -Chart, c_{mg} -Chart and c_{me} -Chart. In the c_g -Chart, we constructed the control limits with the mean and variance of geometric distribution. In c_{mg} -Chart, the mean of geometric distribution for used replaced both the mean and variance in the control limits. In the c_{me} -Chart, we used the median and variance of geometric for replaced in control limits.

Additionally, the simulations have been carried out to compare the performances of the three control charts with the performances of three other charts: c -Chart, c_{ZIP} -Chart and c_J -Chart. The average run length (ARL) and average coverage probability (ACP) have been compared. The results of the comparisons are summarized in table III which gives a list of preferred control charts for both in-control and out-of control states for a range of values of ZIP parameters.

TABLE III
SUMMARY OF PREFERRED CONTROL CHARTS

The mean shift of process	Mean of process (μ_0 / μ_1)	Proportion of zero (ω)	Preferred control charts		
			For ARL value	For ACP value	For both ARL and ACP values
In-control	1.0	0.3 - 0.4	c_g -Chart and c_J -Chart	c_g -Chart and c_J -Chart	c_g -Chart and c_J -Chart
		0.5 - 0.9	c_g -Chart and c_J -Chart	c_{ZIP} -Chart and c_J -Chart	c_J -Chart
	2.0 - 4.0	0.3 - 0.4	c_g -Chart	c_g -Chart and c_J -Chart	c_g -Chart
		0.5 - 0.9	c_J -Chart	c_{ZIP} -Chart and c_J -Chart	c_J -Chart
Out-of-control (all level of ρ)	1.0	0.3 - 0.6	c -Chart, c_{mg} -Chart and c_{me} -Chart	c_{mg} -Chart and c_{me} -Chart	c_{mg} -Chart and c_{me} -Chart
		0.7 - 0.9	c -Chart	c_{mg} -Chart and c_{me} -Chart	-
	2.0 - 4.0	0.3 - 0.9	c -Chart, c_{mg} -Chart and c_{me} -Chart	c_{mg} -Chart and c_{me} -Chart	c_{mg} -Chart and c_{me} -Chart

ACKNOWLEDGMENT

The authors would like to thank for Rajamangala University of Technology Lanna for the financial support during this research

REFERENCES

- [1] A.C. Cohen, *Truncated and Censored Samples : Theory and Applications*. New York, Marcel Dekker, 1991.
- [2] C.H. Sim, and M.H. Lim, "Attribute Charts for Zero - Inflated Processes," *J. Communications in Statistics-Simulation and Computation.*, 2008, pp. 1440 - 1452.
- [3] D. Bohning, E. Dietz, and P. Schlattmann, "The zero-inflated Poisson model and the decayed missing and filled teeth index in dental epidemiology," *J. the Royal Statistical Society-Series.*, 1999, pp. 195 - 209.
- [4] D.C. Montgomery, *Introduction to Statistical Quality Control 5th Edition*. John Wiley & Sons, Inc, United States, 2005, pp. 288 - 290.
- [5] J.D. Gibbons, and S. Chakrabarti, *Nonparametric Statistical Inference 5th Edition*. Marcel Dekker, New York, 2003, pp. 111-130.
- [6] K. Krishnamoorthy, *Handbook of Statistical distribution with applications*. Taylor & Francis Group, New York, 2006, pp. 207 - 209.
- [7] M. Xie, B. He, and T. N. Goh, "Zero-inflated Poisson model in statistical process control." *J. Computational Statistics & Data Analysis*, 2001, pp. 191 - 201.
- [8] P.L. Gupta, R.C. Gupta, and R.C. Tripathi, "Analysis of zero-adjusted count data." *J. Computational Statistics & Data Analysis*, 1996, pp. 207 - 218.
- [9] T. Cai, "One-sided confidence intervals in discrete distributions." *J. Statistical Planning and Inference*, 2005, pp. 63 - 88.
- [10] V. Peerajit, and T. Mayuresawan, "Nonconforming Control Charts for Zero - Inflated Processes," In *Proceeding of 11th Conf. on statistic and applied statistic*, Holiday In, Chiangmai, Thailand, 2010, pp. 76.