On the Performance Evaluation of the Risk-adjusted Poisson Hurdle Cumulative Sum Chart

By

Muhammad Tahir

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

On the Performance Evaluation of the Risk-adjusted Poisson Hurdle Cumulative Sum Chart

By

Muhammad Tahir

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PHILOSOPHY IN STATISTICS

Supervised By

Dr. Sajid Ali

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

Declaration

I “Muhammad Tahir” hereby solemnly declare that this thesis titled, “ On the performance evaluation of the risk-adjusted Poisson hurdle cumulative sum chart ”.

- This work was done wholly in candidature for a degree of M.Phil Statistics at this University.
- Where I got help from the published work of others, this is always clearly stated.
- Where I have quoted from the work of others, the source is always mentioned. Except of such quotations, this thesis is entirely my own research work.
- Where the thesis is based on work done by myself jointly with my supervisor, I have made clear exactly what was done by others and what I have suggested

Dated:____________________ Signature:____________________

CERTIFICATE

On the Performance Evaluation of the Risk-adjusted Poisson Hurdle Cumulative Sum Chart

By

Muhammad Tahir

(Reg. No. 02222113010)

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF M.PHIL. IN

STATISTICS

We accept this thesis as conforming to the required standards

1. Sajid Ali
Dr. Sajid Ali
(Supervisor)

2. Saadia
Dr. Saadia Masood
(External Examiner)

3.
Prof. Dr. Ijaz Hussain
(Chairman)

DEPARTMENT OF STATISTICS
QUAID-I-AZAM UNIVERSITY
ISLAMABAD, PAKISTAN

2023

Dedication

I am feeling great honor and pleasure to dedicate this research work to my parents, whose unwavering support, love, and encouragement have been my guiding light throughout this journey. Your belief in me has been my greatest motivation. I also dedicate this work to my mentors and teachers, who have imparted knowledge and wisdom, shaping my intellectual growth. May this research contribute to the betterment of our society and stand as a tribute to the values you have instilled in me.

Acknowledgments

Boundless praise and gratitude to Allah Almighty, who is the most gracious, most Benevolent, and kind, peace and blessings of Allah be upon the Holy Prophet Hazrat Mohammad (S.A.W) and his pure and pious progeny, who are the source of knowledge and guidance for the entire world forever. I would like to express my deep gratitude and appreciation to my thesis advisor, **Dr. Sajid Ali**, for their invaluable guidance, support, and insightful feedback throughout the course of this research. Their expertise and dedication have been instrumental in shaping the direction and quality of this work. I feel highly privileged to take this opportunity to express my heartiest gratitude and a deep sense of indebtedness to the Honorable Supervisor **Dr. Sajid Ali**. Also, thanks to all the department teachers **Prof. Dr. Ijaz Hussain**, **Dr. Abdul Haq**, **Dr. Manzoor Khan**, **Dr. Ismail Shan**, and **Mam Dr Maryam Asim** for guiding me in all my research work.

My appreciation goes to my family for their unwavering love and encouragement. Your belief in me has been my driving force. I extended my heartfelt thanks to my all colleagues, classmates, and friends for their stimulating discussions, encouragement, and moral support during the ups and downs of this academic endeavor.

Abstract

Many operating environments, including manufacturing and healthcare, became more productive due to advances in science and technology. The effectiveness of a process is increased when a particular cause is precisely and accurately discovered. The hurdle Poisson process is frequently used for cases where the number of zeros is excessive. We suggest using risk-adjusted hurdle Poisson cumulative sum control charts to manage influenza risk.

In a simulation study, we compare how well traditional cumulative sum control charts stack up against risk-adjusted hurdle Poisson control charts. We evaluate their performance using the average run length (ARL) and its standard deviation. We use Tokyo's flu data to demonstrate how the proposed chart is applied. The charts reveal that, in the simulation studies, the unadjusted cumulative sum control chart outperforms the risk-adjusted one in terms of efficiency.

Contents

List of Tables

List of Figures

List of Abbreviations

SPC	Statistical Process Control
HP	Hurdle Poisson
ZIP	Zero-inflated Poisson
CUSUM	Cumulative sum
UCL	Upper Control Limit
LCL	Lower Control Limit
GZIP	Generalized Zero-inflated Poisson
GLM	Generalized Linear Model
MLE	Maximum Likelihood Estimation
SIDS	Sudden Infant Death Syndrome
MCE	Mixed CUSUM-EWMA
AIC	Akaike Information Criteria
AIDS	Acquired Immune Deficiency Syndrome
ARL	Average Run Lenght
OR	Odd Ratio
RR	Relative Risk
ARL_0	In-Control Average Run Length
ARL_1	Out-of-Control Average Run Lenght
ILI	Influenza-like illness

Chapter 1

Introduction

To ensure a process meets the desired quality standards, statistical process control (SPC) methods are used to distinguish between unusual and typical sources of variation. The SPC frequently uses control charts to monitor the process's consistency over time, which lowers process variability. Many different industries, including accounting, the stock market, and healthcare, use control charts. Boucher et al. (2007) discussed that the Poisson distribution's mean and variance are both equal to λ, it is equi-dispersed. Due to this potential weakness of the Poisson distribution, which may restrict its applicability, it is advised to use alternative distributions, such as hurdle models.

Mullahy (1986) was the first to discuss the data models for hurdle counts. With the aid of hurdle models, it is possible to distinguish statistically between individuals (observations) beneath and above the hurdle. To address results that are above a predefined hurdle and to differentiate between a binary outcome where the count falls either below or above this hurdle, a hurdle model is specifically integrated with a truncated model. Because of this, hurdle models are also known as two-part models. A hurdle count data model proves to be highly valuable when the hurdle is set at zero. The presence of excessive zeros can be explained by using the hurdle-at-zero approach. It implies that this model can be applied when the response variable contains a large number of zeros. The first part of the two-part model, P(Y=0), is defined by the hurdle at zero in this scenario. In various studies, different zero percentages in the response variable are discussed. Chen et al. (2008) they introduced a generalized zero-inflated Poisson (GZIP) model to account for multiple random shocks with varying probabilities, all following Poisson distributions. They recommended employing Shewhart charts, and cumulative sum (CUSUM) charts and evaluated control charts for probability as GZIP process monitoring tools. He et al. (2012) recommended the usage of the CUSUM chart to track ZIP processes. They developed the p-λ CUSUM, which combines the p-CUSUM and the λ-CUSUM, the chart will signal when either parameter shifts. They provided clear definitions for p-CUSUM, which monitors parameter p shifts, and λ-CUSUM, which monitors the parameter λ changes individually. For simultaneous monitoring of both parameters, they also suggested using the t-CUSUM. These diverse CUSUM control charts offer flexibility for various application scenarios, as outlined in their comparative study.

1.1 Background of the Problem

To comprehensively address these distinct features, the present study constructs an online quality monitoring and prediction system using the hurdle Poisson model. Tan et al. (2021) discussed that the CUSUM charts adjusted for risk are useful in monitoring to promptly and precisely signal influenza surveillance data. In simulation studies, they conducted a comparison between the ZIP CUSUM chart with risk adjustment and the standard CUSUM chart without adjustments. Furthermore, for the cases with excesses of zeros, the hurdle Poisson model is widely used which has not been considered in the literature for risk-adjusted charts.

1.2 Thesis Objectives

The thesis' main aim is to provide a thorough analysis of how to monitor the inflation of zeros using risk-adjusted hurdle Poisson CUSUM charts. especially to

- Propose an un-adjusted standard cumulative sum control chart.
- Create cumulative sum control charts for hurdle Poisson with risk adjustment.
- Evaluate the risk-adjusted and traditional charts by using the average run length criterion along with its standard deviation and quartiles.

The remainder of the thesis follows the structure: Chapter 2 encompasses the literature review. Chapter 3 delves into the discussion of both the proposed unadjusted standard CUSUM control charts and their risk-adjusted counterparts. Chapter 4 is dedicated to presenting results and engaging in discussions. Finally, Chapter 5 contains concluding remarks.

Chapter 2

Literature Review

Saghir and Lin (2015) proposed control charts are designed for monitoring count data under the assumption of a Poisson distribution. The Poisson distribution is characterized by having its mean and variance equal, making it an equi-dispersion distribution. The most commonly used c and u control charts are constructed based on the Poisson distribution. However, these traditional c and u control charts are not suitable when dealing with count data that either exhibits zero dispersion or inflation. In such cases, control charts using a generalized Poisson distribution are more appropriate. Additionally, the author recommends the use of the MCE (Modified Cumulative Exponential) chart as a novel control chart for monitoring process location. The analysis indicates that the proposed MCE control chart is significantly more sensitive to small and moderate shifts compared to existing charts, demonstrating a more effective structure.

Yeh et al. (2008) discussed that, in numerous cases, the EWMA control charts outperform the CUSUM control charts. Abujiya et al. (2015) suggested a CUSUM control chart for detecting variations in the standard deviation of a normal process across a range of shifts in process variability. Malela-Majika and Rapoo (2016) proposed that to identify mean shifts effectively, distribution-free CUSUM and EWMA control charts have been developed by incorporating the Wilcoxon rank-sum statistic within the context of ranked set sampling. Song et al. (2018) introduced a new individual EWMA control chart that uses the weighted likelihood ratio test, showcasing impressive effectiveness in diverse situations, including the detection of decreased variability and individual observations. Asghar et al. (2023) discussed the Poisson hurdle model as a basis for Shewhart-style control charts within the framework of Generalized Linear Models (GLM), especially for the monitoring of Pearson and deviance residuals.

Saffari et al. (2012a) proposed a hurdle-negative binomial (HNB) to overcome the dispersion of zero as an ordinary Poisson that cannot accommodate this problem. Saffari et al. (2013) suggested using the censored hurdle-generalized

Poisson regression model while examining the maximum likelihood method and the goodness of fit of the regression model is also evaluated in order to estimate the parameters. Alwani and Achmad (2021) proposed the Poisson hurdle model as a solution to address the high frequency of zero values, particularly evident in the context of acquired immune deficiency syndrome (AIDS) cases in Jambi province between 2015 and 2017. Through an evaluation using the Akaike information criteria (AIC), it has been demonstrated that the hurdle Poisson model provides a superior fit compared to the traditional Poisson regression model, offering a more suitable approach to handle the excess of zeros. Legisso et al. (2023) proposed the HP regression model to determine the factors that affect how often blood pressure is checked. They discussed the uses of extensions of the regression model, encompassing the zero-inflated model, the hurdle model, and the negative binomial model, in dental caries research. Additionally, they addressed fundamental aspects of model fitting, including the evaluation of goodness-of-fit. Dalrymple et al. (2003) compared the effects of climatic covariates during months with sudden infant death syndrome (SIDS) and without SIDS using three types of mixture models, namely finite mixture models, zero-inflated Poisson models, and hurdle models.

Zorn (1998) examined and compared the zero-inflated Poisson model with the Poisson hurdle model, using data related to congressional responses to supreme court rulings spanning from 1979 to 1988. Mahmood (2020) proposed the development of Shewhart-type control charts tailored for zero-inflated Poisson and zero-inflated negative binomial distributions, using Pearson residuals (PRs).

Bedrick and Hossain (2013) proposed a conditional test to assess the similarity between zero-inflated Poisson and Poisson-hurdle distributions. In order to forecast post-cardiac surgery mortality among individuals truly at risk, a model was introduced. This model exhibited high discriminative power and outstanding calibration, and it was developed using a retrospective population-based cohort study that used linked administrative data. Raja (2021) studied both, the Poisson distribution and the generalized Poisson distribution alongside the zero-inflated Poisson distribution. After evaluating the goodness of fit for two data sets by utilizing a portion of the zeroth cell, it was determined that the zero-inflated Poisson distribution offered a superior fit compared to both the Poisson distribution and the generalized Poisson distribution.

Mahmood et al. (2021) discussed that the EWMA and the CUSUM control charts indeed rely on the same probability distributions when compared to the Shewhart chart. They both leverage statistical distributions to monitor processes for detecting shifts or deviations from a mean or target value. Park et al. (2020) discussed Shewhart-type control charts based on GLMs for various distributions, such as the normal, binomial, Poisson, negative binomial, COM-Poisson, and

ZIP distributions. Urbieta et al. (2017) proposed EWMA and CUSUM control charts based on the negative binomial distribution. Lai et al. (2023b) proposed a risk-adjusted EWMA chart based on the ZIP process using a generalized likelihood ratio approach.

Chapter 3

Traditional and Risk-adjusted CUSUM charts

3.1 Hurdle Poisson Model

Mullahy (1986) proposed two models, ZIP and HP, which are the two most impactful methods for addressing the issue of excessive zero counts in data. The HP model consists of two parts. The first part introduces the zero-hurdle model, which helps assess the likelihood of observing a zero count. Typically, the probability of an excess of zeros is estimated using a logistic regression model. The covariates incorporated into the logistic regression may consist of factors that influence the likelihood of encountering a zero count, such as demographic variables, ecological variables, or other significant indicators, while the other part of the hurdle Poison model is the count model. The count model is used for the positive counts, which is conditional on the observation being non-zero. The frequently used count model is the Poisson regression model, which assumes that non-zero counts follow a Poisson distribution. Similar to any Poisson regression, covariates can be incorporated to capture the impact of the explanatory variables on the count. By combining the two stages, the Poisson hurdle model calculates the probability of observing a zero count (zero hurdle) and the mean count for non-zero observations (count model) simultaneously. Model boundaries were assessed using the most extreme probability assessment or Bayesian techniques. The benefits of the Poisson hurdle model include its capacity to deal with the abundance of zeros and capture the two-step process of count data generation. The researchers suggested using hurdle models when dealing with count data that exhibits an unusually high number of zeros.

Zuur et al. (2009) discussed the various modeling approaches, including zero-altered models, conditional models, and compatible models, to characterize hurdle models. These hurdle models consist of two key segments. In the initial part, the data is segregated into zero and non-zero categories and the likelihood of observing a zero value is computed using a binomial model. The subsequent section employs a truncated count model to describe the positive counts. The binary components were estimated using a binary model, such as logistic regression. In contrast, for estimating the positive count component, a zero-truncated count model, like the zero-truncated Poisson model, was used. The following is a mathematical representation of the model:

$$P(Q = 0) = s_1(0)$$

$$P(Q = q) = \left(\frac{1 - s_1(0)}{1 - s_2(0)}\right) s_2(q) \ , \qquad q = 1, 2, ...$$

where the probability mass function (PMF) s_2 represents positive counts, and s_1 represents zero counts. The following is how the HP PMF can be expressed:

$$P(X_i = x_i) = \begin{cases} p_i, & x_i = 0 \\ \frac{1-p_i}{1-e^{-\lambda_i}} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i}, & x_i > 0 \end{cases} \tag{3.1}$$

where the probability p_i of the binary component signifies whether the outcome is zero, and λ_i represents the mean response. The HP model's p_i can be generated using the logit model. The expressions for the mean and variance of these models are as follows:

$$E(X_i) = \mu_{HP} = \frac{(1 - p_i)\lambda_i}{1 - e^{\lambda_i}}$$

$$Var(X_i) = \sigma^2_{HP} = \frac{(1 - p_i)\lambda_i}{1 - e^{\lambda_i}} + \frac{\lambda_i^2(1 - \lambda_i)(\lambda_i - e^{\lambda_i})}{1 - e^{\lambda_i^2}}, \qquad 0 < p < 1 \ \ and \ \ \lambda > 0$$

3.1.1 Standard Hurdle Poisson CUSUM Model

The CUSUM control chart is primarily used for detecting small and persistent shifts or changes in a process. Its basic purpose is to provide a sensitive method for identifying gradual or incremental deviations from a target value or mean in a process. The CUSUM control chart was initially proposed by Page (1954). When using count data, ordinary Poisson cannot be used to check for inflation of zeros; instead, the hurdle Poisson CUSUM is used. This is especially helpful when there are more zeros in the data than a standard Poisson distribution can adequately illustrate. Experts can detect changes in count data with extra zeros using the hurdle Poisson CUSUM, enabling them to spot changes in the data and take appropriate action. The typical tabular CUSUM statistic is as follows:

$$S_t = \max(0, S_{t-1} + W_t)$$

where t takes on values 1, 2, 3, and so forth, S_t represents the CUSUM statistic at time t, with S_0 set to 0. Additionally, W_t represents the observational score for the t-th observation, determined using the log-likelihood ratio. Let H_0 be the alternative hypothesis and H_1 be the null hypothesis. Log(H_1/H_0) can be used to express the score W_t. The CUSUM control chart will trigger an alert when S_t exceeds the control limit denoted by h which is set to achieve a specific level of in-control performance. When the CUSUM statistic exceeds this control limit, it indicates that the process is no longer under control. The chart serves as an early warning system for practitioners, drawing attention to abrupt changes as they occur. Steiner et al. (2000) suggested the lower CUSUM statistic

$$M_t = \min(0, M_{t-1} + W_t)$$

where t ranging from 1, 2, 3, and so forth, with M_0 initially set to 0, the log-likelihood ratio forms the basis, while W_t represents the observation's score. The CUSUM control chart will trigger an alert when S_t falls below -h, where h represents the control limit set to attain the desired level of in-control performance. If the CUSUM statistic surpasses this control limit, it signifies that the process is considered out of control. According to the p-CUSUM chart, which was suggested by Tan et al. (2021), p_0 and p_1 are random probabilities in the context of the null and alternative hypotheses, respectively, and p_1 shows a specific shift of p_0.

$$W_t = \begin{cases} \log(\frac{p_1}{p_0}), & X = 0 \\ \log\left(\frac{1-p_1}{1-p_0}\right), & X > 0 \end{cases} \tag{3.2}$$

When p shifts, the process is monitored using this score in combined with the CUSUM statistic. Comparably, HP λ-CUSUM's score W_t is:

$$W_t = \begin{cases} \log(\frac{p_0}{p_0}), & X = 0 \\ \log\left(\frac{\lambda_1^{x_i} e^{-\lambda_i}(1-e^{-\lambda_0})}{\lambda_0^{x_i} e^{-\lambda_0}(1-e^{-\lambda_1})}\right), & X > 0 \end{cases} \tag{3.3}$$

used to monitor the process when λ shifts.

The score W_t of HP t-CUSUM can be characterized as follows:

$$W_t = \begin{cases} \log\left(\frac{p_1}{p_0}\right), & X = 0 \\ \log\left(\frac{\lambda_1^{x_i} e^{-\lambda_1}(1-e^{-\lambda_0})}{\lambda_0^{x_i} e^{-\lambda_0}(1-e^{-\lambda_1})}\right) + log\left(\frac{p_1}{p_0}\right), & X > 0 \end{cases} \tag{3.4}$$

In a simulation study, we take the parameters are known, but in actual situations, we must estimate them from Phase I using maximum likelihood estimation (MLE). The HP model's MLE procedure:

1. Formulate the HP model
2. Construct the likelihood function
3. Take the natural logarithm
4. Maximize the log-likelihood

   After maximizing the log-likelihood, you will obtain estimates for the model parameters

3.2 Hurdle Poisson Regression

Mullahy (1986) discussed that the logistic regression model estimates the log odds (logit) of observing zero as a function of the predictor variables. The predictors can include categorical and continuous variables, and their coefficients represent the effect on the likelihood of observing zero. The Poisson regression model was used to model positive counts. Specifically, the log of the expected positive count is estimated through this regression model, considering the predictor variables. The predictors can be the same or different, and their coefficients represent their effect on the expected count. Saffari et al. (2012b) describe the hurdle regression as follows:

$$logit(p) = \log\left(\frac{p_1}{1-p_1}\right) = \sum_{j=1}^{m} \zeta_i j \delta_j \tag{3.5}$$

where δ represents the vector of unknown parameters in m-dimensional columns and ζ_i represents the ith row of the covariate matrix denoted as Z, where ζ_i=1, ζ_i=2,..., ζ_i=m. In this setup, the logit link function is utilized to model the non-negative function p. In addition, since the value of λ is typically included in a log-linear model, the goal is to capture any systematic variation in that particular value.

$$\log(\lambda) = \sum_{j=1}^{m} X_i j \beta_j \tag{3.6}$$

The regression model incorporates the independent variables denoted as β_j". The overall count of independent variables in this regression model is represented as m. When the coefficients are known, it is possible to compute the values for the parameters p and λ. However, when dealing with applications where these coefficients and intercepts are not known, we can conveniently estimate these parameters by using the *hurdle()* function within the *pscl* package in R to fit the HP data.

3.2.1 Risk-Adjusted HP CUSUM

We designate λ_t as the mean and p_t as the adapted probability of a random shock. Eq.3.5 and Eq.3.6 can be used to determine the values of p_t and λ_t. Based on the odd ratios for shock probability p and relative risk for mean λ, respectively, we define the hypotheses H_0 and H_1. Let OR_0 and OR_1 stand for the respective odds ratios for the null and alternative hypotheses. The ratio $p_t/(1\text{-}p_t)$, where p_t denotes the estimated probability of a shock, represents the likelihood of a shock occurring. The odds of a shock for the t-th observation under the H_0 are given by $OR_0 p_t/(1\text{-}p_t)$, which corresponds to a probability of $OR_0 p_t/(1\text{-}p_t + OR_0 p_t)$. The odds are $OR_1 p_t/(1\text{-}p_t)$, which corresponds to a probability of $OR_1 p_t/(1\text{-}p_t + OR_1 p_t)$, under the H_1. Using the probability density function of the HP distribution, we calculate the log-likelihood ratio score for monitoring individual shifts in each p_t as follows:

$$W_t = \begin{cases} \log\left(\frac{OR_1(1-p_t+p_t OR_0)}{OR_0(1-p_t+p_t OR_1)}\right), & X = 0 \\ \log\left(\frac{(1-p_t+OR_0 p_t)}{(1-p_t+p_t OR_1)}\right), & X > 0 \end{cases} \tag{3.7}$$

This score is combined with CUSUM statistics to obtain p-CUSUM.

Derivation of Eq.3.7 we have let:

H_0: $OR_0 p_t/(1\text{-}p_t + OR_0 p_t)$,

H_1: $OR_1 p_t/(1\text{-}p_t + OR_1 p_t)$.

We have $\log(H_1/H_0)$

$\log(OR_1 p_t/(1\text{-}p_t + OR_1 p_t)/OR_0 p_t/(1\text{-}p_t + OR_0 p_t))$

$\log(OR_1(1\text{-}p_t + p_t OR_0)/OR_0(1\text{-}p_t + p_t OR_1))$.

Let the relative risks for the null and alternative hypotheses be RR_0 and RR_1 respectively,

H_0:$RR_0\lambda_t$, H_1:$RR_1\lambda_t$.

The H_0 and H_1 are associated with mean values of $RR_0\lambda_t$ and $RR_1\lambda_t$, correspondingly, for the mean number of observations at time t. We used the log-likelihood ratio score, which relies on the probability density function of the HP distribution, to monitor specific changes in λ_t.

$$W_t = \begin{cases} \log(\frac{p_t}{p_t}), & X = 0 \\ \log\left(\frac{RR_1 e^{-RR_1\lambda_t}(1-e^{-RR_0\lambda_t})}{RR_0 e^{-RR_0\lambda_t}(1-e^{-RR_1\lambda_t})}\right), & X > 0 \end{cases} \tag{3.8}$$

The score W_t of the t-CUSUM can be formulated as follows to detect changes in the two parameters, p and λ:

$$W_t = \begin{cases} \log\left(\frac{OR_1(1-p_t+p_tOR_0)}{OR_0(1-p_t+p_tOR_1)}\right), & X = 0 \\ \log\left(\frac{RR_1e^{-RR_1\lambda_t}(1-e^{-RR_0\lambda_t})}{RR_0e^{-RR_0\lambda_t}(1-e^{-RR_1\lambda_t})}\right)+\log\left(\frac{(1-p_t+OR_0p_t)}{(1-p_t+p_tOR_1)}\right), & X > 0 \end{cases} \tag{3.9}$$

To assess the performance of the control chart, we use the Average Run Length (ARL), which represents the average number of observations required before the CUSUM statistics initially exceed the control limit. This ARL is referred to as ARL_0 under the in-control condition H_0 and is purposefully set to be large to reduce false alarms. The ARL under H_1, ARL_1, should be as small as possible to allow for quick shift detection.

3.2.2 Algorithm to compute ARL

The value of h in the HP-CUSUM control chart serves as a fundamental determinant influencing the width of the chart. The subsequent steps illustrate the process for obtaining the control chart constants needed for specific charts.

1. Generate a data set from the HP model.
2. Calculate the score function on the basis of probability, mean, odd ratio, and relative risk.
3. Calculate the CUSUM model with their respective score function.
4. Plot the CUSUM statistic against the prefixed control limit and record the sample number at which it crosses the limit.
5. To achieve the prespecified ARL_0, repeat steps 1-4, 10,000 times. If the desired ARL_0 is not achieved, change the value of h and repeat the steps 1-5 until the desired ARL_0 is achieved.

Chapter 4

Results and Discussion

This chapter evaluates the chart detection capabilities using the HP model. The performance is assessed based on the ARL, SDRL, Q_1, Q_2, and Q_3 for $ARL_0 = 200$

4.0.1 Standard CUSUM

We detect upward shifts in the parameter $p_1 > p_0$ as well as the lower shifts $p_1 < p_0$. Similarly, for the detection of the shift of the parameter λ, both cases $\lambda_1 < \lambda_0$ or $\lambda_1 > \lambda_0$ are considered.

Table 4.1: Standard p-CUSUM assuming HP(λ,p) where p_0=0.10, λ_0= 1.14, ARL_0= 200

UCL = 59.5						
Rp	p_1	ARL	SDRL	Q_1	Q_2	Q_3
1.5	0.142	199.85	6.66	196	199	205
2	0.1818	126.68	6.3	122	127	130
2.5	0.217	103.69	6.75	98	103	180
3	0.25	93.211	7.25	88	93	98
LCL = 298.5						
Rp	p_1	ARL	SDRL	Q_1	Q_2	Q_3
0.2	0.021	200.4	2.3	199	200	202
0.4	0.042	366.5	4.189	364	366	369
0.6	0.062	682.011	7.28	677	682	687
0.8	0.081	1616.9	13.49	1608	1617	1625

Table 4.2: Standard λ-CUSUM assuming HP(λ,p) where ARL_0= 200, λ_0= 1.14.

UCL = 4.059						
Rl	λ_1	ARL	SDRL	Q_1	Q_2	Q_3
1.5	1.71	200.34	130.98	106	166	256
2	2.28	61.01	41.63	31	52	81
2.5	2.85	31.46	23	15	26	42
3	3.42	20.8514	15.44	10	17	28

4.0.2 Risk-adjusted CUSUM

The risk-adjusted p-CUSUM chart is made to take into account the various odd ratio levels associated with process changes. The UCL and LCL are calculated based on these odd ratio levels, providing a more sensitive and effective way to monitor the process. If the value of $OR_0 < OR_1$, the UCL is used. However OR_0 $> OR_1$, the LCL is used.

In practical application, we typically estimate the coefficients through the HP regression method, but in this context, we are simply assuming that they are already known. The constant values we have β=-0.5, δ= -1.386, ζ=1.06003, X=0.6695. Using these constant values in Eq.3.5 and Eq.3.6, we get the values of p_t=0.13087, λ_t= 0.7155.

Table 4.3: Risk-adjusted p-CUSUM assuming HP(λ,p), ARL_0= 200,R_0=2

UCL = 29.8					
R_1	ARL	SDRL	Q_1	Q_2	Q_3
2.5	200.7761	5.76362	196	200	204
3	113.58	4.49	111	114	116
3.5	84.53	4.47	81	84	87
4	70.13	4.19	67	70	73
LCL = 375					
R_1	ARL	SDRL	Q_1	Q_2	Q_3
0.2	199.85	4.75	196	200	203
0.4	295.2	5.85	291	295	298.25
0.6	404.16	7.13	400	404	409
0.8	542.4	8.46	536	542	548

Table 4.4: Risk-adjusted λ-CUSUM assuming HP(λ,p), ARL_0= 200,RR_0=2

UCL = 28.5					
RR_1	ARL	SDRL	Q_1	Q_2	Q_3
0.2	199.45	28.15	180	198	218
0.4	219.92	29.78	199	219	239
0.6	249.99	31.29	228	248	270
0.8	290.13	34.53	266	288	312
LCL = 8.94					
RR_1	ARL	SDRL	Q_1	Q_2	Q_3
2.5	200.02	28.2	180	198	218
3	100.14	19.93	86	99	239
3.5	65.29	16.18	53	64	270
4	49.78	13.96	40	48	312

The risk-adjusted λ chart is made to take into account the various relative risk values connected to uncommon events or defects in a process. The UCL and LCL are calculated based on these relative risk values, providing a more sensitive and effective way to monitor the rate of rare events. When $RR_0 > RR_1$ the UCL is calculated. In contrast, when $RR_0 < RR_1$ the LCL is calculated.

Table 4.5: Comparison of standard and risk-adjusted t-CUSUM assuming $R_p \in$ (1.5,2,2.5,3), $R_l \in$ (1.5,2,2.5,3), $OR_1 \in$(1.5,2,2.5,3), $RR_1 \in (1.5, 2, 2.5, 3)$, OR_0= 1, RR_0= 1,ARL_0= 200, $\lambda_1 \in$ (1.71,2.28,2.85,3.42)

		Standard t-CUSUM					Risk-adjusted t-CUSUM			
		UCL $\in$(67,108,137,160)					UCL $\in$(35.1,65.3,85.5,102)			
R_p	R_l	1.5	2	2.5	3	OR_1/RR_1	1.5	2	2.5	3
1.5	ARL	200.8	167.32	132.51	104.27	1.5	200	297.88	603.27	4421.42
	SDRL	10.18	16.56	18.28	17.77		22	53.96	196.41	-
	Q_1	194	156	120	92		184	260	462	-
	Q_2	201	167	132	104		198	292	572	-
	Q_3	208	179	144	116		214	330	708	-
2	ARL	200.21	173.7	143.89	117.3	2	201.32	245	316.38	452.5
	SDRL	7.03	11.93	14.2	14.85		17.31	29.15	29.15	29.15
	Q_1	195	166	134	107		189	224	280	379
	Q_2	200	174	144	117		200	244	321	441
	Q_3	205	182	153	127		213	263	347	511
2.5	ARL	199.56	174.9	147.13	122.18	2.5	199.7	229.8	274.07	341.9
	SDRL	5.966	10.21	12.53	13.12		15.19	23.9	36.57	58.02
	Q_1	196	168	139	113		186	214	248	300
	Q_2	200	175	147	122		199	228	271	336
	Q_3	204	182	156	130		209	245	297	377
3	ARL	200.36	176.78	149.24	124.4	3	200.5	226.08	260.86	309.81
	SDRL	5.32	9.46	11.47	12.12		16.27	22.44	31.63	47.26
	Q_1	197	170	141	116		189	210	239	277
	Q_2	200	177	149	124		200	225	259	306
	Q_3	204	183	157	132		211	241	281	340

4.1 Standard CUSUM Analysis

4.1.1 Shifts in p_1 as $R_p p_0/1+(R_p$ -1)p_0

Table 4.1 summarizes the outcomes of altering R_p, which indirectly impacts the shifts in p_1. Where p_1=$R_p p_0/1+(R_p$-1)p_0 when the value of ARL_0 is 200. A shift size of 0.1818 in p_1 can cause ARL_1 = 126.68 while SDRL-6.66. A decreasing trend in the ARL is observed as R_p increases. This means that the higher R_p values used in the standard p-CUSUM chart lead to quicker detection of shifts.

A lower ARL implies that the chart possesses greater sensitivity in detecting shifts when associated with higher R_p values. In the case of LCL, the shift size 0.042 can cause ARL_1= 366.5 and SDRL= 4.189. The ARL values against the comparing R_p values show an increasing pattern as R_p increases. This implies that larger R_p values used in the standard p-CUSUM chart lead to a higger number of samples needed to detect shifts. A larger ARL shows that the chart turns out to be less sensitive to shifts with higher R_p values.

4.1.2 Shifts in λ_1 as $R_l\lambda_0$

Table 4.2 summarizes the results that either $\lambda_0 < \lambda_1$ or $\lambda_0 > \lambda_1$ and the UCL CUSUM statistic is applicable. We have λ_1 = 2.28, and the ARL_1 is roughly 61.01. The SDRL which measures the variability in detection performance, is approximately 41.63. This implies that higher R_l values used in the standard λ-CUSUM chart lead to speedier identification of changes in the process mean. A smaller ARL means that the chart turns out to be more sensitive in detecting shifts with higher R_l values.

4.2 Risk-adjusted CUSUM Analysis

4.2.1 Shifts in OR_1

The results of Table 4.3 show the direct shifts in OR_1. If we take $R_1 = 3$, the ARL_1 is approximately 113.58, that is the process will, on average, detect a shift in the mean every 113.58 consecutive samples. The SDRL is approximately 5.134, which indicates variability in the detection performance. Similarly, if we assume that $OR_1 = 0.4$ the results will typically identify a shift in the mean every 295.2 consecutive samples with LCL. The ARL values decrease while OR_1 increases from 2.5 to 4. This means that higher odd ratios lead to a speedier identification of changes. In general, the ARL rises as the OR_1 value rises. This suggests that more samples are required to identify changes in the process mean when larger odds ratios are used in the risk-adjusted p-CUSUM chart. A larger ARL indicates that the chart becomes less sensitive to detect shifts with higher odds ratios.

4.2.2 Shifts in RR_1

Table 4.4 summarizes the results of shifts in RR_1. Assuming $RR_1 = 0.4$, the ARL is approximately 219.92, that is, the process will, on average, detect a shift with a relative risk of 0.4 for every 219.92 consecutive samples. The SDRL is approximately 29.78, which suggests the degree of variability in the run length. By analyzing the ARL values in relation to different values of RR_1, it is evident that as RR_1 increases from 0.2 to 0.8, the ARL values also increase. This trend indicates that as the relative risk of identifying a shift in the process mean increases, a greater number of consecutive samples are necessary to detect the shift. In contrast, in the case of LCL where RR_1 is 3, the ARL is approximately 100.14. This indicates that a shift in relative risk of 3 will be detected, on average, after every 100.14 samples. The ARL values show that as the RR_1 increases, the ARL values decrease. This means that higher relative risks lead to faster detection of shifts. Just as with this, lower ARL values suggest that the risk-adjusted λ-CUSUM chart is more sensitive in detecting shifts with higher relative risks.

4.3 Comparison of Traditional and Risk-adjusted t-CUSUM Analysis

4.3.1 Shifts in p_1 and λ_1

Table 4.5 provides an overview of the results stemming from the indirect shifts in both p_1 and λ_1. We consider different values of the parameter Rp and R_l and assume both the parameters are independent. It is noticed that the ARL values show a decreasing trend for different values of Rp and R_l. This means that the trend leads to quicker detection shifts. However, in the case of the standard t-CUSUM, a smaller ARL shows that the chart is more sensitive.

4.3.2 Shifts in RR_1 and OR_1

Similarly, Table 4.5 summarizes the results of RR_1 and OR_1. In these results, the ARL values show an increasing trend which means that a larger number of samples is needed to detect shifts. In addition, a larger value of ARL shows that the chart turns out to be less sensitive. This suggests that more samples are required to identify changes in the process mean when larger odds ratios are used in the risk-adjusted. To establish the presence of a shift, a larger number of samples is required as the relative risk of detecting a shift in the process mean increases

Furthermore, a comparison between standard t-CUSUM and risk-adjusted t-CUSUM results in Table 4.5 indicates that the standard t-CUSUM chart shows greater sensitivity than the risk-adjusted t-CUSUM.

4.4 Real Data Application

This section uses the Tokyo influenza dataset (Imai et al., 2015) to serve as an example of how to use the risk-adjusted CUSUM control chart. The dataset comprises weekly counts of reported cases of influenza-like illnesses obtained from the National Institute of Infectious Diseases (NIID) in Japan. This data contains cases from Tokyo from April 1999 to March 2004. Since influenza season in Japan generally occurs from October through the following March, the epidemic year starts from April and it is used instead of the calendar time year to cover the whole course of each influenza epidemic season.

To assess the effectiveness of our suggested approach for handling influenza data that includes an excess of zero values, we select the weekly influenza data from 1999 to 2002 as the Phase-I dataset. These data are used to create control charts. The remaining data from 2003-2004 is used for monitoring.

Weekly ILI cases from 1999 to 2004 can be observed in Fig. 4.1. The coefficients used in our analysis are derived through HP regression using the 1999-2002 ILI data, as presented in Table 4.8.

We use a modified CUSUM chart and a standard CUSUM chart to monitor the data from 2003 to 2004. Figure 4.2 displays the risk-adjusted CUSUM chart with h= 6.675, while Figure 4.3 presents the standard CUSUM chart with h= 220, where ARL_0=200. In CUSUM charts, the solid line represents the observations, while the red line represents the control limit. It is evident from Figure 4.2 that the risk-adjusted CUSUM chart detects an out-of-control (OOC) signal during the 9th week. Conversely, Fig. 4.3 demonstrates that the standard CUSUM chart detects OOC starting from the 56th week. This study's proposed chart signals an OOC condition during the 9th week, indicating its superior efficiency compared to the traditional chart.

Table 4.6: Summary of real data 1999-2004

	Minimum	Q_1	Median	Mean	Q_3	Maximum
Year	1999	2000	2002	2002	2003	2004
sntl	140	178	178	176	178	188
count_flu	0	0	6.50	49.60	70.25	921
temp_mean	3.329	10.11	17.42	17.16	23.24	30.68

There are three variables described in the summary Table.4.6: "sntl" has a fairly stable distribution around a mean of 176, "count flu" shows a distribution that is significantly skewed, featuring an average of 93.93, and the presence of potential outliers, and "temp mean" represents temperature data with a mean of 17.16 and a moderate spread between quartiles.

Table 4.7: Test Results for Dispersion

Dispersion	z	p
61.8126	2.0432	0.02052

Table 4.7 shows a dispersion test in the hurdle Poisson regression model. The dispersion value is larger than 1 and p value is less than 0.05 indicates overdispersion.

Table 4.8: HP regression coefficients of Phase I data 1999-2002

Count model				
	Estimate	Standard error	z value	Pr(>|z|)
Intercept	4.69	0.2982	15.731	2.00E-16
b1	0.012228	0.00165	7.395	1.41E-13
b2	-0.176081	0.001802	-97.71	2.00E-16
Zero hurdle model				
	Estimate	Standard error	z value	Pr(>|z|)
Intercept	9.03213	3.2038	2.819	0.00481
b1	-0.006255	0.016719	-0.374	0.70831
b2	-0.345138	0.051125	-6.751	1.47E-11

These models are used to examine count data from Table 4.8, which contains a substantial number of zero values. The count model describes the relationship between predictors and the count itself, whereas the zero hurdle model describes the likelihood of observing zero counts. The statistical significance of each coefficient is determined by the p-value, which is set at 0.05; lower p-values than 0.05 indicate greater statistical significance.

Table 4.9 presents a summary of the AIC, BIC, and log-likelihood for each model. The Poisson model may not be suitable for the data due to its elevated

Table 4.9: The log-likelihood, AIC, and BIC, of different model results

Model	AIC	BIC	Log-likelihood
Poisson	77409.86	77423,53	–
ZIP	30758.94	30781.17	-1.5370
ZINB	2227.61	2253.53	-1107
HP	30758.77	30780.99	-1.5370

AIC and BIC values. On the other hand, the ZIP model, with its lower AIC and BIC values, could offer a more favorable fit. The ZINB model, which exhibits the lowest AIC and BIC values, appears to be the most suitable choice for this data. Similarly, the HP model, characterized by its low AIC and BIC values, also presents a better fit.

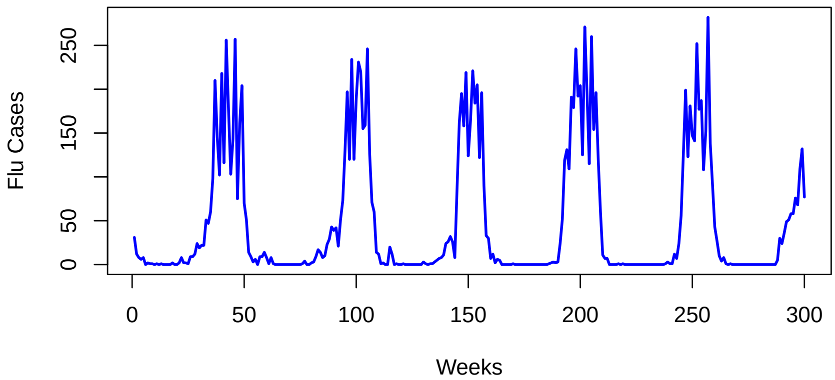

Figure 4.1: weekly Influenza-Like Illness cases during 1999-2004

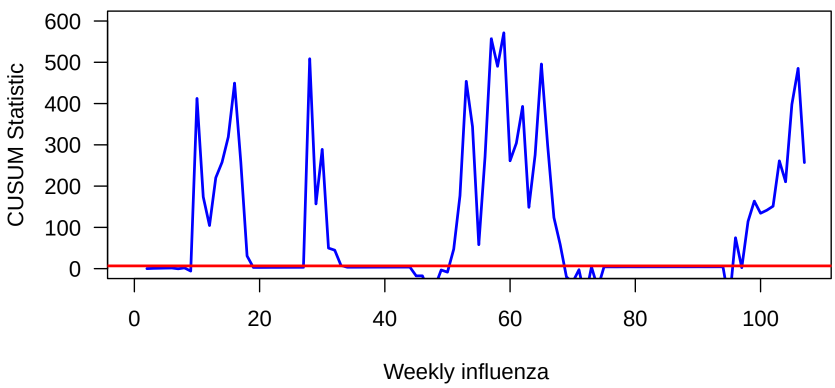

Figure 4.2: Risk-adjusted CUSUM chart for 2003-2004 influenza monitoring

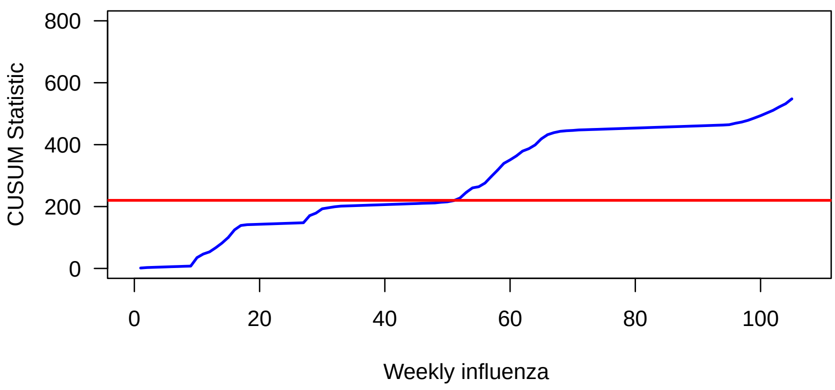

Figure 4.3: Standard CUSUM chart for 2003-2004 influenza monitoring

4.4.1 New Method for Comparison of Traditional and Risk-adjusted CUSUM charts

Lai et al. (2023a) discussed that during Phase I, the in-control parameters remain constant and are estimated using MLE along with the Newton-Raphson method. Let's consider that x_i is collected over time from the change-point model described below.

$$x_i \sim \begin{cases} \text{HP}(p_{0_i}, \lambda_{0_i}) & \text{for } i = 1, \ldots, \tau - 1 \\ \text{HP}(p_{1_i}, \lambda_{1_i}) & \text{for } i = \tau, \tau + 1, \ldots \end{cases} \tag{4.1}$$

τ is the unknown change-point, and p_{1_i} and λ_{1_i} represent the parameters after time τ for observations i. Based on this model, hypotheses can be formulated.
$H_0 : p_{1_i} = p_{0_i}$,λ_{1_i}=λ_{0_i}
$H_1 : p_{1_i} \neq p_{0_i}$, $\lambda_{1_i} \neq \lambda_{0_i}$

The control limits are established for the in-control data to accommodate an acceptable shift, and the CUSUM scheme will monitor the range of shift between predicted and estimated values. We set $\alpha_\lambda = \frac{\lambda_{1_i}}{\lambda_{0_i}}$, $\alpha_p = \frac{p_{1_i}}{p_{0_i}}$ to represent the shift range, and the CUSUM statistics are given by

$$C_i = \max(0, C_{i-1} + \alpha_p)$$

$$D_i = \max(0, D_{i-1} + \alpha_\lambda)$$

where C_0= 1, D_0= 1.

Based on hypotheses the likelihood ratio test statistic for the risk-djusted HP process can be obtained as

$$R = \begin{cases} \log\left(\frac{C_i p_1}{p_0}\right), & X = 0 \\ \log\left(\frac{D_i \lambda_1^{x_i} e^{-D_i \lambda_1}(1-e^{-\lambda_0})}{\lambda_0^{x_i} e^{-\lambda_0}(1-e^{-D_i \lambda_1})}\right) + \log\left(\frac{C_i p_1}{p_0}\right), & X > 0 \end{cases} \tag{4.2}$$

To construct the control chart, the chart statistic R is compared to the control limit. When R exceeds the control limit, alarm signals should be triggered.

Table 4.10: Comparison of standard and risk-adjusted CUSUM assuming p_0= 0.10, λ_0= 1.14 ARL_0= 200, $\lambda_1 \in$ (2.14,3.14,4.14,5.14) $p_1 \in$ (0.20,0.40,0.50,0.60), UCL-s $\in$ (149,178,220), UCL-r $\in$(6.675)

		Standard CUSUM					Risk-adjusted CUSUM			
λ_1	p_1	0.2	0.40	0.5	0.6	λ_1/p_1	0.2	0.40	0.5	0.6
2.14	ARL	200.36	100.36	86.08	76.91	2.14	200.108	51.1	33.09	24.08
	SDRL	7.22	3.566	3.18	2.96		0.346	0.925	0.749	0.821
	Q1	196	98	84	75		200	51	33	24
	Q2	200	100	86	77		200	51	33	24
	Q3	206	103	88	79		203	51	33	24
3.14	ARL	200.77	100.67	85.39	75.4	3.14	200.108	51.117	33.11	24.109
	SDRL	13.73	6.564	5.65	5.06		0.346	0.36	0.479	0.414
	Q1	191	96	82	72		200	51	33	24
	Q2	201	101	85	75		200	51	33	24
	Q3	210	105	89	79		200	51	33	24
4.14	ARL	200.31	99.6	83.45	72.8	4.14	200.104	51.117	33.11	24.11
	SDRL	17.9	8.63	7.15	6.25		0.343	0.36	0.36	0.35
	Q1	186	94	74	69		200	51	33	24
	Q2	198	99	83	73		200	51	33	24
	Q3	211	105	88	77		200	51	33	24
5.14	ARL	199.8	99.87	82.9	71.66	5.14	200.108	51.117	33.11	24.11
	SDRL	20.95	9.8	7.93	6.744		0.3466	0.364	0.36	0.35
	Q1	185	93	77	67		200	51	33	24
	Q2	199	100	83	72		200	51	33	24
	Q3	213	106	88	76		200	51	33	24

Table 4.10 provides an overview of the results arising from direct shifts in both p_1 and λ_1. We consider various values for the parameters p_1 and λ_1, assuming their independence. It is observed that the ARL values exhibit a decreasing trend across different combinations of p_1 and λ_1, indicating quicker detection of shifts. However, the risk-adjusted CUSUM shows greater efficiency compared to a standard CUSUM in terms of both the ARL and SDRL for the specified parameter combinations.

Chapter 5

Conclusion And Recommendations

When dealing with count data analysis, we frequently encounter the challenge of an inflation of zeros. For counting data with too many zeros, we suggest using a risk-adjusted HP CUSUM chart. We assessed the effectiveness of our suggested chart by conducting a detailed simulation study, focusing on its run-length characteristics.

In a simulation study the standard p-CUSUM chart, when Rp is transformed into p_1 using the formula $p_1 = \text{Rp}p_0/(1+(\text{Rp-1})p_0)$ with an initial ARL_0 of 200. A chart with UCL shows quicker shift detection and greater chart sensitivity. Conversely, the LCL chart suggests that more samples are needed for shift detection, indicating reduced chart sensitivity.

The standard λ-CUSUM chart results show quicker shift detection and greater sensitivity. Risk-adjusted p-CUSUM chart with the UCL shows more sensitivity than the LCL chart. Also, higher odds ratios lead to faster detection. In contrast, the LCL chart has larger odds ratios and needs more samples for a change identification, showing reduced sensitivity. The risk-adjusted λ-CUSUM chart with the UCL shows the ARL values increase, which means requiring more consecutive samples to detect shifts with higher relative risk. Thus, the LCL chart shows more sensitivity than the UCL.

The simulation results indicate that in the comparative scenario, risk-adjusted HP CUSUM chart performs better than the traditional HP CUSUM chart. For real-world data sets, we implement the proposed risk-adjusted hurdle Poisson model to demonstrate its suitability and effectiveness in practical situations.

In the future, control charts for data with inflated zero values might be made using different hurdle models. It is feasible to introduce risk-adjusted charts based on the generalized likelihood ratio. It is also possible to introduce the hurdle Poisson EWMA case as well as risk-adjusted zero-inflated Poisson EWMA charts.

Bibliography

Abujiya, M. R., Riaz, M., and Lee, M. H. (2015). Enhanced cumulative sum charts for monitoring process dispersion. *PLoS One*, 10(4):e0124520.

Alwani, N. N. and Achmad, A. I. (2021). Model regresi hurdle Poisson dalam mengatasi permasalahan excess zero untuk kasus aids di provinsi jambi tahun 2015-2017. *Prosiding Statistika*, pages 557–563.

Asghar, M., Ali, S., and Shah, I. (2023). Poisson hurdle model for monitoring the inflation of zeros. *Quality and Reliability Engineering International*.

Bedrick, E. J. and Hossain, A. (2013). Conditional tests for homogeneity of zero-inflated Poisson and Poisson-hurdle distributions. *Computational Statistics & Data Analysis*, 61:99–106.

Boucher, J.-P., Denuit, M., and Guillén, M. (2007). Risk classification for claim counts: a comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal*, 11(4):110–131.

Chen, N., Zhou, S., Chang, T.-S., and Huang, H. (2008). Attribute control charts using generalized zero-inflated Poisson distribution. *Quality and Reliability Engineering International*, 24(7):793–806.

Dalrymple, M. L., Hudson, I. L., and Ford, R. P. K. (2003). Finite mixture, zero-inflated Poisson and hurdle models with application to sids. *Computational Statistics & Data Analysis*, 41(3-4):491–504.

He, S., Huang, W., and Woodall, W. H. (2012). CUSUM charts for monitoring a zero-inflated Poisson process. *Quality and Reliability Engineering International*, 28(2):181–192.

Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., and Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental Research*, 142:319–327.

Lai, X., Lian, X., Liu, L., Wang, J., Liu, Y., and Chong, K. C. (2023a). Generalized likelihood ratio based risk-adjusted control chart for zero-inflated poisson process. *Quality and Reliability Engineering International*, 39(1):363–381.

Lai, X., Lian, X., Liu, L., Wang, J., Liu, Y., and Chong, K. C. (2023b). Generalized likelihood ratio based risk-adjusted control chart for zero-inflated Poisson process. *Quality and Reliability Engineering International*, 39(1):363–381.

Legisso, T. Z., Mamo, B. G., Bimrew, A. M., and Fikadu, T. (2023). Blood pressure examination habit and its determinants among civil servants in Arba Minch Town: A cross-sectional study–using hurdle Poisson regression model. *Integrated Blood Pressure Control*, pages 1–9.

Mahmood, T. (2020). Generalized linear model-based monitoring methods for high-yield processes. *Quality and Reliability Engineering International*, 36(5):1570–1591.

Mahmood, T., Balakrishnan, N., and Xie, M. (2021). The generalized linear model-based exponentially weighted moving average and cumulative sum charts for the monitoring of high-quality processes. *Applied Stochastic Models in Business and Industry*, 37(4):703–724.

Malela-Majika, J.-C. and Rapoo, E. (2016). Distribution-free cumulative sum and exponentially weighted moving average control charts based on the wilcoxon rank-sum statistic using ranked set sampling for monitoring mean shifts. *Journal of Statistical Computation and Simulation*, 86(18):3715–3734.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3):341–365.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Park, K., Jung, D., and Kim, J.-M. (2020). Control charts based on randomized quantile residuals. *Applied Stochastic Models in Business and Industry*, 36(4):716–729.

Raja, T. (2021). Generalized Poisson distribution and zero-inflated Poisson distribution with reference to Poisson distribution. *The Pharma Innovation Journal The Pharma Innovation Journal*, 10(5):1471–1473.

Saffari, S., Adnan, R., and Greene, W. (2012a). Hurdle negative binomial regression model with right censored count data. *SORT*, 36(2):0181–194.

Saffari, S., Adnan, R., and Greene, W. (2012b). Parameter estimation on hurdle Poisson regression model with censored data. *Jurnal Teknologi (Sciences and Engineering)*, 57(SUPPL. 1):189–198.

Saffari, S. E., Adnan, R., and Greene, W. (2013). Investigating the impact of excess zeros on hurdle-generalized Poisson regression model with right censored count data. *Statistica Neerlandica*, 67(1):67–80.

Saghir, A. and Lin, Z. (2015). Control charts for dispersed count data: an overview. *Quality and Reliability Engineering International*, 31(5):725–739.

Song, Z., Liu, Y., Li, Z., and Zhang, J. (2018). A weighted likelihood ratio test-based chart for monitoring process mean and variability. *Journal of Statistical Computation and Simulation*, 88(7):1415–1436.

Steiner, S. H., Cook, R. J., Farewell, V. T., and Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics*, 1(4):441–452.

Tan, Y., Lai, X., Wang, J., Zhang, X., Zhu, X., Chong, K.-C., Chan, P. K., and Tang, J. (2021). Risk-adjusted zero-inflated Poisson CUSUM charts for monitoring influenza surveillance data. *BMC Medical Informatics and Decision Making*, 21(2):1–11.

Urbieta, P., Lee HO, L., and Alencar, A. (2017). CUSUM and EWMA control charts for negative binomial distribution. *Quality and Reliability Engineering International*, 33(4):793–801.

Yeh, A. B., Mcgrath, R. N., Sembower, M. A., and Shen, Q. (2008). EWMA control charts for monitoring high-yield processes based on non-transformed observations. *International Journal of Production Research*, 46(20):5679–5699.

Zorn, C. J. (1998). An analytic and empirical examination of zero-inflated and hurdle Poisson specifications. *Sociological Methods & Research*, 26(3):368–400.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*, volume 574. Springer.