THREE-STAGE ADAPTIVE BATCH TESTING MODEL FOR ESTIMATING PREVALENCE OF A TRAIT WITH ERRORS WITHOUT TRUNCATION

ODUORI VICTOR

A thesis submitted in partial fulfillment of the requirements for the award of degree of Master of Science in Statistics of Masinde Muliro University of Science and Technology.

FEBRUARY, 2021

DECLARATION

This thesis is my original work which has not been presented elsewhere for an award of a degree or any other award.

Signature..... Date .....

Odouri Victor

SES/G/03/2016

CERTIFICATION

The undersigned certify that they have read and hereby recommend for acceptance of Masinde Muliro University of Science and Technology a thesis entitled, "Three-stage adaptive batch testing model for estimating prevalence of a trait with errors without truncation."

Signature..... Date .....

Dr. Annette .W. Okoth

Department of Mathematics

Masinde Muliro University of Science and Technology

Signature..... Date .....

Dr. Ronald .W. Wanyonyi

Department of Mathematics

Egerton University.

COPYRIGHT

DEDICATION

To my beloved brothers Benard, Joseph and Peter for developing a lot of trust and confidence in me throughout this journey.

ACKNOLEDGEMENT

To my supervisors Dr.Okoth A.W and Dr.Wanyonyi R. W, I express my utmost gratitude to you for your invaluable mentorship, distinguished scholarly guidance and technical advice. I would also like to acknowledge Dr. J. K.Nthiiri, C.O.D Mathematics and other members of staff, for their encouragement during this study. Sincere thanks to my parents Ernest and Anjeline for their moral and financial support for without them this work would not have been a reality. My gratitude to God Almighty for his grace and provision on this journey.

ABSTRACT

Batch testing is a fundamental testing scheme that results into substantial saving in terms of cost and time. It is mainly applicable in cases with large population sizes and low prevalence rates. Studies on Batch testing have shown that Adaptive Batch testing is more efficient than Non-Adaptive Batch testing particularly as the number of stages increases. Most recent studies on Batch testing have shown that even with truncation in inspection, Adaptive Batch testing remains more efficient. In this study a Three-Stage Adaptive Batch testing Model with errors without Truncation is presented with the view to establishing whether or not it is more efficient than the truncated estimator. Maximum Likelihood Estimate (MLE) method is used to obtain the estimator and Crammer-Rao Lower Bound method to determine the variance of the estimator. Data is obtained through simulation by the help of R-Software tool .The efficiency of the Estimator relative to the Adaptive estimator with truncation was determined with the view to performing a comparative analysis between the two. Model verification is done and the results show that Three stage-Adaptive Batch testing model with errors without truncation is more efficient than truncated estimator in the presence of errors. This study is significant in the sense that it brings forth a new model in the literature of estimation in batch testing, a Model that would find application in various fields such as HIV/AIDS , Blood donation, quality control among others.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION AND NOTATIONS

Table 0.1: List of abbreviations.

Abbreviation	Meaning
MLE	Maximum likelihood Estimator
C.R.L.B	Crammer Rao-Lower Bound
PCR	Polymerace Chain Reaction
RNA	Ribonucleic Acid
HIV	Human Immunodeficiency Virus
ARE	Asymptotic Relative Efficiency

Table 0.2: List of Notations.

Notation	Meaning
N	Population sample size
n	Number of Batches constructed
p	Probability that an individual tests positive
$\hat{p}$	The estimator of p (Non-adaptive scheme)
$\hat{p}_j$	The estimator of p at stage j (Truncated)
$\hat{p}_A$	The estimator of p at stage Two (Non-Truncated)
$\hat{p}_B$	The estimator of p at stage Three (Non-Truncated)
$\pi_j(p)$	Probability that a batch test positive at stage j
$\text{Var}(\hat{p}_A)$	Variance of A (Non-Truncated)
$\text{Var}(\hat{p}_B)$	Variance of B (Non-Truncated)
$\text{Var}(\hat{p}_j)$	Variance of p_j (Truncated)
k	Batch size (Non-adaptive scheme)
k_j	Batch size at stage j (adaptive scheme)
η	Sensitivity of the test kit
ϕ	Specificity of the test kit
λ	Partitioning parameter (Non-Adaptive scheme)
λ_j	Partitioning parameter at stage j (Adaptive scheme)

CHAPTER ONE

INTRODUCTION

1.1 Background information

Batch testing is a fundamental tool used in identifying defective items from a large population with low prevalence rates. It is designed to reduce the number of tests required to identify defective items, thus very economical in terms of time and cost. In this scheme items from a given population are pooled and tested as a single entity [5]. Batch testing has been known to reduce the variance of the estimator thereby making it more efficient, [13] , [15].

Batch testing is categorized into two; Adaptive and Non-Adaptive. In a Non-Adaptive scheme, a large population is divided into 'n' batches and are subjected to testing, [5]. The results obtained are then used to construct the Non-adaptive scheme. Adaptive scheme on the other hand involves partitioning the population into 'n' Batches depending on the number of stages using a predetermined partitioning parameter. The Batches are then tested in stages and the obtained results are used in constructing the adaptive model, [12].

1.2 Statement of the problem

Efficiency in Batch testing is a fundamental property in determining the applicability of any constructed estimator. Previous studies on Adaptive Batch testing with errors did not consider the effect of truncation on the efficiency of the estimator. Since truncated probabilities are incomplete, it gives us an ability to estimate and relate it with complete probability. It is against this background that this study presents a Three-Stage Adaptive Batch testing model for estimating prevalence of a trait without truncation with errors with the view to

establishing the effect of truncation on the efficiency of the Estimator.

1.3 Objective of the Study

1.3.1 Main Objective

To develop a Three-stage adaptive batch testing estimator of prevalence of a trait without truncation.

1.3.2 Specific Objectives

The specific objectives of this study are:

(i)To construct a Three-stage adaptive batch testing estimator with test errors.

(ii)To determine the efficiency of the constructed estimator.

(iii)To compare the efficiency of the estimator constructed with that of the truncated estimator.

1.4 Significance of the Study

This study does a comparative analysis between two models of estimation with test errors in order to identify the scheme that is better. The identified scheme would find application in the many situations that require estimation of prevalence such as HIV prevalence, quality control among others.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Batch testing has been widely used as a sampling scheme that is very instrumental in reducing the time and cost of testing especially when the items of interest are rare, [5], [13] and [16]. The scheme has had a rich history, dating back to Dorfman [5] and his seminar work during the world war (II) when they estimated the proportion of diseased individuals among the US soldier.

2.2 Non-Adaptive Batch testing

Dorfman,[5] introduced the statistical and mathematical concepts of batch testing. He used it to estimate the proportion of diseased individuals among the US soldiers by dividing a large population of size say, N into n Batches for the purpose of testing blood samples of diaftles to detect syphilis. He Assumed that if n Batches were subjected to test and X found to test positive and let p be the probability of classifying a batch as positive, then X follows a binomial distribution with parameters n and p expressed as

$$X \sim Binomial(n, p) \tag{2.1}$$

.

In his testing, Dorfman,[5] realised savings of upto 80% in terms of cost and time. Thompson, [15] studied estimation using Batch testing procedure and according to him, if X Batches test positive on the test, then X has a binomial distribution with parameters n and 1-(1-p)^k or simply expressed as

$$X \sim Binomial(n, 1-(1-p)^k) \tag{2.2}$$

where,

$$1-(1-p)^k \tag{2.3}$$

is the probability that a batch tests positive. He used this model to obtain the MLE, $\hat{p}$ of p as

$$\hat{p} = \left[1 - \frac{X}{n}\right]^{\frac{1}{k}} \tag{2.4}$$

Recent studies have seen a large number of scholars engage in Batch testing in different fields such as estimation of HIV/AID's prevalence without necessarily identifying the subject [2],[1],[6],[9], quality control process [16] and phytopathology, [7]. Batch testing need not only be applied to a population in identification of a trait but also on other populations with limited intention of re-testing the items. For example, if a batch of fruit items is being tested for contamination , the interest may not be on identifying the particular items which are defective. The need may instead be on estimating the ratio of the positive items so as to remove them from a given market population.

Brookmeyer, [3] presented a simple stage determination analysis of the general Multi-Stage Batch testing model for estimating disease prevalence. He altered a Single-Stage Batch testing study originally proposed by Thompson,[15] to estimate HIV incidence rates from prevalence studies of early HIV infections by use of a Polymerance Chain Reaction (PCR) assay for HIV Ribonucleic Acid (RNA). He established that if the disease is significantly rare, pooling increases efficiency of the estimator of prevalence unlike the individual testing, because lesser tests are necessary [3].

Nyongesa, [10] introduced the idea of error terms thereby altering the Thompson

model [15]. He obtained the MLE $\hat{p}$, of prevalence p as

$$\hat{p} = 1 - \left[\frac{\eta - \frac{x}{n}}{\eta + \phi - 1}\right]^{\frac{1}{k}} \tag{2.5}$$

and established that the model was more efficient in situations where test kits had low sensitivity and specificity.

Wanyonyi, [16] carried out estimation of proportion of a trait by Batch testing with errors in inspection as applied to quality control process and established that in a given population the probability of detecting a defective Batch is much affected by the cut off value and the total size of the Batch. He illustrated that this probability increases with increase in the total size of the batch at a given value of sensitivity and specificity. However this rate reduces at high cut off value.

2.3 Adaptive Batch Testing Scheme

Oliver-Hughes and Swallow, [13] proposed a Two-Stage Adaptive Batch testing model which they used to estimate small proportions. They used Maximum Likelihood Estimation (MLE) method to estimate the proportion and Crammer-Rao Lower Bound method to obtain the variance of the estimator. They divided the population in two Batches, that is to say λn Batches tested at Stage one and (n-λn) Batches tested at stage two; where λ is a partitioning parameter used to partition the Batches. The MLE's at stages one $\hat{p_1}$ and two $\hat{p_2}$ were established and the results obtained were more impressive compared to the single stage estimator $\hat{p}$ because the estimator $\hat{p_2}$ at stage two, $\hat{p_2}$ (adaptive), was found to be more sufficient and efficient than the single stage estimator (Non-adaptive). They further established that, efficiency increases with increase in stages from one to two.

Okoth, [11] generalised the Oliver-Hughes and Swallow model by introducing error terms. With introduction of the error terms the probabilities that a batch

tests positive at stage one and two were established as,

$$\pi_1(p) = \eta(1-(1-p)^{k_1} + (1-\phi)(1-p)^{k_1} \tag{2.6}$$

and

$$\pi_2(p) = \eta(1-(1-p)^{k_2}) + (1-\phi)(1-p)^{k_2} \tag{2.7}$$

respectively. Utilizing these probabilities the estimator at stage two was established as the solution to

$$\begin{aligned}&\frac{x_1 k_1 q^{k_1}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_1}} + \frac{x_2 k_2(x_1) q^{k_2(x_1)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_2(x_1)}} = \\ &\frac{(n\lambda_1 - x_1)k_1 q^{k_1}[(1-\phi)+\eta]}{1-[\eta-(\eta+(1-\phi))q^{k_1}]} + \frac{(n\lambda_2 - x_2)k_2 q^{k_1(x_1)}[(1-\phi)+\eta]}{1-[\eta-(\eta+(1-\phi))q^{k_2(x_1)}]}\end{aligned} \tag{2.8}$$

and its variance as ($\mathrm{Var}(\hat{p}_A)$),

$$Var(\hat{p_A}) = \frac{\pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))}{R} \tag{2.9}$$

where,

$$\begin{aligned}R &= (\eta+\phi-1)^2[\pi_2(p)(1-\pi_2(p))\lambda N k_1^2(1-p)^{2k_1-2} \\ &+ \pi_1(p)(1-\pi_1(p))\lambda n k_2^2(x_1)(1-p)^{2k_1-2}]\end{aligned}$$

2.4 Truncated Models

Truncated probability distributions are an example of conditional distributions. They arise from restricting the range of a given probability distribution. They are used in many scientific practical statistics, mostly in areas where there is need to give a record or understand occurrences which are bound to result from values that lie below or above a given threshold [8]. The use of processes that

require simulation has gained fame in many different areas of industry. With the high rate of development in computers computation methods are becoming easier [4].

Truncation also happens when a function is expanded and only a specified number of terms considered in the succeeding work. In the study upon which this study is bench marked a truncated model in which a simple binomial expansion procedure is applied, was considered.

Recall that Equation (2.3) is the probability that a batch test positive. To obtain this probability we first need to determine the probability that a batch test negative since this is only the case when all the constituents of a batch are negative. Suppose that in a batch we have k constituent objects each with a probability (1-p) of testing negative, then the probability that the batch test negative will be given by,

$$\prod_{i=1}^{k}(1-p) = (1-p)^k \tag{2.10}$$

Equation (2.10) can be binomially expanded and truncated to the required number of terms as,

$$(1-p)^k = 1 - kp + \frac{1}{2}k(k-1)p^2 + O(P^3) \tag{2.11}$$

Equation (2.11) which is truncated after the term of order two is the truncated probability that a batch tests negative from which truncated models have been constructed [12].

Okoth *et.al* [12] extended this work to a Multi-Stage case. In their development, they truncated the model and established that the adaptive testing scheme is more efficient than the non-adaptive testing scheme with truncation incorporated.

One notable model in their work was the Estimator at stage-three where MLE was found the solution to

$$
\begin{aligned}
& \frac{k_1 X_1(1-pk_1+p)}{\pi_1} \\
+ \; & \frac{k_2(X_1)x_2(1-pk_2(X_1)+p)}{\pi_2} \\
+ \; & \frac{k_3(X_2)X_3(1-pk_2(X_2)+p)}{\pi_3} \\
= \; & \frac{(\lambda_1 n - X_1)k_1(1-pk_1+p)}{(1-\pi_1)} \\
+ \; & \frac{(\lambda_2 n - X_2)k_2(X_1)(1-pk_2(X_1)+p)}{(1-\pi_2)} \\
+ \; & \frac{((1-\lambda_1-\lambda_2)n - X_3)k_3(X_2)(1-pk_3(X_2)+p)}{(1-\pi_3)} \qquad (2.12)
\end{aligned}
$$

and whose variance was found as

$$
Var(\hat{p}_3) = \frac{\pi_3(p)\pi_2(p)\pi_1(p)(1-\pi_2(p))(1-\pi_2(p))(1-\pi_3(p))}{B} \qquad (2.13)
$$

where

$$
\begin{aligned}
B = \; & (\eta+\phi-1)^2[\lambda_1 n\pi_2(p)\pi_3(p)(1-\pi_2(p))1-\pi_3(p)(k_1-pk_1^2+pk_1)^2 \qquad (2.14) \\
+ \; & \lambda_2 n\pi_1(p)\pi_3(p)(1-\pi_1(p))(1-\pi_3(p))(k_2(x_1)-pk_2^2(x_2)+pk_2(x_1))^2 \\
+ \; & (1-\lambda_1-\lambda_2)n\pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))(k_3(x_1)-pk_3^2(x_1)+pk_3(x_1))^2]
\end{aligned}
$$

The truncation in Equation (2.11) can be solved by use of Equation (2.10) which is complete. Three-Stage adaptive estimator of prevalence of a trait without truncation is developed in this study and studies the effect of truncation by comparing with Okoth *et.al* [12] model at stage three where truncation was done.

CHAPTER THREE

MODEL FORMULATION

3.1 Introduction

In this chapter a three stage adaptive estimator $\hat{p}_B$ without truncation by incorporating errors is constructed. Further, its efficiency relative to the three stage adaptive estimator with truncation is determined. The chapter is outlined as follows;

A general overview of Adaptive batch testing is presented in Section 3.2 while Three-stage Adaptive batch testing is discussed in Section 3.3 and it's asymptotic variance in Section 3.4

3.2 Adaptive batch testing

Adaptive batch testing involves testing batches in stages after partitioning the population. The batch size is updated from one stage to the next.

We note at this point that the batch size at any given stage depends on the results obtained at previous stages. The Batches are tested to establish defective Batches across the stages. The outcome obtained at a stage depends on the outcome at a previous stage. The following methods were used in the current study.

3.2.1 Method of Maximum likelihood Estimation

The estimator of the prevalence trait is obtained by use of Maximum Likelihood Estimation (MLE) method. The likelihood function of random variable X_j where j=1,2,...n is defined to be the joint density of the n random variables which is considered to be a function of θ and denoted by L(θ).

i.e

$$L(\theta) = f(X_j, \theta) = \prod_{i=1}^{n} f(X_j, \theta)$$

$$L(\theta) = f(\underline{X}, \theta) = \prod_{i=1}^{n} f(X_j, \theta)$$

The maximum likelihood estimate of θ is the value of θ that maximises the likelihood function L(θ),thus

$$\frac{\partial}{\partial\theta} L(\theta) = 0 \tag{3.1}$$

to which is equivalent to maximising the log likelihood ,i.e

$$\frac{\partial}{\partial\theta} logL(\theta) = 0 \tag{3.2}$$

.

3.2.2 Cramer-Rao Lower Bound (CRLB) Method

The CRLB method is used a method for finding the lower bound of the variance of an unbiased estimator of prevalence trait constructed estimator. Suppose X_j is random variable from a given population having density function f(X_j,θ)with respect to the measure μ where θ ϵ $\bigodot$. Let T=t(X_j) be an unbiased estimator for θ under the regularity conditions :

(i) θ lies in the open interval $\bigodot$ of the real line.

(ii)$\frac{\partial}{\partial\theta} f(X_j, \theta)$ exist $\forall$ $\theta\varepsilon$ $\bigodot$, $\forall$ x

(iii)$\int T(\underline{X}) \prod_{j=1}^{n} \partial\mu$ can be differentiated under the integral sign for any $T(\underline{X})$

(iv) $0 < E[\frac{\partial}{\partial\theta} \ln f(X_j, \theta)]^2 < \infty \ \forall \ \theta \epsilon \bigodot$ and $\exists \forall \theta$ if the above assumptions hold, then

$$Var(T) \geq \frac{[\tau'(\theta)]^2}{nE_\theta[\frac{\partial}{\delta\theta} \ln f(X_j, \theta)]^2} \tag{3.3}$$

if T is unbiased then the numerator is 1

$\therefore$ under the assumptions, if $E_\theta(T) = \theta + b(\theta)$, then

$$Var_\theta(T) \geq \frac{(1 + b'(\theta))^2}{nE[\frac{\partial}{\delta\theta} \log f(X_j, \theta)]^2} \tag{3.4}$$

T is unbiased

$$\therefore Var_\theta(T) \geq \frac{1}{nE[\frac{\partial}{\delta\theta} \log f(X_j, \theta)]^2} \tag{3.5}$$

$$\therefore Var_\theta(T) \geq \frac{1}{I(\theta)} \tag{3.6}$$

obtaind from the fishers information

Equation (3.6) is Cramer-Rao inequality and the right hand side of this inequality is known as Cramer-Rao Lower Bound for the variance of the unbiased estimators. Numerical simulations of this model were carried out in R to graphically illustrate the behaviour of the solutions of the estimated model.The efficiency of the estimator relative to the adaptive estimator with truncation shall be determined with the view to performing a comparative analysis between the two.

3.3 Three-stage adaptive batch testing model

In this scheme batches are partitioned into three and tested at three stages. Suppose we define N as the size of population which is partitioned into n homogeneous batches each of which of size k. In this case $\lambda_1 n$, $\lambda_2 n$ and $\lambda_3 n$ batches of sizes each k_1 ,k_2 and k_3 are tested at stages one, two and three respectively. The diagrammatic representation of this model is as shown in Figure 3.1

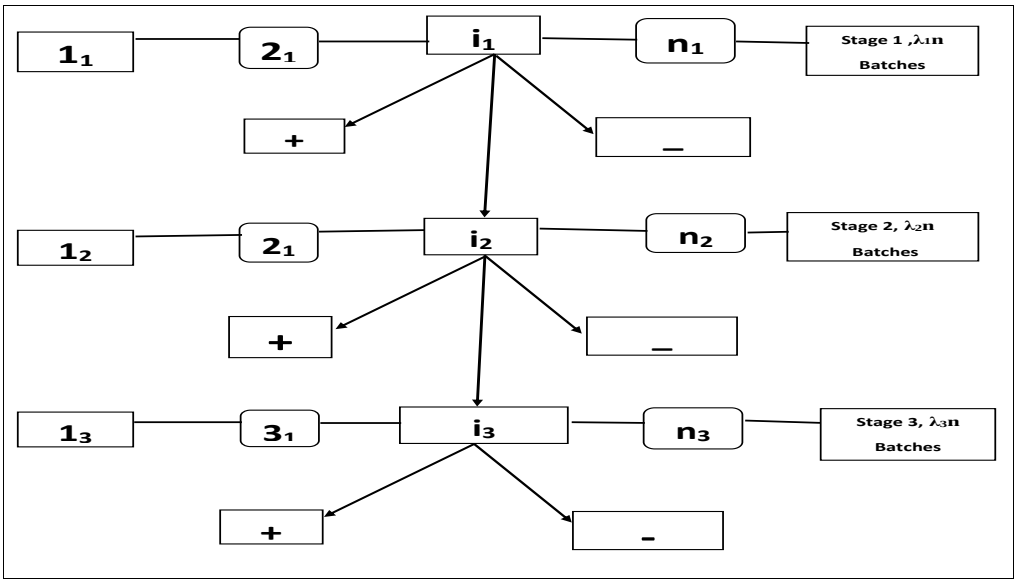

Figure 3.1: Schematic representation of Three stage Adaptive Batch Testing Model

From Figure 3.1, when a batch is tested, it either yields positive or negative results; with the probability that a batch tests positive as $1-(1-p)^k$ and $(1-p)^k$ as the probability that a batch tests negative [11].

The batch size at this stage is k_3 and it depends on k_2 and k_1. This is determined as ,

$$k_3 = argmin(Var\hat{p}_2) \tag{3.7}$$

If X_3 is the number of defective batches at stage 3, conditioned on X_1 and X_2 then X_3 follows a Binomial distribution with parameters $\lambda_3 n$ and π_3 simply written as

$$X_3|X_1X_2 \sim Bi(\lambda_3 n, \pi_3) \tag{3.8}$$

where π_3 is the probability that a batch tests positive at stage three and is given by,

$$\pi_3 = \eta[1-(1-p)^{k_3(x_2)}] + (1-\phi)(1-p)^{k_3(x_2)} \tag{3.9}$$

The final Three stage adaptive estimator of p, $\hat{p}_B$ is the MLE based on the joint distribution of X_1,X_2 and X_3.

Here η and ϕ are the error terms referred to us sensitivity and specificity of the test kit respectively. In this case sensitivity is the probability of a positive batch truly tests as positive and sensitivity is the probability of a negative batch truly tests as negative.

This joint distribution is given by,

$$\begin{aligned} f(X_3,X_2,X_1) &= Bin(\lambda_1 n, \eta[1-(1-p)^{k_1}] + (1-\phi)(1-p)^{k_1}) \\ &\times Bin(\lambda_2 n, \eta[1-(1-p)^{k_1(x_1)}] + (1-\phi)(1-p)^{k_2(x_1)}) \\ &\times Bin(\lambda_3 n, \eta[1-(1-p)^{k_3(x_2)}] + (1-\phi)(1-p)^{k_3(x_2)}) \end{aligned} \tag{3.10}$$

To find MLE we find the likelihood function of equation (3.10) which is given by

$$\begin{aligned} f(X_3,X_2,X_1) &= \begin{pmatrix} n\lambda_1 \\ x_1 \end{pmatrix} [\eta(1-(1-p)^{k_1}) + (1-\phi)(1-p)^{k_1}]^{x_1} \\ &\times [1-(\eta-\eta(1-p)^{k_1} + (1-\phi)(1-p)^{k_1})]^{n\lambda_1 - x_1} \\ &\times \begin{pmatrix} n\lambda_2 \\ x_2 \end{pmatrix} [\eta(1-(1-p)^{k_2(x_1)}) + 1 - \phi)(1-p)^{k_2(x_1)}]^{x_2} \\ &\times [1-(\eta-\eta(1-p)^{k_2(x_1)}) + (1-\phi)(1-p)^{k_2(x_1)})]^{n\lambda_2 - x_2} \\ &\times \begin{pmatrix} n\lambda_3 \\ x_3 \end{pmatrix} [\eta(1-(1-p)^{k_3(x_2)}) + 1 - \phi)(1-p)^{k_3(x_2)}]^{x_3} \\ &\times [1-(\eta-\eta(1-p)^{k_3(x_2)}) + (1-\phi)(1-p)^{k_3(x_2)})]^{n\lambda_3 - x_3} \end{aligned} \tag{3.11}$$

Introducing the constant of proportionality, Equation (3.11) becomes

$$
\begin{aligned}
f(X_3, X_2, X_1) &\propto [\eta(1-(1-p)^{k_1}) + (1-\phi)(1-p)^{k_1}]^{x_1} \\
&\times [1-(\eta-\eta(1-p)^{k_1} + (1-\phi)(1-p)^{k_1})]^{n\lambda_1 - x_1} \\
&\times [\eta(1-(1-p)^{k_2(x_1)}) + (1-\phi)(1-p)^{k_2(x_1)}]^{x_2} \qquad (3.12) \\
&\times [1-(\eta-\eta(1-p)^{k_2(x_1)} + (1-\phi)(1-p)^{k_2(x_1)})]^{n\lambda_2 - x_2} \\
&\times [\eta(1-(1-p)^{k_3(x_2)}) + (1-\phi)(1-p)^{k_3(x_2)}]^{x_3} \\
&\times [1-(\eta-\eta(1-p)^{k_3(x_2)} + (1-\phi)(1-p)^{k_3(x_2)})]^{n\lambda_3 - x_3}
\end{aligned}
$$

Next we find the natural logs

$$
\begin{aligned}
Inf(X_3, X_2, X_1) &\propto x_1 In[\eta - \eta(1-p)^{k_1}) + (1-\phi)(1-p)^{k_1}] \\
&+ (n\lambda_1 - x_1) In[1-[\eta - \eta(1-p)^{k_1} + (1-\phi)(1-p)^{k_1}]] \\
&+ x_2 In[\eta - \eta(1-p)^{k_2(x_1)}) + (1-\phi)(1-p)^{k_2(x_1)}] \qquad (3.13) \\
&+ (n\lambda_2 - x_2) In[1-[\eta - \eta(1-p)^{k_2(x_1)} + (1-\phi)(1-p)^{k(x_1)}]] \\
&+ x_3 In[\eta - \eta(1-p)^{k_3(x_2)}) + (1-\phi)(1-p)^{k_3(x_2)}] \\
&+ (n\lambda_3 - x_3) In[1-[\eta - \eta(1-p)^{k_3(x_2)} + (1-\phi)(1-p)^{k_3(x_2)}]]
\end{aligned}
$$

The derivative with respect to p of Equation (3.13) is obtained as

$$
\begin{aligned}
\frac{\partial}{\partial p} f(X_3, X_2, X_1) &= \frac{x_1[\eta k_1(1-p)^{k_1-1} + k_1(1-\phi)(1-p)^{k_1-1}]}{\eta - \eta(1-p)^{k_1} + (1-\phi)(1-p)^{k_1}} \\
&- \frac{(n\lambda_1 - x_1)[\eta k_1(1-p)^{k_1-1} + k_1(1-\phi)(1-p)^{k_1-1}]}{1-[\eta - \eta(1-p)_1^k + (1-\phi)(1-p)^{k_1}]} \\
&+ \frac{x_2[\eta k_2(x_1)(1-p)^{k_2(x_1)-1} + k_2(1-\phi)(1-p)^{k_2(x_1)-1}]}{\eta - \eta(1-p)^{k_2(x_1)} + (1-\phi)(1-p)^{k_2(x_1)}} \qquad (3.14) \\
&- \frac{(n\lambda_2 - x_2)[k_2\eta(1-p)^{k_2(x_1)-1} + k_2(1-\phi)(1-p)^{k_2(x_1)}]}{1-[\eta - \eta(1-p)^{k_2(x_1)} + (1-\phi)(1-p)^{k_2(x_1)}]} \\
&+ \frac{x_3[\eta k_3(x_2)(1-p)^{k_3(x_2)-1} + k_3(1-\phi)(1-p)^{k_3(x_2)-1}]}{\eta - \eta(1-p)^{k_3(x_2)} + (1-\phi)(1-p)^{k_3(x_2)}} \\
&- \frac{(n\lambda_3 - x_3)[k_3\eta(1-p)^{k_3(x_2)-1} + k_3(1-\phi)(1-p)^{k_3(x_2)}]}{1-[\eta - \eta(1-p)^{k_3(x_2)} + (1-\phi)(1-p)^{k_3(x_1)}]}
\end{aligned}
$$

The MLE of p_B, $\hat{p}_B$, is given by the solution to Equation(3.15)

$$\begin{aligned}
& \frac{x_1 k_1(1-p)^{k_1}[(1-\phi)-\eta]}{\eta-\eta(1-p)^{k_1}+(1-\phi)(1-p)^{k_1}} + \frac{x_2 k_1(x_1)(1-p)^{k_2(x_1)}[(1-\phi)-\eta]}{\eta-\eta(1-p)^{k_2(x_1)}+(1-\phi)(1-p)^{k_2(x_1)}} \\
+ & \frac{x_3 k_3(x_1)(1-p)^{k_3(x_1)}[(1-\phi)-\eta]}{\eta-\eta(1-p)^{k_3(x_1)}+(1-\phi)(1-p)^{k_3(x_1)}} \\
= & \frac{(n\lambda_1-x_1)k_1(1-p)^{k_1}[(1-\phi)+\eta]}{1-[\eta-\eta(1-p)^{k_1}+(1-\phi)(1-p)^{k_1}]} + \frac{(n\lambda_2-x_2)k_2(1-p)^{k_1(x_1)}[(1-\phi)+\eta]}{1-[\eta-\eta(1-p)^{k_2(x_1)}+(1-\phi)(1-p)^{k_2(x_1)}]} \\
+ & \frac{(n\lambda_3-x_3)k_3(1-p)^{k_3(x_2)}[(1-\phi)+\eta]}{1-[\eta-\eta(1-p)^{k_3(x_2)}+(1-\phi)(1-p)^{k_3(x_2)}]} \qquad (3.15)
\end{aligned}$$

Next we establish the asymptotic variance of p_B, $\text{var}(\hat{p}_B)$ which does not require that we find $\hat{p}_B$

3.4 Asymptotic variance of $\hat{p}_B$

Asymptotic variance is useful in determining Asymptotic Relative Efficiency (ARE) of an estimator. To determine variance, we recall Equation (3.12) and let

$$\pi_1(p) = [\eta(1-(1-p)^{k_1}) + (1-\phi)(1-p)^{k_1}] \qquad (3.16)$$

$$\pi_2(p) = \eta(1-(1-p)^{k_2(x_1)}) + (1-\phi)(1-p)^{k_2(x_1)} \qquad (3.17)$$

$$\pi_3(p) = \eta(1-(1-p)^{k_3(x_2)}) + (1-\phi)(1-p)^{k_3(x_2)} \qquad (3.18)$$

$$(1-\pi_1(p)) = 1-[\eta-\eta(1-p)^{k_1}+(1-\phi)(1-p)^{k_1}] \qquad (3.19)$$

$$(1-\pi_2(p)) = 1-[\eta-\eta(1-p)^{k_2(x_1)}+(1-\phi)(1-p)^{k_2(x_1)}] \qquad (3.20)$$

$$(1-\pi_3(p)) = 1-[\eta-\eta(1-p)^{k_3(x_1)}+(1-\phi)(1-p)^{k_3(x_1)}] \qquad (3.21)$$

The first and second derivatives of Equations (3.16), (3.17), (3.18), (3.19), (3.20) and (3.21) are given as

$$\frac{\partial}{\partial p}\pi_1(p) = \eta k_1(1-p)^{k_1-1} - k_1(1-\phi)(1-p)^{k_1-1} \tag{3.22}$$

$$\frac{\partial^2}{\partial p^2}\pi_1(p) = k_1(k_1-1)(1-\phi)(1-p)^{k_1-2} - \eta k_1(k_1-1)(1-p)^{k_1-2} \tag{3.23}$$

$$\frac{\partial}{\partial p}(1-\pi_1(p)) = k_1(1-\phi)(1-p)^{k_1-1} - \eta k_1(1-p)^{k_1-1} \tag{3.24}$$

$$\frac{\partial^2}{\partial p^2}(1-\pi_1(p)) = \eta k_1(k_1-1)(1-p)^{k_1-2} - k_1(k_1-1)(1-\phi)(1-p)^{k_1-2} \tag{3.25}$$

$$\frac{\partial}{\partial p}\pi_2(p) = \eta k_2(x_1)(1-p)^{k_2(x_1)-1} - k_2(x_1)(1-\phi)(1-p)^{k_2(x_1)-1} \tag{3.26}$$

$$\frac{\partial^2}{\partial p^2}\pi_2(p) = k_2(x_1)(k_2(x_1)-1)(1-\phi)(1-p)^{k_2(x_1)-2} - \eta k_2(x_1)(k_2(x_1)-1)(1-p)^{k_2(x_1)-2} \tag{3.27}$$

$$\frac{\partial}{\partial p}(1-\pi_2(p)) = k_2(x_1)(1-\phi)(1-p)^{k_2(x_1)-1} - \eta k_2(x_1)(1-p)^{k_2(x_1)-1} \tag{3.28}$$

$$\frac{\partial^2}{\partial p^2}(1-\pi_2(p)) = \eta k_2(x_1)(k_2(x_1)-1)(1-p)^{k_2(x_1)-2} - k_2(x_1)(k_2(x_1)-1)(1-\phi)(1-p)^{k_2(x_1)-2} \tag{3.29}$$

$$\frac{\partial}{\partial p}\pi_3 p) = \eta k_3(x_2)(1-p)^{k_3(x_2)-1} - k_3(x_2)(1-\phi)(1-p)^{k_3(x_2)-1} \tag{3.30}$$

$$\frac{\partial^2}{\partial p^2}\pi_3(p) = k_3(x_2)(k_3(x_2)-1)(1-\phi)(1-p)^{k_3(x_2)-2} - \eta k_3(x_2)(k_3(x_2)-1)(1-p)^{k_3(x_2)-2} \tag{3.31}$$

$$\frac{\partial}{\partial p}(1-\pi_3(p)) \;=\; k_3(x_2)(1-\phi)(1-p)^{k_3(x_2)-2} - \eta k_3(x_2)(1-p)^{k_3(x_1)-1} \tag{3.32}$$

$$\frac{\partial^2}{\partial p^2}(1-\pi_3(p)) \;=\; \eta k_3(x_2)(k_3(x_1)-1)(1-p)^{k_3(x_2)-2} - k_3(x_2)(k_3(x_2)-1)(1-\phi)(1-p)^{k_3(x_2)-2} \tag{3.33}$$

Substituting Equations (3.16), (3.17) and (3.18), in Equation (3.12) we obtain,

$$f(X_3,X_2,X_1) \propto \pi_1(p)^{(x_1)}(1-\pi_1(p))^{n\lambda_1-x_1}\pi_2(P)^{x_2}(1-\pi_2(p))^{n\lambda_2-x_2}\pi_3(P)^{x_3}(1-\pi_3(p))^{n\lambda_3-x_3} \tag{3.34}$$

Next we find the natural log on Equation (3.34)

$$\begin{aligned} lnf(X_3,X_2,X_1) \;&\propto\; (x_1)ln\pi_1(p) + (n\lambda_1-x_1)ln(1-\pi_1(p)) \\ &+\; x_2 ln\pi_2(p) + (n\lambda_2-x_2)ln(1-\pi_2(p)) \\ &+\; x_3 ln\pi_3(p) + (n\lambda_3-x_3)ln(1-\pi_3(p)) \end{aligned} \tag{3.35}$$

The first and second derivatives of Equation (3.35) with respect to p are obtained as,

$$\begin{aligned} \frac{\partial}{\partial p}lnf(x_3,x_2,x_1) \;&=\; \frac{x_1}{\pi_1(p)}\frac{\partial}{\partial p}\pi_1(p) + \frac{(n\lambda_1-x_1)}{(1-\pi_1(p))}\frac{\partial}{\partial p}(1-\pi_1(p)) \\ &+\; \frac{x_2}{\pi_2(p)}\frac{\partial}{\partial p}\pi_2(p) + \frac{(n\lambda_2-x_2)}{(1-\pi_2(p))}\frac{\partial}{\partial p}(1-\Pi_2(p)) \\ &+\; \frac{x_3}{\pi_3(p)}\frac{\partial}{\partial p}\pi_3(p) + \frac{(n\lambda_3-x_3)}{(1-\pi_3(p))}\frac{\partial}{\partial p}(1-\pi_3(p)) \end{aligned} \tag{3.36}$$

and

$$
\begin{aligned}
[\frac{\partial^2}{\partial p^2} lnf(X_3, X_2, X_1)] &= [x_1[\frac{\pi_1(p)\frac{\partial^2}{\partial p^2}\pi_1(p) - \frac{\partial}{\partial p}\pi_1(p)\frac{\partial}{\partial p}\pi_1(p)}{\pi_1^2(p)}] \\
&+ (n\lambda_1 - x_1)[\frac{(1-\pi_1(p))\frac{\partial^2}{\partial p^2}(1-\pi_1(p)) - \frac{\partial}{\partial p}(1-\pi_1(p))\frac{\partial}{\partial p}(1-\pi_1(p))}{(1-\pi_1(p))^2}] \\
&+ x_2[\frac{\pi_2(p)\frac{\partial^2}{\partial p^2}\pi_2(p) - \frac{\partial}{\partial p}\pi_2(p)\frac{\partial}{\partial p}\pi_2(p)}{\pi_2^2(p)}] \\
&+ (n\lambda_2 - x_2)[\frac{(1-\pi_2(p))\frac{\partial^2}{\partial p^2}(1-\pi_2(p)) - \frac{\partial}{\partial p}(1-\pi_2(p))\frac{\partial}{\partial p}(1-\pi_2(p))}{(1-\pi_2(p))^2}] \\
&+ x_3[\frac{\pi_3(p)\frac{\partial^2}{\partial p^2}\pi_3(p) - \frac{\partial}{\partial p}\pi_3(p)\frac{\partial}{\partial p}\pi_3(p)}{\pi_3^2(p)}] \\
&+ (n\lambda_3 - x_3)[\frac{(1-\pi_3(p))\frac{\partial^2}{\partial p^2}(1-\pi_3(p)) - \frac{\partial}{\partial p}(1-\pi_3(p))\frac{\partial}{\partial p}(1-\pi_3(p))}{(1-\pi_3(p))^2}]]
\end{aligned}
\tag{3.37}
$$

respectively.

To find the expectation of X_1, X_2, X_3 in Equation (3.37), we note that for a binomial distribution, E(x)=np. In this case $E(x_1) = n\lambda_1\pi_1(p), E(x_2) = n\lambda_2\pi_2(p)$ and $E(x_3) = n\lambda_3\pi_3(p)$

Substituting for $E(x_1), E(x_2) and E(x_3)$ in Equation (3.37), factoring and simplifying we obtain

$$
\begin{aligned}
E[\frac{\partial^2}{\partial p^2} lnf(X_3, X_2, X_1] &= (n\lambda_1)[\frac{\pi_1(p)\frac{\partial^2}{\partial p^2}\pi_1(p) - \frac{\partial}{\partial p}\pi_1(p)\frac{\partial}{\partial p}\pi_1(p)}{\pi_1(p)}] \\
&+ [\frac{(1-\pi_1(p))\frac{\partial^2}{\partial p^2}(1-\pi_1(p)) - \frac{\partial}{\partial p}(1-\Pi_1(p))\frac{\partial}{\partial p}(1-\pi_1(p))}{(1-\pi_1(p))}] \\
&+ (n\lambda_2)[\frac{\pi_2(p)\frac{\partial^2}{\partial p^2}\pi_2(p) - \frac{\partial}{\partial p}\pi_2(p)\frac{\partial}{\partial p}\pi_2(p)}{\pi_2(p)}] \\
&+ [\frac{(1-\pi_2(p))\frac{\partial^2}{\partial p^2}(1-\pi_2(p)) - \frac{\partial}{\partial p}(1-\Pi_2(p))\frac{\partial}{\partial p}(1-\pi_2(p))}{(1-\pi_2(p))}] \\
&+ (n\lambda_3)[\frac{\pi_3(p)\frac{\partial^2}{\partial p^2}\pi_3(p) - \frac{\partial}{\partial p}\pi_3(p)\frac{\partial}{\partial p}\pi_3(p)}{\pi_3(p)}] \\
&+ [\frac{(1-\pi_3(p))\frac{\partial^2}{\partial p^2}(1-\pi_3(p)) - \frac{\partial}{\partial p}(1-\pi_3(p))\frac{\partial}{\partial p}(1-\pi_3(p))}{(1-\pi_3(p))}]
\end{aligned}
$$

(3.38)

Next we substitute for $\frac{\partial^2}{\partial p^2}\pi_2(p)$, $\frac{\partial^2}{\partial p^2}(1-\pi_1(p))$, $\frac{\partial^2}{\partial p^2}\pi_2(p)$, $\frac{\partial^2}{\partial p^2}(1-\pi_2(p))$, $\frac{\partial^2}{\partial p^2}\pi_3(p)$, and $\frac{\partial^2}{\partial p^2}(1-\pi_3(p))$ in Equation (3.38) and obtain

$$\begin{aligned}
E[\frac{\partial^2}{\partial p^2}lnf(x_3,x_2,x_1)] &= -n\lambda_1 k_1(1-p)^{2k_1-2}[(\frac{\eta^2 k_1 - 2\eta l_1(1-\phi)+k_1(1-\phi)^2}{[\eta-\eta(1-p)^{k_1}+(1-\phi)(1-p)^{k_1}]} \\
&+ \frac{[k_1(1-\phi)^2-2\eta k_1(1-\phi)-\eta^2 k_1]}{1-[\eta-\eta(1-p)^{k_1}+(1-\phi)(1-p)^{k_1}]}] \\
&- n\lambda_2 k_2(x_1)(1-p)^{2k_2(x_1)-2}[(\frac{\eta^2 k_2(x_1)-2\eta k_2(x_1)(1-\phi)+k_2(x_1)(1-\phi)^2}{[\eta-\eta(1-p)^{k_2(x_1)}+(1-\phi)(1-p)^{k_2(x_1)}]} \\
&+ (\frac{[k_2(x_1)(1-\phi)^2-2\eta k_2(x_1)(1-\phi)+\eta^2 k_2(x_1)]}{1-[\eta-\eta(1-p)^{k_2(x_1)}+(1-\phi)(1-p)^{k_2(x_1)}]})] \\
&- n\lambda_3 k_3(x_1)(1-p)^{2k_3(x_2)-2}[(\frac{\eta^2 k_3(x_2)-2\eta k_3(x_1)(1-\phi)+k_3(x_2)(1-\phi)^2}{[\eta-\eta(1-p)^{k_3(x_2)}+(1-\phi)(1-p)^{k_3(x_1)}]} \\
&+ (\frac{[k_3(x_2)(1-\phi)^2-2\eta k_3(x_2)(1-\phi)+\eta^2 k_3(x_2)]}{1-[\eta-\eta(1-p)^{k_3(x_2)}+(1-\phi)(1-p)^{k_3(x_1)}]})] \qquad (3.39
\end{aligned}$$

Which reduces to:

$$\begin{aligned}
E[\frac{\partial^2}{\partial p^2}lnf(X_3,X_2,X_1)] &= -n\lambda_1 k_1(1-p)^{2k_1-2}[(\frac{\eta^2 k_1-2\eta k_1(1-\phi)+k_1(1-\phi)^2}{\pi_1(p)} \\
&+ \frac{[k_1(1-\phi)^2-2\eta k_1(1-\phi)-\eta^2 k_1]}{1-\pi_1(p)}] \\
&+ -n\lambda_2 k_2(x_1)(1-p)^{2k_2(x_1)-2} \\
&\times [(\frac{\eta^2 k_2(x_1)-2\eta k_2(x_1)(1-\phi)+k_2(x_1)(1-\phi)^2}{\pi_2(p)}) \\
&+ (\frac{[k_2(x_1)(1-\phi)^2-2\eta k_2(x_1)(1-\phi)+\eta^2 k_2(x_1)]}{1-\pi_2(p)]})] \\
&+ -n\lambda_3 k_3(x_1)(1-p)^{2k_3(x_2)-2} \\
&\times [(\frac{\eta^2 k_3(x_2)-2\eta k_3(x_2)(1-\phi)+k_3(x_2)(1-\phi)^2}{\pi_3(p)}) \\
&+ (\frac{[k_3(x_2)(1-\phi)^2-2\eta k_3(x_2)(1-\phi)+\eta^2 k_3(x_2)]}{1-\pi_3(p)})] \qquad (3.40)
\end{aligned}$$

Factorising and rearranging Equation (3.40) we obtain ,

$$\begin{aligned} E[\frac{\partial^2}{\partial p^2} lnf(X_3, X_2, X_1)] &= -n\lambda_1 k_1^2 (1-p)^{2k_1-2}[\frac{\eta^2 - 2\eta(1-\phi) + (1-\phi)^2}{\pi_1(p)(1-\pi_1(p))}] \\ &+ -n\lambda_2 k_2^2(x_1)(1-p)^{2k_2(x_1)-2}[\frac{\eta^2 - 2\eta(1-\phi) + (1-\phi)^2}{\pi_2(p)(1-\pi_2(p))}] \\ &+ -n\lambda_3 k_3^2(x_2)(1-p)^{2k_3(x_2)-2}[\frac{\eta^2 - 2\eta(1-\phi) + (1-\phi)^2}{\pi_3(p)(1-\pi_3(p))}] \end{aligned} \tag{3.41}$$

Now

$$\begin{aligned} E[\frac{\partial^2}{\partial p^2} lnf(X_3, X_2, X_1)] &= -\frac{n\lambda_1 k_1^2 (1-p)^{2k_1-2}(\eta+\phi-1))^2}{\pi_1(p)(1-\pi_1(p))} \\ &+ -\frac{n\lambda_2 k_2^2(x_1)(1-p)^{2k_2(x_1)-2}(\eta+\phi-1)^2}{\pi_2(p)(1-\pi_2(p))} \\ &+ -\frac{n\lambda_3 k_3^2(x_2)(1-p)^{2k_3(x_2)-2}(\eta+\phi-1)^2}{\pi_3(p)(1-\pi_3(p))} \end{aligned} \tag{3.42}$$

Therefore,

$$\begin{aligned} -E[\frac{\partial^2}{\partial p^2} lnf(X_3, X_2, X_1)] &= (\eta+\phi-1))^2[\frac{n\lambda_1 k_1^2 (1-p)^{2k_1-2}}{\pi_1(p)(1-\pi_1(p))} + \frac{n\alpha_2 k_2^2(x_1)(1-p)^{2k_2(x_1)-2}}{\pi_2(p)(1-\pi_2(p))} \\ &+ \frac{n\lambda_3 k_3^2(x_2)(1-p)^{2k_3(x_1)-2}}{\pi_3(p)(1-\pi_3(p))}] \end{aligned} \tag{3.43}$$

Writing Equation (3.43) as a single fraction we have

$$-E[\frac{\partial^2}{\partial p^2} lnf(X_3, X_2, X_1)] = \frac{A}{\pi_1(p)\pi_2(p)\pi_3(p)(1-\pi_1(p))(1-p_2(p))(1-\pi_3(p))}] \tag{3.44}$$

Where

$$\begin{aligned} A &= (\eta+\phi-1))^2[\pi_2(p)\pi_3(p)(1-\pi_2(p))(1-\pi_3(p))n\lambda_1 k_1^2(1-p)^{2k_1-2} \\ &+ \pi_1(p)\pi_3(p)(1-\pi_1(p))(1-\pi_3(p))n\lambda_2 k_2^2(x_1)(1-p)^{2k_2(x_1)-2} \\ &+ \pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))n\lambda_3 k_3^2(x_3)(1-p)^{2k_3(x_2)-2} \end{aligned} \tag{3.45}$$

$$Var(\hat{p_B}) = \frac{1}{I_N} = [\frac{1}{-E[\frac{\partial^2}{\partial p^2} lnf(X_3, X_2, X_1)]}] \tag{3.46}$$

$$Var(\hat{p_B}) = \frac{\pi_1(p)\pi_2(p)\pi_3(p)(1-\pi_1(p))(1-\pi_2(p))(1-\pi_3(p))}{A} \tag{3.47}$$

Equation (3.44) is fisher information I_N

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

In this chapter the results of Asymptotic Relative Efficiency (ARE) generated by R software tool are discussed. The highlights of this study will enable us to make necessary conclusion and recommendations to this study.

4.2 Asymptotic Relative Efficiency (ARE)

In this section results on ARE of the three-stage adaptive Batch Testing Estimator without truncation with errors relative to the three-stage Adaptive batch testing Estimator with truncation with error is presented.

The computation of ARE was accomplished by dividing Equations (2.13) by (3.47) given by

$$\frac{Var(\hat{p}_3)}{Var(\hat{p}_B)}$$

Upon simplification we obtain

$$ARE = \frac{A}{B} \tag{4.1}$$

where ,

$$\begin{aligned} A &= (\eta+\phi-1))^2[\pi_2(p)\pi_3(p)(1-\pi_2(p))(1-\pi_3(p))n\lambda_1 k_1^2(1-p)^{2k_1-2} \quad (4.2)\\ &+ \pi_1(p)\pi_3(p)(1-\pi_1(p))(1-\pi_3(p))n\lambda_2 k_2^2(x_1)(1-p)^{2k_2(x_1)-2}\\ &+ \pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))n\lambda_3 k_3^2(x_3)(1-p)^{2k_3(x_2)-2} \end{aligned}$$

and

$$
\begin{aligned}
B &= (\eta+\phi-1)^2[\lambda_1 n\pi_2(p)\pi_3(p)(1-\pi_2(p))1-\pi_3(p)(k_1-pk_1^2+pk_1)^2 \qquad (4.3)\\
&+ \lambda_2 n\pi_1(p)\pi_3(p)(1-\pi_1(p))(1-\pi_3(p))(k_2(x_1)-pk_2^2(x_2)+pk_2(x_1))^2\\
&+ (1-\lambda_1-\lambda_2)n\pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))(k_3(x_1)-pk_3^2(x_1)+pk_3(x_1))^2]
\end{aligned}
$$

Utilizing Equation (4.1) in R, Tables 4.1, 4.2 and 4.3 were obtained.

Table 4.1: ARE values of $\hat{p}_B$ relative to $\hat{p}_3$ when η=ϕ at specified values of p

p	$\eta=\phi=0.99$	$\eta=\phi=0.95$	$\eta=\phi=0.90$	$\eta=\phi=0.85$	$\eta=\phi=0.80$
0.10	1.5383	1.4764	1.4999	1.4864	1.4764
0.14	2.4339	2.3424	2.3811	2.3596	2.3424
0.18	6.5765	6.0859	6.3029	6.1842	6.0859
0.20	12.3877	10.9777	11.5792	11.2462	10.9777
0.22	7.9072	7.9582	7.95312	7.9451	7.9582
0.24	2.5095	2.8015	2.6657	2.7385	2.8015
0.28	0.4102	0.4984	0.4585	0.4801	0.4984
0.30	0.2025	0.2564	0.23254	0.2455	0.2564

Table 4.1 provides ARE values at different values of p when $\eta=\phi$. From the table it is evident that ARE values are all greater than one across the various values of p at low prevalence rate, (i.e p $<$ 0.28). This means that the estimator without truncation is more efficient than the Truncated Estimator. ARE values increase between 0.10 and 0.20 and then they start to decrease at p $>$ 0.20. This implies that the non-truncated estimator performs better at lower values of p. This is a good observation for this model because batch testing is done on population with low prevalence rates.

Table 4.2: ARE values of $\hat{p}_B$ relative to $\hat{p}_3$ when η is constant at 99% and ϕ varies

p	$\eta = 0.99$, $\phi = 0.80$	$\eta = 0.99$, $\phi = 0.85$	$\eta = 0.99$, $\phi = 0.90$	$\eta = 0.99$, $\phi = 0.95$	$\eta = 0.99$, $\phi = 0.98$
0.10	1.5107	1.527	1.5489	1.5797	1.6167
0.14	2.3322	2.343	2.4356	2.4907	2.5545
0.18	6.0268	6.1926	6.4060	6.6887	7.02319
0.20	11.2352	11.8175	12.5899	13.6564	15.0480
0.22	8.5861	8.9824	9.4850	10.1411	11.07124
0.24	2.8993	2.9609	3.0329	3.1179	3.2582
0.28	0.4718	0.4758	0.4803	0.4852	0.4996
0.30	0.2323	0.23413	0.2358	0.23766	0.2451

Table 4.2 provides ARE values at different values of p, when η is held constant and ϕ varies. From the table all the ARE values are greater than one at low prevalence rates (p<0.28). The efficiency increases as the values of ϕ increase. This implies that in the presence of test errors non-truncated estimator are more efficient than truncated estimator with increase in specificity of the test kits.

Table 4.3: ARE values of $\hat{p}_B$ relative to $\hat{p}_3$ when ϕ is constant at 99% and η varies

p	$\eta = 0.80$, $\phi = 0.99$	$\eta = 0.85$, $\phi = 0.99$	$\eta = 0.90$ $\phi = 0.99$	$\eta = 0.95$ $\phi = 0.99$	$\eta = 0.98$, $\phi = 0.99$
0.10	1.5107	1.5270	1.5488	1.5796	1.6049
0.14	2.3623	2.3942	2.4357	2.4907	2.5330
0.18	6.0267	6.1924	6.4060	6.6887	6.9056
0.20	11.2353	11.81755	12.5899	13.6564	14.5333
0.22	8.5860	8.9824	9.4850	10.1411	10.6395
0.24	2.8992	2.9609	3.0329	3.1179	3.1765
0.28	0.4718	0.475	0.4809	0.4852	0.4884
0.30	0.2325	0.2341	0.2358	0.2376	0.2388

Table 4.3 provides generated ARE values at different values of p, when ϕ is held constant and η varies. Results show that the ARE values are greater than one at low values of p. The efficiency increases as η increases. The observations on Tables 4.2 and 4.3 shows that the model performs better when sensitivity and specificity are high. These observations are graphically illustrated as in Figures 4.1, 4.2 and 4.3 below.

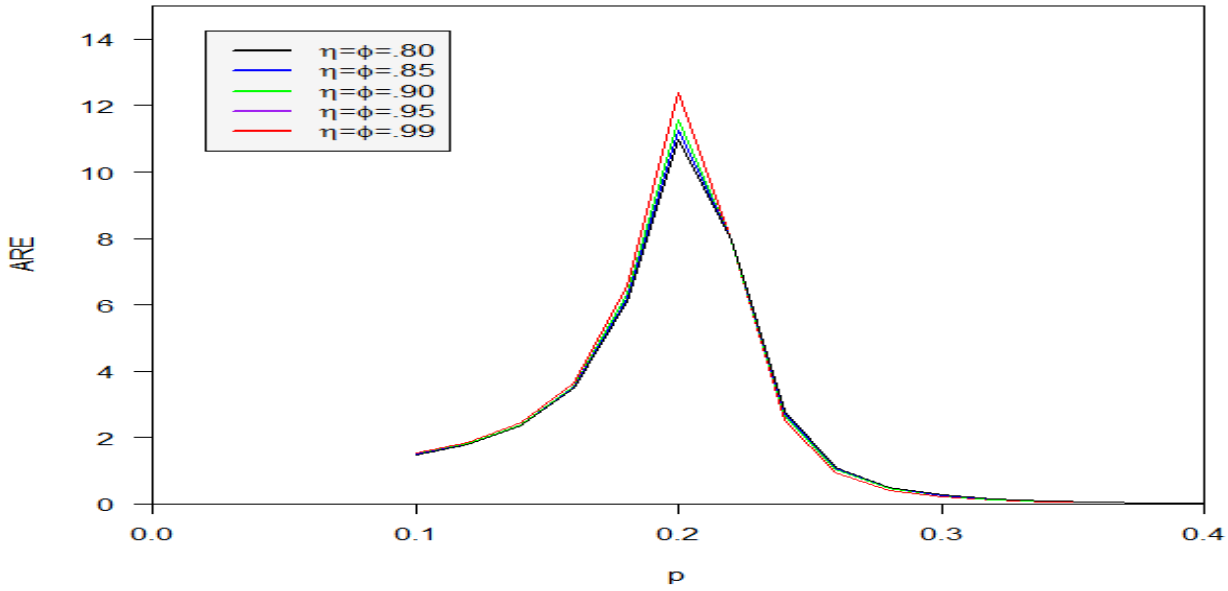

Figure 4.1: Plot of ARE values vs p when $\eta = \phi$ at specified values of p

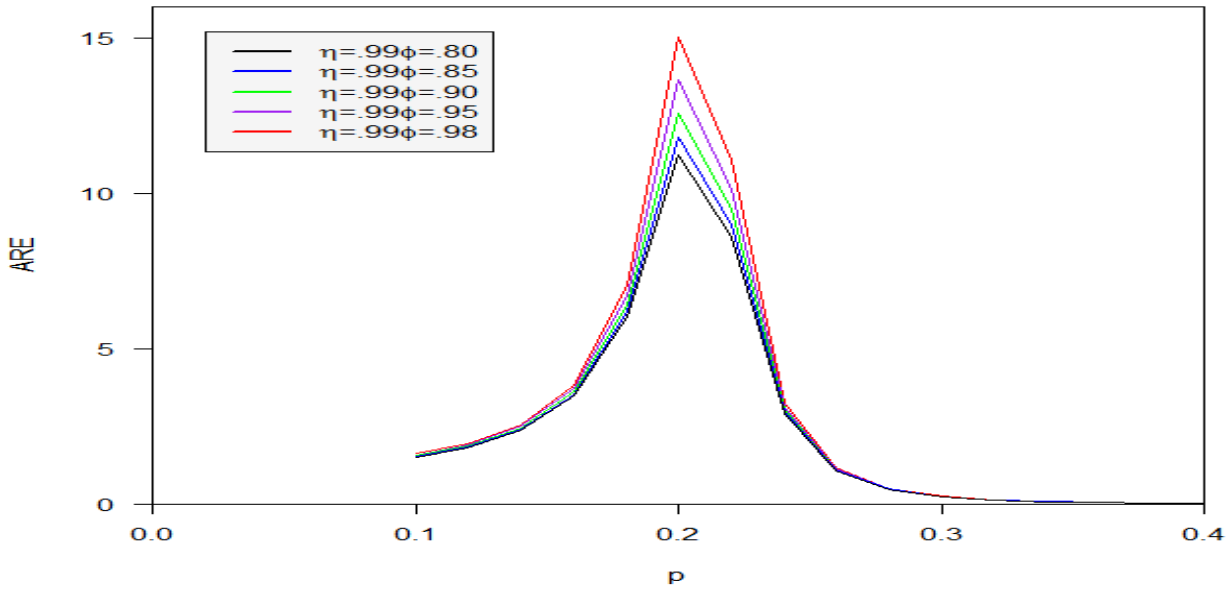

Figure 4.2: Plot of ARE values vs p when η is constant at 99% and ϕ varies

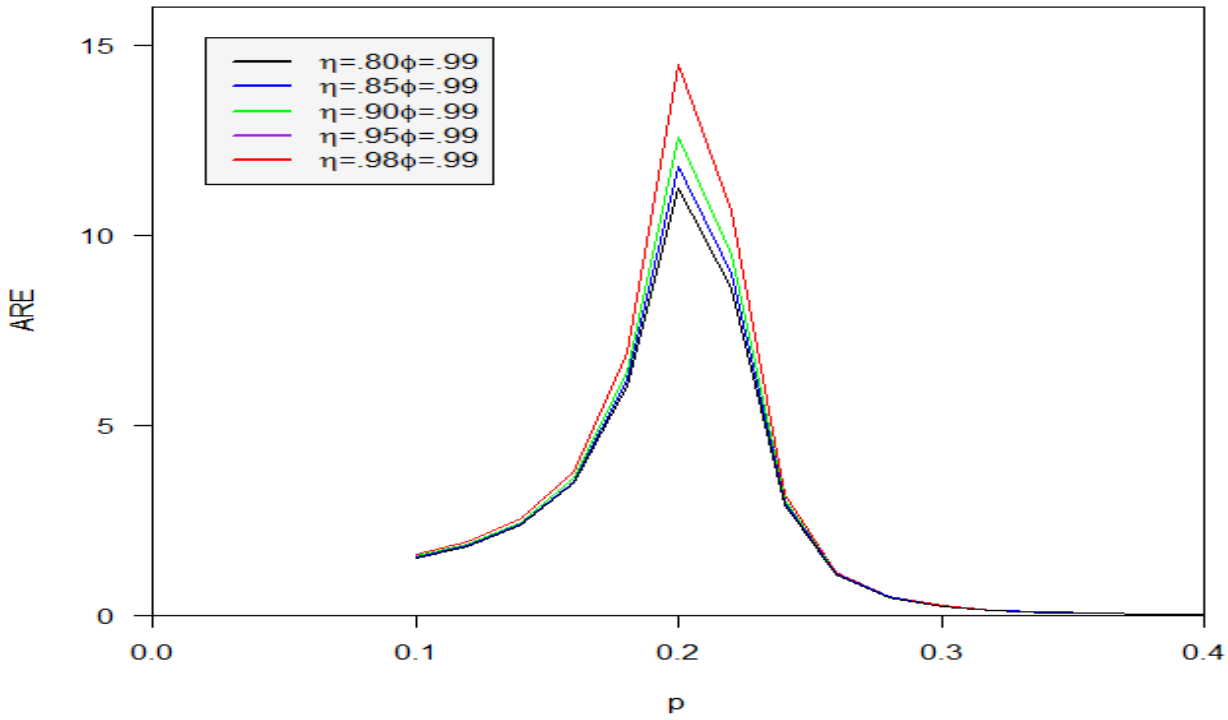

Figure 4.3: Plot of ARE values vs p when ϕ is constant at 99% and η varies

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATION

5.1 Introduction

Our study focussed on the effect of truncation on the efficiency of a constructed estimator. In this chapter we give the conclusion as per the objectives of the study and necessary recommendations for future work.

5.2 Conclusions

Evidently from the above discussions the three stage adaptive batch testing estimator without truncation with errors is more efficient than truncated estimator in presence of errors as we notice that all the ARE values are all greater than one at lower values of p. The Estimator performs better at higher values of sensitivity and specificity (η=ϕ=0.99) than at lower values (η=ϕ=0.80). It also performs better at low prevalence rates (p$\leq$ 0.28). Since Batch testing targets low prevalence rates, the model in this study befits this kind of scenario (low prevalence rates) given that it performs better than the Truncated Model.

5.3 Recommendation

After comparison, we realise that Non-truncated models outperforms Truncated models. We therefore recommend a generalised use of Non-truncated models of up to 'n' stages in statistical fields that require batch testing as it yields better results. However, in the absence of errors will the same scenario depict across the n stages? This perhaps presents an avenue for further research.

REFERENCES

[1] Berger T., Mandell J.W., and Subrahmanya P. *Maximally efficient two-stage screening*. Biometrics, 56:833840(8), September 2000.

[2] Biggerstaff, B.(2008). *Confidence intervals for the Difference of two proportions Estimated from pooled samples* .Jaurnal of Agricultural ,Biological and Environmetal statistics, 13,478-496.

[3] Brookmeyer, R,(1999). *Analysis of multistage Pooling studies of biological specimens for estimating desease incidence and privalence*. Biometrics 55, 608-612.

[4] C.Cohen,Jr (1989). *Estimating the mean and variance of Normal populations from singly Truncated and Doubly Truncated samples*, The Annals of mathematical statistics, vol.21, No.4, Edwards Brothers.

[5] Dorfman R (1943). *The detection of defective members of a large population* ,Annals of mathematical statistics, 14, 436-440.

[6] E. Barillot, B. Lacroix, and D. Cohen. *Theoretical analysis of library screening using an ndimensional pooling strategy*. Nucleic Acids Research, pages 62416247, 1991.

[7] Hughes,G., and Gottwald, T.R (1998). *Survey methods for assessment of citrus tristeza virus incidence* .Phytopathology 88. 715-723.

[8] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994) *Continuous Univariate Distributions*, Volume 1, Wiley. ISBN 0-471-58495-9

[9] Macula A.J. *Probabilistic nonadaptive and two-stage group testing with relatively small pools and DNA library screening.* Journal of Combinatorial Optimization, 2:385397(13), 1998.

[10] Nyongesa L.K (2011). *Dual estimation of prevalence and disease incidence in pool testing strategy.* Communication in statistics theory and methods 40,1-12.

[11] Okoth .A.W .(2012). *Two stage Adaptive Pool Testing for estimating prevalence of atrait in the presence of test errors.* IOSR-JM,Vol 13, Issue 1,P-ISSN: 2319-765X

[12] Okoth.A.W.(2017). *Multi-stsge Pool Testing Model With Test Errors:Improved Efficiency .Group testing a new goal Estimation* . biometrika, 52-181-194.

[13] Oliver-Hughes ,JM and Swallow W.H (1994). *A two stage adaptive group testing procedure for estimating small proportion.* American statistical association. 89,982-993.

[14] Oliver-Hughes J.M (2006). *Pooling for Experiments Blood screening and Drug Discovery screning; methods for experimentation in industry, Drug Discovery and Genectics.* New York, springer.

[15] Thompson K.M (1962). *Estimating of proportion of vector in anatural population of insects* .Biometics 18, 568-578.

[16] Wanyonyi W.R (2015). *Estimation of proportion of a trait by batch testing with Errors in Inspection in a quality control process.* Biometrics, 62:123-130.