PARAMETER ESTIMATION USING GENERATING FUNCTION BASED MINIMUM POWER DIVERGENCE MEASURE

TAY SIEW YING

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

PARAMETER ESTIMATION USING GENERATING FUNCTION BASED MINIMUM POWER DIVERGENCE MEASURE

TAY SIEW YING

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: TAY SIEW YING

Matric No: SGP 130006

Name of Degree: MASTER OF SCIENCE

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

PARAMETER ESTIMATION USING GENERATING FUNCTION BASED MINIMUM POWER DIVERGENCE MEASURE

Field of Study: STATISTICAL INFERENCE AND MODELLING (STATISTICS)

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

PARAMETER ESTIMATION USING GENERATING FUNCTION BASED MINIMUM POWER DIVERGENCE MEASURE

ABSTRACT

This research proposes a parameter estimation method that minimizes a probability generating function (pgf) based power divergence with a tuning parameter to mitigate the impact of data contamination. Special cases arise when the tuning parameter approaches zero, resulting in a Kullback-Leibler type divergence, and when it takes on the value of one, resulting in a pgf-based L_2 distance. The proposed estimator, BHHJ-PGF is linked to the M-estimators and therefore inherits the properties of consistency and asymptotic normality. The behaviour and performance of the proposed divergence was studied through simulations using Poisson and negative binomial distributions. Comparison was made with the maximum likelihood method (MLE), the pgf-based minimum Hellinger distance and also the pgf-based Jeffreys divergence. In terms of estimation bias and mean squared error from the results of simulations, the proposed estimation method performed better for smaller values of the tuning parameter as data contamination percentage increases. Application of the proposed method on four sets of real life data showed an improvement of fit and also its ability to mitigate the impact of outliers.

Keywords: asymptotic normality, density power divergence, M-estimators, probability generating function, robustness

PENGANGGARAN PARAMETER MENGGUNAKAN SUKATAN KUASA PENCAPAHAN MINIMUM BERASASKAN FUNGSI PENJANA

ABSTRAK

Kajian ini mencadangkan satu kaedah penganggaran parameter yang meminimumkan pencapahan kuasa berdasarkan fungsi penjana kebarangkalian (fpk) dengan parameter penala untuk mengurangkan kesan pencemaran data. Kes-kes khas timbul apabila parameter penala mendekati sifar, menghasilkan perbezaan jenis Kullback-Leibler, dan apabila ia mengambil nilai satu, menghasilkan jarak L_2 berasaskan fpk. Penganggar yang dicadangkan dikaitkan dengan penganggar-M dan oleh itu mempunyai sifat-sifat konsistensi dan asimtotik normal. Tingkah laku dan prestasi penyelewangan yang dicadangkan dikaji melalui simulasi menggunakan taburan Poisson dan binomial negatif. Perbandingan dibuat dengan penganggar kebolehjadian maksimum, jarak Hellinger minimum dan pencapahan Jeffrey berasaskan fpk. Dari segi bias parameter dan min perbezaan kuasa dua dari hasil simulasi, kaedah anggaran yang dicadangkan dalam kajian ini adalah lebih baik untuk nilai parameter penala yang lebih kecil dengan peningkatan peratusan pencemaran data. Penggunaan kaedah yang dicadangkan pada empat set data kehidupan sebenar menunjukkan peningkatan yang baik dan juga keupayaannya untuk mengurangkan impak daripada outlier.

Katakunci: normal asimptot, pencapahan ketumpatan kuasa, penganggar-M, fungsi penjana kebarangkalian, keteguhan

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude to my supervisors Dr. Ng Choung Min and Prof. Dr. Ong Seng Huat, who have encouraged, given advice and supported me throughout the entire journey of this research. Your patience in providing guidance and insights has enabled my growth as a student and researcher. I am deeply thankful to have both of you as my supervisors.

I would like to thank my dissertation examiners for taking their time to read this dissertation and giving crucial comments as well as suggestions to improve it.

Special thanks to my family for their love, encouragement and support throughout my study.

Last but not least, I would like to extend my appreciation and blessings to those who have supported me in any way for the successful completion of this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

T	:	A function of the sample $(X_1, X_2, \ldots, X_n)$
θ^c	:	Best fitting parameter
$d_\alpha\big(f(t), g(t)\big)$	:	BHHJ divergence between two pgf(s)
χ^2	:	Chi-square
$\xrightarrow{d}$	:	Convergence in distribution
$d\big(g(x), f(x)\big)$	:	Distance/ divergence between $g(x)$ and $f(x)$
$g_n(t)$	:	epgf
$\hat{\theta}$	:	Estimator
$f_X(x)$	:	Function of X
$L(\theta; \boldsymbol{x})$	:	Likelihood function with respect to θ
$x_1, \ldots, x_n$	:	Observation of size n
θ	:	Parameter
N_p	:	p-dimensional normal distribution
Θ	:	p-dimensional parameter space
$g_X(t)$	:	pgf of X
$g_\theta(t)$	:	pgf of X for parameter θ
$g_{(r,p)}(t)$	:	pgf of X where X is a negative binomial distribution
$g_\lambda(t)$	:	pgf of X where X is a Poisson distribution
$(r)_x$	:	Pochhammer symbol
$g(x), f(x)$	:	Probability distribution
$P(X = x_i),\ F_X(x)$	:	Probability function for X
X	:	Random variable
n	:	Sample size
$X_1, X_2, \ldots, X_n$	:	Sample of size n

$\to$	:	Tends to
α	:	Tuning parameter
cf	:	Characteristic function
epgf	:	Empirical probability generating function
JD	:	Jeffreys' divergence
lim	:	Limit
MLE	:	Maximum likelihood estimation
MSE	:	Mean squared error
BHHJ	:	Minimum density power divergence
MHD	:	Minimum Hellinger distance
mgf	:	Moment generating function
$NB(r,p)$	:	Negative binomial distribution with parameters r and p
BHHJ-PGF	:	pgf-based BHHJ
JD-PGF	:	pgf-based JD
MHD-PGF	:	pgf-based MHD
$Po(\lambda)$	:	Poisson distribution with parameter λ
pdf	:	Probability distribution function
pgf	:	Probability generating function
pmf	:	Probability mass function

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Parameter estimation is a process of constructing a statistic, typically under an optimization method such as maximization of the likelihood function, which can be used to estimate the parameters of a statistical model. For example, the sample mean and sample variance are statistics, also known as estimators for the population average and dispersion, respectively. A method to obtain a good estimator is to minimize a certain measure of discrepancy between the estimated parameter and the true parameter (of the population model). Desirable qualities of an estimator include unbiasedness, consistency and efficiency.

Well-known classical methods of parameter estimation are method of moments, due to Karl Peason (Pearson, 1894, 1902) and maximum likelihood estimation (MLE) method formalised by R.A Fisher in 1922 (Fisher, 1922). The method of moments estimates are simple to compute and consistent yet may be biased and not efficient. MLE is consistent, unbiased for large sample size and asymptotically efficient. Its advantages are hampered by the fact that MLE is not robust as it is easily affected by the presence of outliers in the data. Other more recent parameter estimation methods that have been widely adopted due to the advancement of computing power include the Bayesian estimation method. This method differs from the classical methods as it considers the parameters as random variables having some distributions, instead of unknown constants. This approach requires prior knowledge of the distribution for the parameters, which may be updated based on new information from time to time.

An outlier refers to an observation that is atypical or lies isolated from others in a random sample. Retention or exclusion of an outlier is a delicate matter; hence, an estimator which is less affected by the presence of outlier is highly desirable. This motivates effort to obtain robust estimators. One of the ways to measure robustness of an estimator is through the breakdown point (Hampel, 1971). A higher breakdown point indicates higher proportion of abnormal data that an estimator can handle before giving an incorrect result.

As an effort to obtain robust estimators, modifications are proposed to improve the performance of existing methods. Field and Smith (1994) suggested the weighted maximum likelihood approach, while weighted least squares method is used to forecast flood (Zhao et al., 2008). A generalization of MLE known as M-estimation proposed by Huber (1964) has greatly enriched the field of robust estimation due to its consistency and asymptotic properties.

Apart from the aforementioned methods, parameter estimation by way of minimizing density-based distance is also considered. Hellinger distance proposed in the context of count data (Simpson, 1987) proves to be effective in handling possible outliers. Using the minimum Hellinger distance, Beran (1977), pioneered the use of density-based minimum distance estimation in continuous models. This procedure requires nonparametric estimation of the probability density function with a kernel density estimator. In order to avoid this, (Basu et al., (1998) introduced the density power divergence for estimation to obtain a balance between robustness and asymptotic efficiency of parameter estimators through the use of a tuning parameter.

The use of probability generating function (pgf) in statistical inference was proposed by Kemp and Kemp (1988) as a tool for estimation due to its simplicity, especially when the corresponding probability mass function is complicated or intractable. The pgf, just as probability mass function (pmf) and probability density function (pdf), is unique to each distribution. The idea to equate empirical pgf (epgf) to pgf on a fixed finite set of values is investigated (Kemp & Kemp, 1988), and is then extended (Dowling & Nakamura, 1997) to include the asymptotic theory for the estimators.

Recently, parameter estimation by pgf-based Hellinger distance has been applied in univariate discrete case (Sim & Ong, 2010) as well as in multivariate discrete case (Ng et al., 2013).

In this research, the aim is to obtain a new pgf-based estimator for the parameters of two selected univariate discrete distributions. The method proposed here incorporates the pgf into the power divergence measure of Basu et al. (1998) to produce a consistent and robust estimator for the model parameters.

1.1 Aims and objectives

The objectives of this study are to i) obtain new estimation method that is insensitive to outliers in the data, ii) develop a generating function based power divergence measure for estimation, iii) examine the consistency and robustness of the estimator from the new estimation method and iv) compare the proposed new method with existing estimation methods.

1.2 Scope of the research

This study dwells into the area of parameter estimation for statistical distributions. In particular, a new approach using power divergence measure and probability generating function in parameter estimation for discrete distributions is explored.

Properties associated with the proposed estimator such as consistency and robustness against outliers will be investigated. Monte Carlo simulations will be employed to assess the performance of the proposed estimator against other well-known estimators such as the maximum likelihood estimator and pgf-based Hellinger distance estimator. The performance of the estimators will be assessed based on the mean squared errors and relative biases of the estimates.

1.3 Thesis structure

Chapter 2 contains the literature review, and also brief explanation of terms and concepts involved in this research.

Chapter 3 proposes a new parameter estimation method using a pgf-based power divergence measure. Proof of the proposed measure being a divergence is provided. Theoretical properties of the proposed estimator such as its link with M-estimators and the derivation of its asymptotic properties are considered in this chapter.

Next, Chapter 4 investigates the behaviour and performance of the proposed estimator through simulations. Comparison of performance, in terms of mean squared errors and relative biases, against other estimators are carried out and described. The efficiency of proposed estimator relative to that of MLE is also investigated.

Various situations are considered during the simulation runs, including the use of different sets of parameter values to give different shapes of a distribution as well as the addition of sample contamination in different percentages for estimation to test the ability of the proposed estimator in handling outliers.

Chapter 5 describes the application of proposed estimator in real life data set. The applicability of the proposed measure for inference is determined through the application to 4 sets of real life data. Calculation of χ^2 goodness of fit test-value is adopted as a measure to compare the goodness of fit.

Finally, Chapter 6 concludes and summarizes the findings of the research. Suggestions for future work are also included in this section.

CHAPTER 2: PRELIMINARIES AND LITERATURE REVIEW

Parameter estimation is an important process to facilitate the selection of the best-fitting model according to desired criterion for a data set of interest. Parameter estimation is applied in various fields ranging from space studies, engineering, sciences to social sciences.

2.1 Introduction

In the process of estimating parameters, sample data are first collected to represent the population before being fitted with an appropriate distribution. One of the two classes of distributions is the class of discrete distributions. A discrete distribution is a step function with only an enumerable (a one-to-one mapping with the set of all positive integers) number of steps, and can be represented by (Johnson et al., 2005)

$$P(X = x_i) = p_i.$$

where p_i is the probability function and the set $\{x_i\}$, the support of the random variable X. Random variables belonging to this class are called discrete random variables. Probability function of a discrete random variable is referred to as probability mass function (pmf). Some well-known discrete distributions include the binomial distribution, Poisson distribution (Po), negative binomial (NB) distribution and hypergeometric distribution.

As for the class of continuous distributions, its probability function is absolutely continuous and can be represented by (Johnson et al., 2005)

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\ dx.$$

Any function $f_X(x)$ for which the equation above is true for every x is a probability distribution of X. Random variables belonging to this class are called continuous random variables, having the support as a set of possible values, also known as a range. Every distribution, regardless of being discrete or continuous, has its own distinct pmf or pdf.

In each distribution, there will be one or more parameters. A suitable estimator is hence, needed to evaluate the parameters of the distribution that was fitted over the aforementioned data. An estimator serves as a rule to evaluate an estimate for the parameters based on observed values (data). Some examples of estimators are the maximum likelihood estimators (MLE), method of moments and distance based estimators. This process of estimating parameters enables us to make inferences regarding the population.

2.2 Desirable properties of an estimator: Consistency and robustness

Properties such as unbiasedness, consistency and efficiency are the basic properties sought after in an estimator because these properties determine how good an estimate is. Consider a random sample $X_1, X_2, \ldots, X_n$ on a random variable, each having a distribution with parameter θ and let $T_n = T(X_1, X_2, \ldots, X_n)$ be a statistic. T_n is said to be

i) an unbiased estimator of θ if $E(T_n) = \theta$.

ii) a consistent estimator of θ if $\lim_{n\to\infty} P(|T_n - \theta| \geq \varepsilon) = 0$ for $\varepsilon > 0$.

iii) an efficient estimator of θ if it is unbiased and the variance of T_n attains the Rao-Cramér lower bound, $Var(T) = \frac{1}{n} I^{-1}(\theta)$, where $I(\theta) = E\left[\left(\frac{\partial \ln f(x;\theta)}{\partial \theta}\right)^2\right]$.

In addition, robustness is fast in being recognized as a desirable property of an estimator. In a straightforward definition, an estimator is said to be robust if it is not sensitive to outliers in the data. In other words, the ability of an estimator to perform reliably despite the existence of extreme values is referred to as the robustness of that estimator. Certain scenarios such as mistakes or rounding errors present in collected data of an experiment are inevitable regardless of how cautious one is during the process. It became crucial to obtain robust estimators. Robustness is a compromise; it usually comes with a cost to the efficiency of estimator. However as Anscombe and Guttman (1960) put it, it is better to sacrifice some efficiency at the model to be insured against deviations from the model.

Robustness of an estimator can be discussed in terms of influence function or influence curve and breakdown points. The influence function (Hampel, 1974) of an estimator measures the sensitivity or effect of a contamination at one point x on the estimate. The influence function of the estimator $\hat{\theta}$ at x is defined as $\mathrm{IF}(x;\hat{\theta}) = \lim_{\varepsilon \to 0} \frac{T[(1-\varepsilon)F+\varepsilon\delta_x]-T(F)}{\varepsilon}$, where T is a functional, δ_x represents the contamination, and $(1-\varepsilon)F+\varepsilon\delta_x$ is a mixture distribution of F and δ_x. An estimator is robust if $|\mathrm{IF}(x;\hat{\theta})|$ is bounded for all x.

Other than influence function, robustness is also determined by obtaining the breakdown point of a sequence of estimators (Hampel, 1971). A simple explanation for breakdown point is, the point where the smallest number of alteration, k, to the original sample $x_1, \dots, x_n$ that an estimator can withstand before the distance between empirical distribution P_n and that of altered data $Q_{n,k}$ become unacceptable (Davies & Gather, 2005).

Some estimators and their properties will be discussed in the following sections. For example, the MLE is a consistent and asymptotically efficient estimator but sensitive to outliers. There have been attempts to improve the performance of MLE such as weighted MLE (Field & Smith, 1994) and generalized MLE known as M-estimator (Huber, 1964). Truncation or trimmed data is also applied to data to remove extreme values so as to get a better estimation. Another estimation method, BHHJ divergence relies on a tuning parameter to achieve robustness (Basu et al., 1998).

2.3 Maximum likelihood estimators (MLE)

One of the most important methods for parameter estimation is the maximum likelihood method (Fisher, 1922) . The maximum likelihood estimator has many desirable properties, such as consistency, asymptotic normality and efficiency. The idea is to search for a value in the parameter space that maximises the likelihood function. Consider the random variables, $X_1, \dots, X_n$, from an unknown distribution $f(\boldsymbol{x};\theta)$ where $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ represent observations. The estimate for the unknown parameter θ is obtained by maximizing the natural logarithm of the likelihood function, $L(\theta;\boldsymbol{x})$ with respect to θ.

2.4 Method of moments

Method of moments developed by Karl Pearson in 1894 (Pearson, 1894, 1902), involves equating theoretical moments (expected value of the powers of random variable) to its sample moments before solving for the parameters. In order to do so, the number of equations must be more than or equal to the dimensions of parameter. However, this method cannot be used on distributions where the moments do not exist such as the case for Cauchy distribution. Also, in applied work, the method of moments is very seldom used in comparison to the maximum likelihood and Bayesian estimations.

2.5 Minimum distance based estimators

Minimum distance based estimators are useful in parameter inference and in the area of goodness-of-fit (Basu et al., 2011). The estimators are widely applied in parametric inference of continuous and discrete data, particularly in the presence of outliers. This method estimates parameters by minimizing the distance between empirical and theoretical density functions.

In the area of parametric inference, one of the familiar measures is the minimum Hellinger distance (MHD) estimator, for which Beran (1977) studied its asymptotic efficiency and robustness for a continuous parametric model in compact support. MHD estimator between two probability distributions, $g(x)$ and $f(x)$, of a random variable, X, is defined as, $d\big(g(x), f(x)\big) = 2\int \left[g^{\frac{1}{2}}(x) - f^{\frac{1}{2}}(x)\right]^2 \, dx$. In the context of count data, MHD estimator is shown to have a breakdown point of 50% at the model (Simpson, 1987). Karlis and Xekalaki (1998) applied MHD estimation for finite Poisson mixtures and compared its properties with that of MLE to conclude that MHD method gives robust estimates.

The Kullback-Liebler (KL) divergence is an extension from the idea of treating information as a statistical concept to measure the difference or loss in information content between two probability density functions, $g(x)$ and $f(x)$, of a random variable, X (Kullback & Leibler, 1951). KL divergence, given by $d\big(g(x), f(x)\big) = \int f(x) \ln[f(x)/g(x)] \, dx$, is non-symmetrical, meaning that, $d(g(x), f(x)) \neq d(f(x), g(x))$. In discrete cases, the integration merely becomes summation.

A symmetrical form of KL divergence was proposed as Jeffreys' divergence (JD) (Jeffreys, 1946). Let $g(x)$ and $f(x)$ be two probability density functions, JD is given by $d\big(g(x), f(x)\big) = \int [f(x) - g(x)] \ln[f(x)/g(x)] \, dx$.

Apart from measuring the distance between probability density functions as studied in the mentioned research works, distance between theoretical and empirical distribution functions was also investigated. Some well-known ones are the Kolmogorov-Smirnov distance and Cramér-von Mises criterion. Both distances are widely applied in goodness-of-fit tests (Darling, 1957).

The terms ‘distance’, ‘disparity’, ‘divergence’ and ‘discrepancy’ are often used interchangeably to refer to the quantification of the degree of closeness, which is nonnegative, and equals to zero if and only if the empirical distribution fits its theoretical distribution exactly.

2.6 Minimum density power divergence

Basu et al. (1998) proposed an estimator that minimises the density power divergence between two probability distributions, $g(x)$ and $f(x)$. This divergence, defined as $d_\alpha\big(g(x), f(x)\big) = \int \left\{ f^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f^{\alpha}(x) + \frac{1}{\alpha} g^{1+\alpha}(x) \right\} dx$ contains a tuning parameter, α, which takes on nonnegative values. The advantage of minimum density power divergence estimator, abbreviated as BHHJ estimator, after the name of the authors, Basu, Harris, Hjort and Jones (1998), is that the use of nonparametric smoothing can be avoided besides being a robust and asymptotically efficient estimator, relative to MLE. Apart from the influence function, the breakdown point of the estimator has been used as one of the indicator for the robustness of the estimation method (Basu et al., 1998).

Nonparametric smoothing such as kernel density estimation is used to estimate the empirical density of a continuous random variable. This procedure involves determining the appropriate bandwidth for the observations which complicates the parameter estimation process. Minimization of BHHJ divergence leads to an expression where the empirical density is linear and hence, smoothing by nonparametric density is unnecessary.

The tuning parameter, α, in BHHJ divergence can be adjusted to provide a balance between robustness and efficiency of the estimator. Increase in α value increases robustness of the estimator but at the same time, decreases its efficiency. Although it can take on any real values, it is found that there is no substantial gain in robustness or efficiency for negative values of α, compared to the MLE method (Patra et al., 2013). As for values of $\alpha > 1$, estimator with high robustness at the cost of its efficiency is obtained. Hence, it is preferable to have $0 \leq \alpha \leq 1$.

Since then, there are several research works proposing different methods to identify an optimal α value. One of the methods involves equating estimated mean squared error to the asymptotic approximation of variance and bias, before minimizing it with respect to α (Warwick, 2005; Warwick & Jones, 2005). This approach results in the best α value that is distinct for each data set. Hong and Kim (2001) proposed similar technique, only without involving asymptotic approximation of bias.

There are two α values that merit special consideration. The BHHJ divergence became KL-type divergence when α approaches 0 and L_2 distance estimator when $\alpha = 1$. A mathematical connection was found to exist between the power divergences (Cressie & Read, 1984) and BHHJ density power divergence. Both divergences contain a tuning parameter,α, and a common member, KL divergence when α approaches 0 (Patra et al., 2013).

Apart from having a tuning parameter to control the trade-off between efficiency and robustness, BHHJ divergence also benefits from its link with M-estimator. Generally, M-estimators are estimators that solve the equation $\sum_i \psi(X_i, t) = 0$ where ψ is a function of the random variables, X_i with parameter t. M-estimators are asymptotically normally distributed, which makes it a desirable estimator. MLEs are classified as M-estimators too (Huber, 1964). With that in mind, BHHJ divergence, with established link to M-estimator, inherits the property of being asymptotically, normally distributed.

BHHJ divergence has been applied to estimate parameters in autoregressive model (Kang & Lee, 2014) and in lognormal distribution (Pak, 2014). In the case where restriction on observations are relaxed, independent but not identically distributed, BHHJ divergence was applied to linear regression (Ghosh & Basu, 2013). For estimation in real life data where outliers are often expected, BHHJ divergence has been developed for a generalized linear model by Ghosh and Basu (2016).

2.7 Probability generating function based estimators

Generating functions such as characteristic function (cf), moment generating function (mgf) and pgf are considered as statistical transforms. They are linked to each other; one could obtain cf and mgf from pgf. Feuerverger and McDunnough (1984) provided conditions for the estimation methods based on empirical generating functions to achieve asymptotic efficiency.

Let X be a discrete random variable with nonnegative integer values where $P(X = x_i)$ represents the probability mass function (pmf) of random variable X. The probability generating function (pgf) of X is defined in (Johnson et al., 2005) as

$$g_X(t) = \sum_{x=0}^{\infty} (t^x) P(X = x) = E[t^X] \, .$$

Every discrete distribution has its own unique pgf. The range of values of t corresponds to the radius of convergence of $g_X(t)$, which is $-1 \leq t \leq 1$.

There are several useful properties of pgf that makes it versatile and desirable to work with, such as the ease in calculating expectations and probabilities via differentiation, obtaining pmf from pgf and vice versa as well as simplifying sum of independent and identically distributed random variables.

One of the earlier research works involves empirical pgf (epgf) obtained by replacing the probability mass function in pgf by observed relative frequency. Kemp and Kemp (1988) developed a straight forward method of equating theoretical pgf to epgf on fixed, finite set of values, within the radius of convergence. It is found that for each part of a parameter space, different set of values is required.

Dowling and Nakamura (1997) extended the idea and developed asymptotic theory for the estimators. This theory assists, as a guidance, in choosing appropriate values for t. In the example for zero-inflated negative binomial distribution, the choices of t is suggested to be in the range of $0 < t < 1$.

Incorporating pgf and its empirical counterpart into divergences for the purpose of estimation was considered. Sim and Ong (2010) applied pgf in a total of 6 generalized Hellinger-type divergences in two cases of discrete data, i.e data with and without contamination for parameter estimation. Simulation results using the negative binomial (NB) distribution for these pgf-based estimation methods were compared to the performances of MLE and MHD estimator. Extension to multivariate discrete case for a pgf-based Hellinger-type distance produces robust and consistent estimators (Ng et al., 2013).

Apart from the area of parameter estimation, pgf is also proposed to be applied in goodness-of-fit tests. A distance based test statistic with pgf and epgf was employed for the fitting of Poisson distribution (Rueda et al., 1991) and NB distribution (Rueda & O' Reilly, 1999), with both research concluding that the distance test statistics perform as good as, if not better than χ^2 test.

Sharifdoust at al., (2016) introduced the pgf-based Jeffreys divergence (JD-PGF) estimator, which can be shown to be consistent due to its link to M-estimation. Simulation was performed to assess the performance of JD-PGF against that of the pgf-based MHD (MHD-PGF) and MLE in terms of biases and mean squared errors (MSEs). They concluded that the JD-PGF demonstrates better performance than the MHD-PGF in cases with outliers and close to the MLE in cases without outliers. In goodness-of-fit test, the JD-PGF performs better than χ^2 test.

CHAPTER 3: FORMULATION AND PROPERTIES OF BHHJ-PGF ESTIMATOR

This chapter introduces a pgf-based BHHJ estimator, which utilises a power divergence measure. The desired properties of an estimator such as consistency and robustness against outliers are investigated for this estimator.

3.1 Formulation of BHHJ-PGF estimator

Let the divergence between two pgf's, $f(t)$ and $g(t)$, be defined as

$$d_\alpha(f(t), g(t)) = \int_0^1 \left[g^{1+\alpha}(t) - \left(1 + \frac{1}{\alpha}\right) f(t) g^\alpha(t) + \left(\frac{1}{\alpha}\right) f^{1+\alpha}(t)\right] dt, \ \ \alpha > 0. \tag{3.1}$$

Here, the pgf's have a probabilistic interpretation and are well behaved when $t \in (0,1)$ (Rade, 1972).

Following Theorem 1 in Basu et al. (1998), this proposed measure (3.1) can be shown to be a valid pgf-based divergence measure.

Theorem: The measure proposed in equation (3.1) is a pgf-based divergence, where it is always non-negative, and equals to zero if and only if $f(t)$ and $g(t)$ are exactly identical.

Proof: Rearranging and simplifying the integrand in equation (3.1), we have

$$g^{1+\alpha}(t)\left\{1 - \left(1 + \frac{1}{\alpha}\right)\frac{f(t)}{g(t)} + \left(\frac{1}{\alpha}\right)\left(\frac{f(t)}{g(t)}\right)^{1+\alpha}\right\}.$$

Since $g(t)$ is always positive, it remains to show that $1 - \left(1 + \frac{1}{\alpha}\right)\frac{f(t)}{g(t)} + \left(\frac{1}{\alpha}\right)\left(\frac{f(t)}{g(t)}\right)^{1+\alpha}$ is always non-negative. For fixed t, we let $\frac{f(t)}{g(t)} = h$, giving

$$1 - \left(1 + \frac{1}{\alpha}\right) h + \left(\frac{1}{\alpha}\right) h^{1+\alpha}.$$

By differentiating this expression with respect to h and equating to zero, we obtain the point $h = 1$. Upon second order differentiation, we can identify that this point $h = 1$, which gives $f(t) = g(t)$, is a minimum point that $\left\{1 - \left(1 + \frac{1}{\alpha}\right)\frac{f(t)}{g(t)} + \left(\frac{1}{\alpha}\right)\left(\frac{f(t)}{g(t)}\right)^{1+\alpha}\right\}$ has a value of 0. This implies that the integrand is always non-negative. Hence, the proposed measure that integrates over $t \in (0,1)$ is always non-negative and equals to zero if and only if $f(t) \equiv g(t)$. □

The integrand is undefined when $\alpha = 0$. However, as $\alpha \to 0$, by l'Hopital's rule, equation (3.1) takes the form of a Kullback-Leibler type divergence as the limiting case, that is,

$$\lim_{\alpha \to 0} d_\alpha(f(t), g(t)) = \int_0^1 f(t) \ln \frac{f(t)}{g(t)} + g(t) - f(t) dt \,.$$

In the case of $\alpha = 1$, equation (3.1) becomes

$$d_1(f(t), g(t)) = \int_0^1 [g(t) - f(t)]^2 \; dt \,.$$

This is the pgf-based L_2-distance. The density based L_2-distance gives estimators with solid robustness properties against outliers (Patra et al., 2013).

Consider $X_1, X_2, \dots, X_n$ as a random sample of size n from a discrete distribution with true pgf $f(t)$. Let the pgf $g_\theta(t) = E_\theta[t^X]$, $\theta = (\theta_1, \theta_2, \dots, \theta_p) \in \Theta$, $0 < t < 1$, where Θ is the p-dimensional continuous open parameter space. Also, let $g_n(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i}$, $0 < t < 1$ be the epgf. Here, the pgf-based BHHJ (BHHJ-PGF) measure is proposed as

$$d_\alpha\big(g_n(t), g_\theta(t)\big) = \int_0^1 \left[g_\theta^{1+\alpha}(t) - \left(1 + \frac{1}{\alpha}\right) g_n(t) g_\theta^{\alpha}(t) + \left(\frac{1}{\alpha}\right) {g_n}^{1+\alpha}(t)\right] dt \,, \alpha > 0. \tag{3.2}$$

The proposed BHHJ-PGF estimator $\hat{\theta}$ will then minimise the measure in (3.2), that is

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \{d_\alpha(g_n(t)\,, g_\theta(t))\}.$$

3.2 Relation to M-estimation

M-estimators have attractive asymptotic properties such as consistency and asymptotic normality of the estimators (Huber, 1964). From Stefanski & Boos (2002), M-estimators $\hat{T}_n$ satisfies $\frac{1}{n}\sum_{i=1}^{n}\psi(X_i,\hat{T}_n) = 0$, where ψ must be a known function. By assuming that the true parameter T_0 is unique and $E[\psi(X_i\,,T_0)] = 0$, then there exists a sequence of M-estimators $\hat{T}_n$ such that $\hat{T}_n \to T_0$ as $n \to \infty$.

Taking the derivative of equation (3.2) with respect to θ after substitution with the epgf, $g_n(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i}$, we have

$$\frac{\partial[d_\alpha(g_n(t)\,,g_\theta(t))]}{\partial\theta} = (1+\alpha)\left[\int_0^1 g_\theta^\alpha(t)g_\theta'(t)\,dt - \frac{1}{n}\sum_{i=1}^{n}\int_0^1 t^{X_i}g_\theta^{\alpha-1}(t)g_\theta'(t)dt\right],$$

where $g_\theta'(t) = \frac{\partial g_\theta(t)}{\partial\theta}$. Upon equating to the null vector $\mathbf{0}$, then

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^1 g_\theta^\alpha(t)g_\theta'(t) - t^{X_i}g_\theta^{\alpha-1}(t)g_\theta'(t)dt\right]. \qquad (3.3)$$

The BHHJ-PGF estimator $\hat{\theta}$ is a solution to equation (3.3) since it minimises equation (3.2) and hence, satisfies $\frac{1}{n}\sum_{i=1}^{n}\psi(X_i,\hat{\theta}) = 0$, where $\psi(X_i,\hat{\theta}) = \int_0^1 g_{\hat{\theta}}^\alpha(t)\,g_{\hat{\theta}}'(t) - t^{X_i}g_{\hat{\theta}}^{\alpha-1}(t)g_{\hat{\theta}}'(t)dt$. Assuming that there is a unique minimum θ_0 and $E[\psi(X_i,\theta_0)] = 0$, this enables BHHJ-PGF to be cast as an M-estimator, where $\hat{\theta} \to \theta_0$ as $n \to \infty$. With this link, the BHHJ-PGF estimator, $\hat{\theta}$ also possesses the properties of consistency and asymptotic normality.

3.3 Asymptotic properties of estimator

This section shows the proof of consistency and asymptotic normality for the estimator $\hat{\theta}$. For brevity, $g_\theta(t)$, $g_n(t)$ and true pgf $f(t)$ will be represented by g_θ, g_n, and f, respectively.

The following assumptions will be made throughout.

A1. The pgf, g_θ has common support of all θ.

A2. There is an open subset of $\theta \in \Theta$, containing the best fitting parameter θ^c such that g_θ is three times differentiable with respect to θ.

A3. The integrals $\int_0^1 g_\theta^{1+\alpha} dt$ and $\int_0^1 g_n g_\theta^\alpha dt$ can be differentiated three times with respect to θ. The divergence is three times differentiable with respect to the parameter θ.

A4. The derivatives, expectations and summations can be taken under the integral sign.

A5. There exists a function $M_{jkl}(x)$ such that $\left|\nabla_{jkl}\left[\int_0^1 g_\theta^{1+\alpha}\, dt - \left(\frac{\alpha+1}{\alpha}\right)\int_0^1 t^x g_\theta^\alpha\, dt\right]\right| \leq M_{jkl}(x)$ for all $\theta \in w$, where $E_g\left[M_{jkl}(X)\right] = m_{jkl} < \infty$ for all j, k and l, and ∇_{jkl} represents the third order partial derivatives with respect to θ.

Rewriting equation (3.2), we have

$$d_\alpha(g_n, g_\theta) = \int_0^1 g_\theta^{1+\alpha}(t) dt - \left(1 + \frac{1}{\alpha}\right)\left(\frac{1}{n}\right)\sum_{i=1}^{n}\int_0^1 t^{X_i} g_\theta^\alpha\, dt + \left(\frac{1}{\alpha}\right)\left(\frac{1}{n}\right)\sum_{i=1}^{n}\int_0^1 t^{X_i}\, dt.$$

Hence, minimizing equation (3.2) is equivalent to minimizing

$$H_n(\theta) = \int_0^1 g_\theta^{1+\alpha} dt - \left(1 + \frac{1}{\alpha}\right)\left(\frac{1}{n}\right)\sum_{i=1}^{n}\int_0^1 t^{X_i} g_\theta^\alpha\, dt. \qquad (3.4)$$

To show consistency and asymptotic normality, the following notations will be adopted for the rest of the chapter.

θ $= (\theta_1, \theta_2, .., \theta_p)$, p-dimensional vector of parameters

$H_n^j(\theta)$ = first order partial derivative of $H_n(\theta)$ with respect to θ_j

$H_n^{jk}(\theta)$ = second order partial derivatives of $H_n(\theta)$ with respect to θ_j and θ_k

$H_n^{jkl}(\theta)=$ third order partial derivatives of $H_n(\theta)$ with respect to θ_j, θ_k and θ_l

θ^c = best fitting parameter

3.3.1 Consistency

With reference to Basu et al. (2011), the consistency of BHHJ-PGF estimator can be established in a similar way. To prove the existence of a sequence of solutions that is consistent, consider a sphere Q_a with radius a and center at the best fitting parameter θ^c.

We need to show that $P(H_n(\theta) > H_n(\theta^c)) \to 1$ for all points θ on the surface of Q_a, because if that is true, then it meant that $H_n(\theta)$ has a local minimum in Q_a.

At a local minimum, the minimum density power divergence estimating equation $H_n^j(\theta)$ must be satisfied, for any $a > 0$, with probability tending to 1. As $n \to \infty$, $H_n^j(\theta)$ has a root or estimate $\hat{\theta}(a)$ within Q_a.

Using Taylor's expansion to study the behaviour of $H_n(\theta)$ on Q_a, we expand $H_n(\theta)$ around θ^c and obtain

$$\frac{[H_n(\theta^c) - H_n(\theta)]}{1+\alpha}$$

$$= \sum_j (-A_j)(\theta_j - \theta_j^c) + \frac{1}{2}\sum_j \sum_k (-B_{jk})(\theta_j - \theta_j^c)(\theta_k - \theta_k^c)$$

$$+ \frac{1}{6}\sum_j \sum_k \sum_l (\theta_j - \theta_j^c)(\theta_k - \theta_k^c)(\theta_l - \theta_l^c)\frac{1}{n}\sum_{i=1}^{n} \gamma_{jkl}(x_i) M_{jkl}(x_i)$$

where $A_j = \frac{1}{1+\alpha} H_n^j(\theta)|_{\theta=\theta^c}$, $B_{jk} = \frac{1}{1+\alpha} H_n^{jk}(\theta)|_{\theta=\theta^c}$ and $\frac{1}{n}\sum_{i=1}^{n} \gamma_{jkl}(x_i) M_{jkl}(x_i) = -\frac{1}{1+\alpha} H_n^{jkl}(\theta)|_{\theta=\theta^c}$ with $0 \le |\gamma_{jkl}(x_i)| \le 1$ and $M_{jkl}(x_i)$ as stated in Assumption A5.

Following the work in Basu et al. (2011), it is found that there exists a sequence of solutions $\hat{\theta}$ such that $P(\|\hat{\theta}-\theta\|_2 < a) \to 1$ for sufficiently small a. Let the limit of the sequence of solutions be $\ddot{\theta}$, and the root closest to θ. Then it can be said that $P(\|\ddot{\theta}-\theta\|_2 < a) \to 1$ for all $a > 0$. This proved the existence of a consistent sequence of roots to $H_n^j(\theta)$ with probability tending to 1.

Detailed proof is included in Appendix A.

3.3.2 Asymptotic normality

This section proves the asymptotic normality of the proposed estimator, $\hat{\theta}$. First, the limits for $H_n(\theta)$, $H_n^j(\theta)$ and $H_n^{jk}(\theta)$ are obtained. Let $u_{j\theta} = u_{j\theta}(t) = \frac{\partial}{\partial\theta_j}\ln g_\theta(t)$. By the law of large number, as $n \to \infty$,

$$H_n(\theta) = \int_0^1 g_\theta^{1+\alpha} dt - \left(1+\frac{1}{\alpha}\right)\left(\frac{1}{n}\right)\sum_{i=1}^{n}\int_0^1 t^{X_i} g_\theta^{\alpha}\, dt$$

$$\to \int_0^1 g_\theta^{1+\alpha} dt - \left(1+\frac{1}{\alpha}\right)\int_0^1 g_\theta^{\alpha} f dt = H(\theta)$$

$$H_n^j(\theta^c) = \int_0^1 (1+\alpha) g_\theta^{\alpha}\frac{\partial g_\theta}{\partial\theta_j} dt - \left(\frac{\alpha+1}{\alpha}\right)\frac{1}{n}\sum_{i=1}^{n}\int_0^1 t^{X_i}\alpha g_\theta^{\alpha-1}\frac{\partial g_\theta}{\partial\theta_j} dt\Bigg|_{\theta=\theta^c}$$

$$= (1+\alpha)\left\{\int_0^1 g_\theta^{\alpha+1} u_{j\theta}\, dt - \frac{1}{n}\sum_{i=1}^{n}\int_0^1 t^{X_i} g_\theta^{\alpha} u_{j\theta}\, dt\right\}\Bigg|_{\theta=\theta^c}$$

$$\to (1+\alpha)\int_0^1 g_\theta^{\alpha+1} u_{j\theta}\, dt - \int_0^1 f\, g_\theta^{\alpha} u_{j\theta}\, dt\Bigg|_{\theta=\theta^c}$$

$$= \frac{\partial H(\theta)}{\partial\theta_j}\Bigg|_{\theta=\theta^c} = H^j(\theta^c)$$

$$= 0$$

since θ^c minimises equation (3.4) as well as $H(\theta)$.

$$H_n^{jk}(\theta^c) = (1+\alpha)\left\{\int_0^1 (\alpha+1) g_\theta^{\alpha+1} u_{k\theta} u_{j\theta} + \, g_\theta^{\alpha+1} \frac{\partial u_{j\theta}}{\partial \theta_k} \, dt \right.$$

$$\left. - \, \frac{1}{n}\sum_{i=1}^{n} \int_0^1 t^{X_i} \left(\alpha g_\theta^\alpha u_{k\theta} u_{j\theta} + \, g_\theta^\alpha \frac{\partial u_{j\theta}}{\partial \theta_k} \right) dt \right\}\Bigg|_{\theta=\theta^c}$$

$$\to (1+\alpha)\left\{\int_0^1 (\alpha+1) g_\theta^{\alpha+1} u_{k\theta} u_{j\theta} + \, g_\theta^{\alpha+1} \frac{\partial u_{j\theta}}{\partial \theta_k} \, dt \right.$$

$$\left. - \, \int_0^1 \alpha f g_\theta^\alpha u_{k\theta} u_{j\theta} + \, f g_\theta^\alpha \frac{\partial u_{j\theta}}{\partial \theta_k} \, dt \right\}\Bigg|_{\theta=\theta^c}$$

$$= (1+\alpha) J_{jk}(\theta^c)$$

where $J_{jk}(\theta^c) = \left\{\int_0^1 g_\theta^{\alpha+1} u_{k\theta} u_{j\theta} \, dt + \int_0^1 \left(i_{jk\theta} - \alpha u_{k\theta} u_{j\theta}\right)(f - g_\theta) g_\theta^\alpha \, dt\right\}\Big|_{\theta=\theta^c}$ and $i_{jk\theta} = i_{jk\theta}(t) = -\frac{\partial}{\partial \theta_k} u_{j\theta}(t)$.

Next, expand $H_n^j(\theta)$ about the best fitting parameter θ^c using Taylor's expansion to obtain

$$H_n^j(\theta) = H_n^j(\theta^c) + \sum_k (\theta_k - \theta_k^c)\, H_n^{jk}(\theta') + \frac{1}{2}\sum_k \sum_l (\theta_k - \theta_k^c)\,(\theta_l - \theta_l^c) H_n^{jkl}(\theta^*)\,.$$

Evaluating at $\theta = \hat{\theta}$, we have

$$0 = H_n^j(\theta^c) + \sum_k (\hat{\theta}_k - \theta_k^c)\, H_n^{jk}(\theta^c) + \frac{1}{2}\sum_k \sum_l (\hat{\theta}_k - \theta_k^c)\,(\hat{\theta}_l - \theta_l^c) H_n^{jkl}(\theta^*)\,,$$

where θ^* is a point on the line segment connecting θ and θ^c. Rearranging the expansion above, we have

$$\sqrt{n}\sum_k (\hat{\theta}_k - \theta_k^c)\left\{H_n^{jk}(\theta^c) + \frac{1}{2}\sum_l (\hat{\theta}_l - \theta_l^c) H_n^{jkl}(\theta^*)\right\} = -\sqrt{n} H_n^j(\theta^c),$$

which can be expressed as

$$\sum_{k=1}^{p} A_{jkn} Y_{kn} = T_{jn}, \tag{3.5}$$

where

$$Y_{kn} = \sqrt{n}\left(\hat{\theta}_k - \theta_k^c\right)$$

$$A_{jkn} = \left\{ H_n^{jk}(\theta^c) + \frac{1}{2}\sum_l (\hat{\theta}_l - \theta_l^c) H_n^{jkl}(\theta^*) \right\}$$

$$T_{jn} = -\sqrt{n} H_n^j(\theta^c). \tag{3.6}$$

As noted in Basu et al. (2011), the solution $\left(Y_{1n}, \dots, Y_{pn}\right)$ of equation (3.5) tends in law to the solutions of

$$\sum_{k=1}^{p} a_{jk}\, Y_k = T_j\,, (j = 1, \dots, p),$$

where $\left(T_{1n}, T_{2n}, \dots, T_{pn}\right)$ converges weakly to $\mathbf{T} = \left(T_1, T_2, \dots, T_p\right)$ and A_{jkn} converges in probability to a_{jk} for fixed $j, k = 1,2, \dots, p$. Hence, in the form of matrices, the solution $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$ satisfies

$$\mathbf{AY} = \mathbf{T} \quad \text{or} \quad \mathbf{Y} = \mathbf{A}^{-1}\mathbf{T} \tag{3.7}$$

where the non-singular matrix $\mathbf{A} = \left(a_{jk}\right)_{p \times p}$.

From the limit of $H_n^{jk}(\theta^c)$ and the fact that $H_n^{jkl}(\theta^*)$ is bounded with probability tending to 1 (from Condition A5), the limit for A_{jkn} as $n \to \infty$ is given by

$$\mathbf{A} = \left(\lim_{n\to\infty} A_{jkn}\right)_{p\times p} = \left(\lim_{n\to\infty} \left\{ H_n^{jk}(\theta^c) + \frac{1}{2}\sum_l \left(\hat{\theta}_l - \theta_l^c\right) H_n^{jkl}(\theta^*) \right\}\right)_{p\times p}$$

$$= \left(\lim_{n\to\infty} H_n^{jk}(\theta^c)\right)_{p\times p}$$

$$= (1+\alpha)\mathbf{J}, \qquad (3.8)$$

where $\mathbf{J} = (1+\alpha)\left(J_{jk}(\theta^c)\right)_{p\times p}$.

From equation (3.6), we have

$$T_{jn} = -\sqrt{n}\, H_n^j(\theta^c)$$

$$= -\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n}(1+\alpha)\left\{\int_0^1 g_{\theta^c}^{\alpha+1} u_{j\theta^c}\ dt - \int_0^1 t^{X_i} g_{\theta^c}^{\alpha} u_{j\theta^c}\ dt\right\}$$

$$= -\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n} V_{j\theta^c}(X_i)$$

The mean of $V_{j\theta^c}(X_i)$ is

$$E\left[V_{j\theta^c}(X_i)\right] = (1+\alpha)\left\{\int_0^1 g_{\theta^c}^{1+\alpha} u_{j\theta^c}\, dt - \int_0^1 f g_{\theta^c}^{\alpha} u_{j\theta^c}\, dt\right\}$$

$$= H^j(\theta^c) = 0, \qquad j = 1, \dots, p,$$

which implies that $E\left[T_{jn}\right] = 0$.

To find the covariance of $\left(V_{j\theta^c}(X_i), V_{k\theta^c}(X_i)\right)$, we obtained

$$V_{j\theta^c}(X_i)V_{k\theta^c}(X_i)$$

$$= (1+\alpha)^2\left\{\int_0^1 g_{\theta^c}^{\alpha+1} u_{j\theta^c}\, dt \int_0^1 g_{\theta^c}^{\alpha+1} u_{k\theta^c}\, dt - \int_0^1 g_{\theta^c}^{\alpha+1} u_{j\theta^c}\, dt \int_0^1 t^{X_i} g_{\theta^c}^{\alpha} u_{k\theta^c}\ dt\right.$$

$$- \int_0^1 t^{X_i} g_{\theta^c}^{\alpha} u_{j\theta^c}\, dt \int_0^1 g_{\theta^c}^{\alpha+1} u_{k\theta^c}\, dt$$

$$\left. + \int_0^1 t^{X_i} g_{\theta^c}^{\alpha} u_{j\theta^c}\, dt \int_0^1 t^{X_i} g_{\theta^c}^{\alpha} u_{k\theta^c}\, dt\right\}.$$

Then,

$$Cov\left(V_{j\theta^c}(X_i), V_{k\theta^c}(X_i)\right)$$

$$= E\left[V_{j\theta^c}(X_i)V_{k\theta^c}(X_i)\right]$$

$$= (1+\alpha)^2\left\{\int_0^1 g_{\theta^c}^{\alpha+1}u_{j\theta^c}\,dt\,\left[\int_0^1 g_{\theta^c}^{\alpha+1}u_{k\theta^c}\,dt - \int_0^1 E(t^{X_i})g_{\theta^c}^{\alpha}u_{k\theta^c}\,dt\right]\right.$$

$$- \int_0^1 E(t^{X_i})g_{\theta^c}^{\alpha}u_{j\theta^c}\,dt\int_0^1 g_{\theta^c}^{\alpha+1}u_{k\theta^c}\,dt$$

$$\left. + \; E\left(\int_0^1 t^{X_i}g_{\theta^c}^{\alpha}u_{j\theta^c}\,dt\int_0^1 t^{X_i}g_{\theta^c}^{\alpha}u_{k\theta^c}\,dt\right)\right\}$$

$$= (1+\alpha)^2\left\{E\left[\int_0^1 t^{X_i}g_{\theta^c}^{\alpha}u_{j\theta^c}\,dt\int_0^1 t^{X_i}g_{\theta^c}^{\alpha}u_{k\theta^c}\,dt\right] - \int_0^1 g_{\theta^c}^{\alpha}fu_{j\theta^c}\,dt\int_0^1 g_{\theta^c}^{\alpha}fu_{j\theta^c}\,dt\right\}$$

$$= (1+\alpha)^2K_{jk},$$

since $H^j(\theta^c) = 0$ implies that $\int_0^1 g_{\theta^c}^{\alpha+1}u_{j\theta^c}\,dt = \int_0^1 g_{\theta^c}^{\alpha}fu_{j\theta^c}\,dt$.

By the Central Limit Theorem and the results above, for sufficiently large n, $\left(T_{1n}, T_{2n}, \ldots, T_{pn}\right) \to \mathbf{T}$, which has a multivariate normal distribution with mean vector **0** and covariance matrix $(1+\alpha)^2\mathbf{K}$, where $\mathbf{K} = \left(K_{jk}\right)_{p\times p}$.

Since we have the solution **Y** satisfying equation (3.7) and the result from equation (3.8), as $n \to \infty$, $\left(Y_{1n}, Y_{2n}, \ldots, Y_{pn}\right) \to \mathbf{Y}$ also has a multivariate normal distribution with mean vector **0** and covariance matrix

$$\mathbf{A}^{-1}(1+\alpha)^2\mathbf{K}(\mathbf{A}^{-1})^T = [(1+\alpha)\mathbf{J}]^{-1}(1+\alpha)^2\mathbf{K}[(1+\alpha)\mathbf{J}]^{-1}$$

$$= \mathbf{J}^{-1}\mathbf{K}\mathbf{J}^{-1},$$

denoted as $N_p(\mathbf{0}, \mathbf{J}^{-1}\mathbf{K}\mathbf{J}^{-1})$ distribution. This indicates that the BHHJ-PGF estimator $\hat{\theta}$ is asymptotically normal, that is, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\theta} - \theta^c\right) \xrightarrow{d} N_p(\mathbf{0}, \mathbf{J}^{-1}\mathbf{K}\mathbf{J}^{-1}),$$

where $\xrightarrow{d}$ denotes convergence in distribution.

CHAPTER 4: SIMULATION AND DISCUSSION

In this chapter, the behaviour of the proposed BHHJ-PGF estimator is studied through simulations. The MHD-PGF estimator, JD-PGF estimator and MLE are included as well in the simulation study for comparison purposes. For ease of reference, the mathematical formulas of the measure for these estimators of the parameter θ are as follows.

(i) *MHD-PGF* (Sim & Ong, 2010):

$$\tilde{\theta}_{MHD-PGF} = \arg\min_{\theta\in\Theta} T(\theta;\alpha,n) = \arg\min_{\theta\in\Theta} \int_0^1 [g_n(t)^\alpha - g_\theta(t)^\alpha]^2 \beta(t)\, dt\,,\ \ 0 < \alpha \le 1$$

where $g_n(t) = \frac{1}{n}\sum_{i=1}^{n} t^{x_i}$ is the empirical probability generating function (epgf), $g_\theta(t)$ denotes the probability generating function (pgf) and Θ is the parameter space. The form of MHD-PGF estimator adopted here has $\alpha = 1/2$ and the weight function, $\beta(t) = 1$ as is preferred in (Sim & Ong, 2010) due to shorter computation time and simplicity.

(ii) *JD-PGF* (Sharifdoust et al., 2016):

$$\tilde{\theta}_{JD-PGF} = \arg\min_{\theta\in\Theta} J(g_\theta, g_n) = \arg\min_{\theta\in\Theta} \int_0^1 [g_n(t) - g_\theta(t)] \ln[g_n(t)/g_\theta(t)]\, dt$$

where $g_n(t)$ is the epgf, $g_\theta(t)$ represents the pgf and Θ is the parameter space.

MLE:

$$\hat{\theta} = \arg\max_{\theta\in\Theta}\ \ln L(\theta)$$

where $\ln L(\theta)$ is the log likelihood function.

The distributions considered in the simulations are the Poisson distribution, $Po(\lambda)$, and the negative binomial distribution, $\mathrm{NB}(r,p)$, with their respective probability mass function (pmf) and pgf given as follows.

Poisson distribution, $Po(\lambda)$:

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x=0,1,2,\dots, \qquad \lambda>0$$

$$g_{\lambda}(t)=e^{\lambda(t-1)}$$

Negative binomial distribution, $\mathrm{NB}(r,p)$:

$$P(X=x)=\frac{(r)_x}{x!}p^{r}(1-p)^{x}, \quad x=0,1,2,\dots, \qquad r>0, \qquad 0<p<1$$

$$g_{(r,p)}(t)=(p/[1-(1-p)t])^{r}$$

where $(r)_x$ is the Pochhammer symbol.

All simulations are carried out through a Windows 7 (16GB RAM) operated computer. Optimization is performed using ‘optim’ package in R programming language (version 3.1.0 “Spring Dance”) through its default ‘Nelder-Mead’ optimization technique for the pgf-based estimators with NB distribution. This technique considers minimization of the function value. As for pgf-based estimators with Poisson distribution, ‘Brent’ which is a better method for one parameter case, is adopted. MLE is done through ‘mle’ package built in R. In order to evaluate the integrations involved, a six-point Legendre quadrature method is used because it is found that increasing the number of nodes (up to 12 points) does not improve computation accuracy.

At the end of each set of simulations, the mean squared error (MSE) and relative biases (in bracket) for the parameter(s) are computed as a measure of performance for all estimators.

4.1 Simulation using Poisson distribution

First and foremost, a one-parameter distribution, the Poisson distribution $Po(\lambda)$, is considered to start off the simulation study. Here, $\lambda = 1,2,3, \ldots, 10$ are taken into consideration with the sample size, n fixed at 500 which is considered to be a large enough sample size and the number of simulation runs fixed at 8000. Other choices of numbers for simulation runs are not considered in this preliminary study.

4.1.1 BHHJ-PGF(α) and other estimators

In Basu et al. (1998) the parameter α provides a bridge between robustness and efficiency of the BHHJ estimator. With that as a motivation, simulation is carried out to investigate if the same conclusion could be attained in its pgf-based counterpart. Here, the performance is explored by varying the values of α in the proposed BHHJ-PGF estimator. As comparisons, aforementioned estimators, namely MHD-PGF, JD-PGF and MLE, are also included. Sample data with and without contamination are taken into consideration. For ease of reference, only certain α values are shown here, the complete tables and figures of results can be found in Appendix B.

4.1.1.1 Sample data without contamination

The simulation results for samples without any contamination are presented in Table 4.1 and Figure 4.1. A general trend can be seen from the MSE values that MLE is the most efficient yet negatively biased throughout all the distributions considered. Among BHHJ-PGF estimators, the ones with greater α leads to better efficiency and smaller relative biases. JD-PGF performs close to that of BHHJ-PGF with small α.

Table 4.1: MSE and relative biases (in bracket) for estimators with samples of size $n = 500$, from $Po(\lambda)$, without contamination. (See Table B1 in Appendix B for complete set of simulated α values.)

Po(λ)	BHHJ-PGF(α)					MLE	MHD-PGF	JD-PGF
	α = 0	α = 0.001	α = 0.5	α = 1.0	α = 2.0			
Po(1)	0.0028	0.0028	0.0027	0.0026	0.0026	0.0020	0.0028	0.0028
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(-0.0006)	(0.0001)	(0.0001)
Po(2)	0.0075	0.0075	0.0071	0.0064	0.0064	0.0040	0.0075	0.0075
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(-0.0005)	(0.0000)	(0.0000)
Po(3)	0.0150	0.0150	0.0132	0.0108	0.0108	0.0061	0.0151	0.0151
	(0.0007)	(0.0007)	(0.0003)	(0.0000)	(0.0000)	(-0.0003)	(0.0007)	(0.0007)
Po(4)	0.0246	0.0245	0.0194	0.0144	0.0144	0.0080	0.0246	0.0246
	(0.0010)	(0.0010)	(0.0003)	(0.0000)	(0.0000)	(-0.0005)	(0.0010)	(0.0010)
Po(5)	0.0357	0.0356	0.0247	0.0201	0.0171	0.0100	0.0360	0.0360
	(0.0012)	(0.0012)	(0.0004)	(0.0000)	(0.0000)	(-0.0004)	(0.0016)	(0.0016)
Po(6)	0.0461	0.0460	0.0282	0.0191	0.0191	0.0121	0.0467	0.0469
	(0.0012)	(0.0012)	(0.0000)	(-0.0002)	(-0.0002)	(-0.0003)	(0.0017)	(0.0018)
Po(7)	0.0532	0.0531	0.0296	0.0234	0.0205	0.0141	0.0533	0.0537
	(0.0013)	(0.0013)	(0.0001)	(0.0000)	(-0.0001)	(-0.0003)	(0.0021)	(0.0021)
Po(8)	0.0566	0.0565	0.0301	0.0218	0.0218	0.0162	0.0553	0.0560
	(0.0008)	(0.0008)	(-0.0001)	(-0.0001)	(-0.0001)	(-0.0002)	(0.0017)	(0.0019)
Po(9)	0.0559	0.0558	0.0300	0.0230	0.0230	0.0182	0.0531	0.0541
	(0.0007)	(0.0007)	(0.0000)	(-0.0001)	(-0.0001)	(-0.0002)	(0.0016)	(0.0017)
Po(10)	0.0522	0.0521	0.0299	0.0241	0.0241	0.0197	0.0491	0.0502
	(0.0006)	(0.0006)	(0.0001)	(-0.0001)	(-0.0001)	(-0.0002)	(0.0013)	(0.0014)

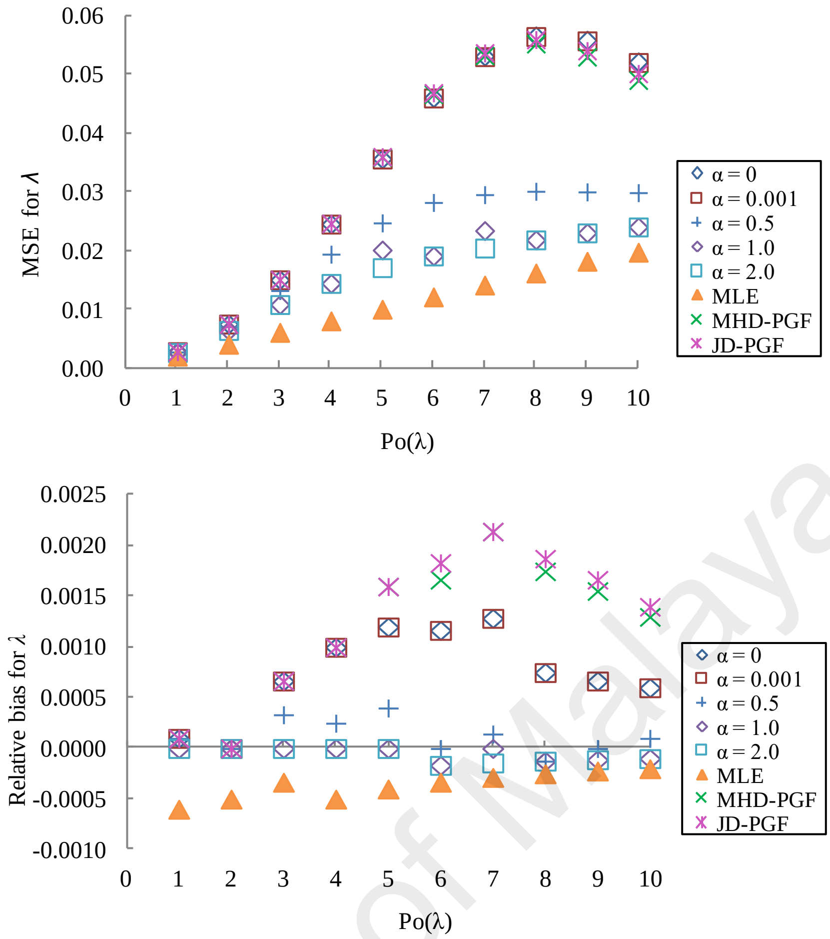

Figure 4.1: MSE and relative biases for estimators with samples of size $n = 500$, from $Po(\lambda)$, without contamination.

4.1.1.2 Sample data with contamination

A 5% contamination is introduced in the sample by data generating from a Poisson distribution with mean, λ_c four times that of the original ($\lambda_c = 4\lambda$). This is chosen so that the generated data are suitably far away from the other data to be considered as contamination. The performance of all estimators are presented in Table 4.2 and Figure 4.2.

Table 4.2: MSE and relative biases (in bracket) for estimators with samples of size $\boldsymbol{n = 500}$, from $\boldsymbol{Po(\lambda)}$, with 5% contamination. (See Table B2 in Appendix B for complete set of simulated $\boldsymbol{\alpha}$ values.)

Po(λ)	BHHJ-PGF(α)					MLE	MHD-PGF	JD-PGF
	α = 0	α = 0.001	α = 0.5	α = 1.0	α = 2.0			
Po(1)	0.0068	0.0068	0.0070	0.0072	0.0074	0.0246	0.0068	0.0068
	(0.0626)	(0.0626)	(0.0640)	(0.0656)	(0.0678)	(0.1490)	(0.0626)	(0.0626)
Po(2)	0.0140	0.0140	0.0144	0.0149	0.0158	0.0940	0.0140	0.0140
	(0.0390)	(0.0390)	(0.0415)	(0.0440)	(0.0475)	(0.1495)	(0.0390)	(0.0390)
Po(3)	0.0246	0.0246	0.0247	0.0255	0.0277	0.2087	0.0247	0.0247
	(0.0307)	(0.0307)	(0.0340)	(0.0373)	(0.0420)	(0.1497)	(0.0307)	(0.0307)
Po(4)	0.0381	0.0381	0.0368	0.0382	0.0428	0.3678	0.0382	0.0382
	(0.0270)	(0.0270)	(0.0313)	(0.0353)	(0.0410)	(0.1498)	(0.0270)	(0.0270)
Po(5)	0.0550	0.0550	0.0506	0.0536	0.0625	0.5727	0.0555	0.0556
	(0.0254)	(0.0254)	(0.0306)	(0.0354)	(0.0416)	(0.1498)	(0.0258)	(0.0258)
Po(6)	0.0721	0.0721	0.0647	0.0713	0.0874	0.8207	0.0732	0.0735
	(0.0245)	(0.0245)	(0.0305)	(0.0360)	(0.0428)	(0.1497)	(0.0252)	(0.0253)
Po(7)	0.0878	0.0878	0.0803	0.0933	0.1197	1.1153	0.0891	0.0896
	(0.0244)	(0.0244)	(0.0311)	(0.0370)	(0.0444)	(0.1497)	(0.0254)	(0.0254)
Po(8)	0.0992	0.0992	0.0970	0.1189	0.1589	1.4538	0.0996	0.1005
	(0.0241)	(0.0241)	(0.0316)	(0.0379)	(0.0458)	(0.1498)	(0.0251)	(0.0251)
Po(9)	0.1089	0.1088	0.1172	0.1505	0.2075	1.8378	0.1081	0.1094
	(0.0242)	(0.0242)	(0.0322)	(0.0389)	(0.0473)	(0.1498)	(0.0252)	(0.0252)
Po(10)	0.1163	0.1163	0.1406	0.1884	0.2654	2.2659	0.1154	0.1168
	(0.0244)	(0.0244)	(0.0328)	(0.0399)	(0.0488)	(0.1500)	(0.0251)	(0.0251)

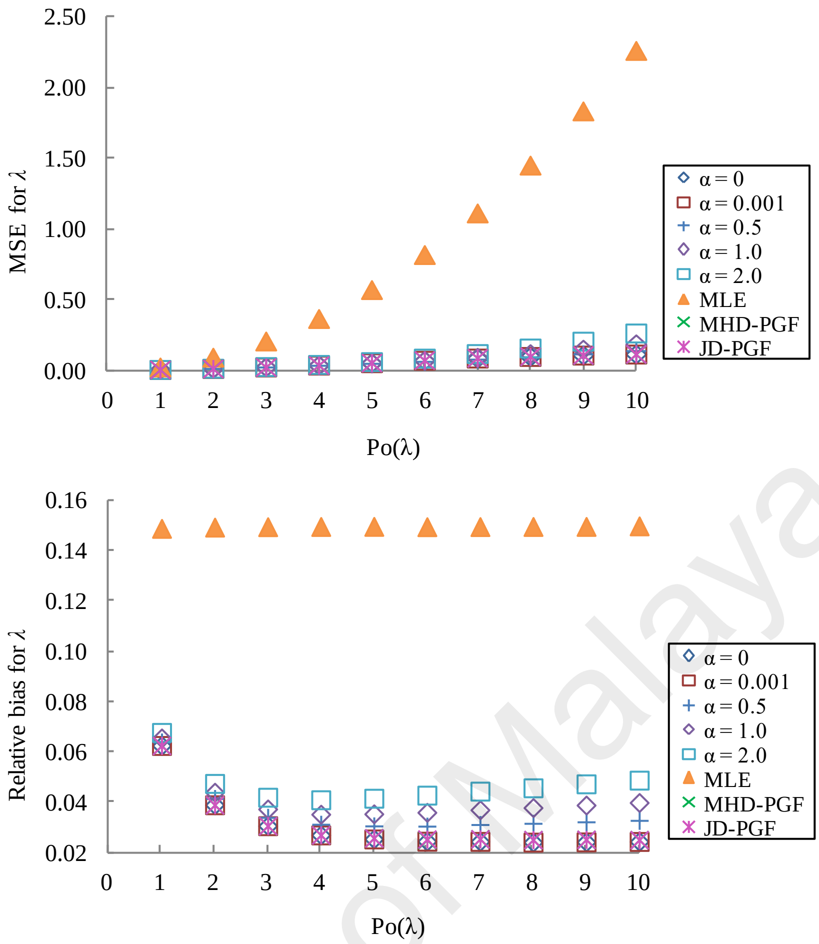

Figure 4.2: MSE and relative biases for estimators with samples of size $n = 500$, from $Po(\lambda)$, with 5% contamination.

With 5% contamination to the sample data, MLE is significantly affected in terms of its MSE and relative biases, as expected, due to the fact that MLE is not a robust estimator. BHHJ-PGF with smaller α values, however, perform better compare to those with greater α values. BHHJ-PGF(0) performs similarly, if not better than, JD-PGF and MHD-PGF.

Upon investigation of BHHJ-PGF for various α values in a Poisson distribution, it appears that α might serve as a tuning parameter, different from that of BHHJ, in the sense that here, greater α values perform better in samples without contamination, whereas smaller α values perform better in contaminated samples. A more comprehensive study involving multiple percentages of contamination and α values is carried out using the negative binomial distribution.

4.2 Simulation using NB distribution

Negative binomial distribution, with its pgf, $g_{(r,p)}(t) = (p/[1-(1-p)t])^r$, is selected for this section of simulation to represent a two-parameter discrete distribution. It has been successfully used to model data from a wide range of topics, ranging from accidents (Arbous & Kerrich, 1951), natural disasters (Banik & Kibria, 2009), and biology (Gurland, 1959), making it a versatile and important model. The parameters $r = 2$ and $p = 0.2$ are chosen arbitrarily for this section. Sample size is set at $n = 500$. Simulation is done for samples with and without outlier. Similar to the work in Sharifdoust et al. (2016), contamination is generated from Poisson distribution with $\lambda_c = 32$, which is four times the mean of NB(2, 0.2).

Several options (1000, 2000, 5000, 8000 and 10000) were considered before deciding on a fixed number of simulation runs. It was found that the difference between parameters estimated by 8000 runs and 10000 runs is within 10^{-4} (this is true for both parameters r and p), however the time elapsed increases by 20 to 50 minutes depending on the percentages of contamination. Therefore, simulation runs of 8000 is deemed sufficient and is adopted throughout the study involving NB distribution.

4.2.1 BHHJ-PGF(α) and other estimators

This section would convey how different α values affect the performance of BHHJ-PGF estimator in situation where there is no contamination and also when contamination is added in percentages of 1%, 5%, 10%, 20%, 30%, 40% and 50%.

The values of α, ranging from 0 to 2.0, are investigated here. Due to space constraint, only selected results are tabulated here in Table 4.3 and illustrated in Figure 4.3. The complete table of results can be found in Appendix C. Estimators are compared in terms of MSE and relative biases for the parameters r and p of the NB distribution.

Table 4.3: MSE and relative biases (in bracket) for estimators with samples of size $n = 500$, from $NB(r = 2.0, p = 0.2)$ for varying α and percentages of contamination.

Contamination	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.5)		BHHJ-PGF(1.0)		BHHJ-PGF(2.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p	r	p
0%	0.1141	0.0011	0.1141	0.0011	0.0945	0.0009	0.0942	0.0009	0.0738	0.0007	0.0274	0.0002	0.1144	0.0011	0.1142	0.0011
	(0.0270)	(0.0220)	(0.0270)	(0.0217)	(0.0228)	(0.0176)	(0.0152)	(0.0098)	(0.0092)	(0.0043)	(0.0080)	(0.0055)	(0.0275)	(0.0222)	(0.0275)	(0.0222)
1%	0.1085	0.0011	0.1082	0.0011	0.0895	0.0009	0.0880	0.0009	0.0722	0.0007	0.03058	0.00031	0.1092	0.0011	0.1087	0.0011
	(0.0165)	(0.0015)	(0.0165)	(0.0016)	(0.0105)	(-0.0050)	(0.0015)	(-0.0148)	(-0.0102)	(-0.0271)	(-0.0480)	(-0.0625)	(0.0170)	(0.0020)	(0.0175)	(0.0025)
5%	0.0959	0.0012	0.0958	0.0012	0.0804	0.0011	0.0802	0.0011	0.0930	0.0014	0.1565	0.0026	0.0963	0.0012	0.0963	0.0012
	(-0.0230)	(-0.0765)	(-0.0230)	(-0.0765)	(-0.0350)	(-0.0913)	(-0.0498)	(-0.1083)	(-0.0774)	(-0.1399)	(-0.1905)	(-0.2525)	(-0.0228)	(-0.0763)	(-0.0230)	(-0.0764)
10%	0.0997	0.0019	0.0997	0.0019	0.0931	0.0020	0.1061	0.0024	0.1487	0.0033	0.3200	0.0062	0.1015	0.0019	0.1000	0.0019
	(-0.0670)	(-0.1669)	(-0.0675)	(-0.1669)	(-0.0857)	(-0.1895)	(-0.1068)	(-0.2145)	(-0.1461)	(-0.2600)	(-0.2795)	(-0.3920)	(-0.0675)	(-0.1671)	(-0.0670)	(-0.1668)
20%	0.1460	0.0048	0.1459	0.0048	0.1631	0.0055	0.1934	0.0064	0.2746	0.0080	0.4834	0.0115	0.1461	0.0048	0.1460	0.0048
	(-0.1415)	(-0.3259)	(-0.1415)	(-0.3260)	(-0.1681)	(-0.3585)	(-0.1939)	(-0.3885)	(-0.2397)	(-0.4395)	(-0.3455)	(-0.5366)	(-0.1413)	(-0.3257)	(-0.1415)	(-0.3258)
30%	0.2143	0.0088	0.2143	0.0088	0.2543	0.0100	0.3016	0.0111	0.4452	0.0135	0.4823	0.0146	0.2146	0.0088	0.2143	0.0088
	(-0.2000)	(-0.4586)	(-0.2000)	(-0.4586)	(-0.2303)	(-0.4930)	(-0.2573)	(-0.5222)	(-0.3100)	(-0.5730)	(-0.3454)	(-0.6037)	(-0.2000)	(-0.4310)	(-0.2000)	(-0.4585)
40%	0.2854	0.0132	0.2855	0.0132	0.3302	0.0144	0.3904	0.0156	0.5647	0.0179	0.3720	0.0160	0.2881	0.0132	0.2853	0.0132
	(-0.2446)	(-0.5673)	(-0.2447)	(-0.5674)	(-0.2726)	(-0.5969)	(-0.2988)	(-0.6000)	(-0.3508)	(-0.6638)	(-0.3025)	(-0.6330)	(-0.2451)	(-0.5677)	(-0.2444)	(-0.5671)
50%	0.3462	0.0174	0.3463	0.0174	0.3931	0.0185	0.4578	0.0195	0.6198	0.0213	0.1919	0.0162	0.3470	0.0174	0.3459	0.0174
	(-0.2764)	(-0.6553)	(-0.2764)	(-0.6553)	(-0.3007)	(-0.6778)	(-0.3236)	(-0.6958)	(-0.3681)	(-0.7256)	(-0.2140)	(-0.6359)	(-0.2763)	(-0.6553)	(-0.2760)	(-0.6550)

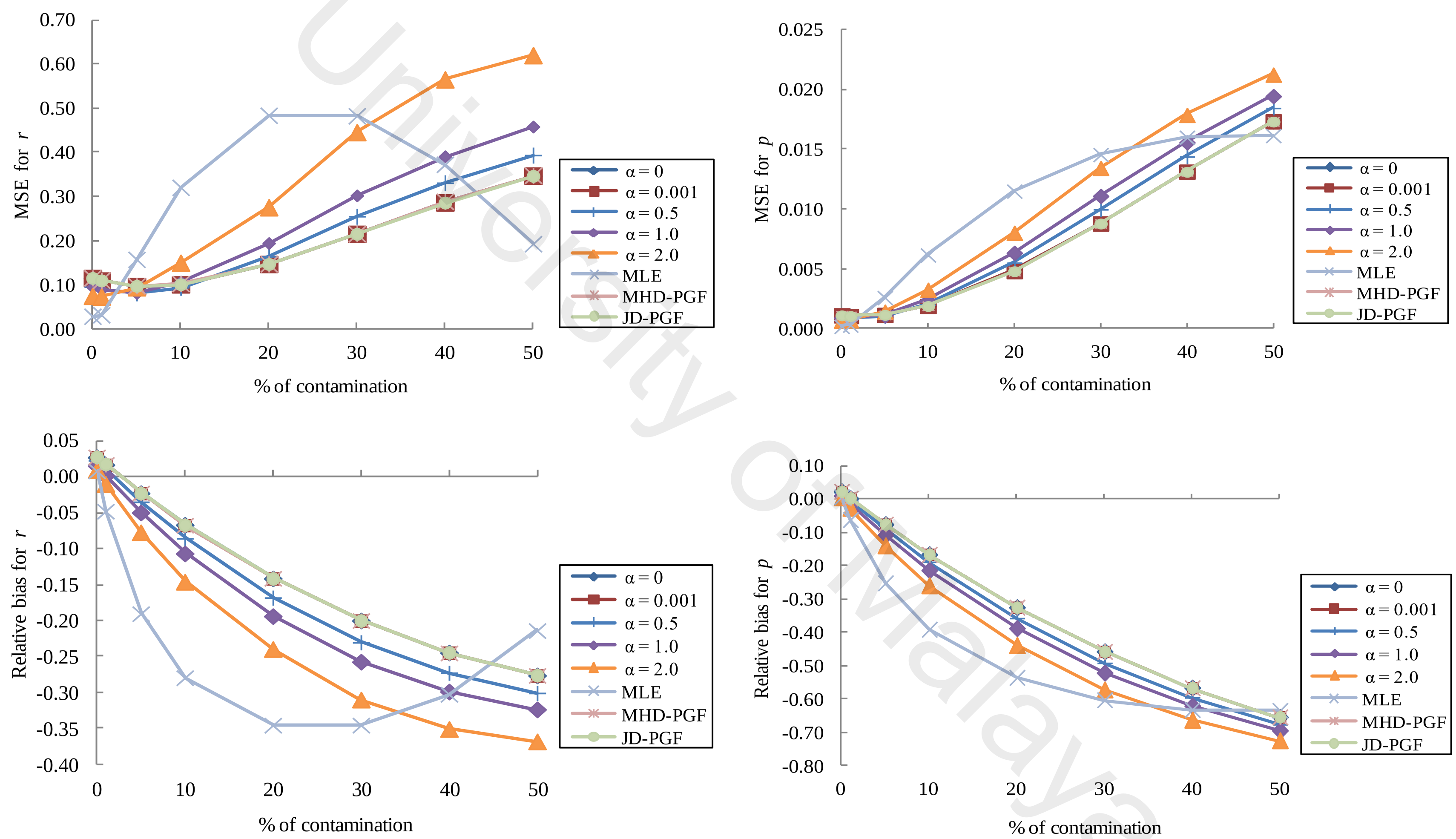

Figure 4.3: MSE and relative biases for estimators with samples of size $n = 500$, from $NB(r = 2.0, p = 0.2)$ for varying α and percentages of contamination.

At 0% contamination, among all estimators, MLE has the best performance for both parameters. BHHJ-PGF with greater α values result in better estimates, nearer to that of MLE, compared to smaller α values. MHD-PGF and JD-PGF are almost identical to each other in terms of their MSEs and relative biases, which resonates with BHHJ-PGF(0). All estimators perform similarly to the case without contamination when 1% of outlier is introduced to the samples.

When contamination is introduced at 5%, MLE is visibly affected as expected. It is known that MLE is an efficient estimator but unable to handle the existence of outlier in the data. It can be observed that as the percentage of contamination increases (5%-50%), the MSEs and relative biases of BHHJ-PGF estimates decrease when the value of α becomes smaller.

The advantage of BHHJ-PGF(α) occurs in its ability to produce good estimates when there is no outlier by using greater α values, as well as robust estimates in the presence of contamination through the use of smaller α values.

4.2.2 Simulation using different sample size

In this section, the performance of proposed estimator BHHJ-PGF(α) is examined in various sample sizes and $\alpha = 0, 0.001, 0.5, 1.0$. A range from $n = 20$, representing small sample size, to $n = 1000$, representing large sample size is considered. The number of simulation runs is set to remain as 8000. Simulation is carried out for samples without contamination as well as with 5%, 10%, 20% and 30% contamination. Also included are the performance of MHD-PGF, JD-PGF and MLE for comparison purposes. Simulations are done on the sample data with 40% and 50% contamination levels. However, it is noted that at that extreme percentage of contamination, parameter estimation ceases to bring meaningful conclusion to the sample data.

4.2.2.1 Sample data without contamination

In the absence of outlier, represented by 0% contamination, MLE outperformed the other estimators by giving the least MSEs and relative biases, as expected for all sample sizes. A comparison for BHHJ-PGF with different α values results in the observation that $\alpha = 1$ is generally a better estimator throughout all the sample sizes. As sample size increases, all pgf-based estimators have similar performances although BHHJ-PGF estimator still gives slightly smaller values for MSEs and relative biases. This shows numerically that BHHJ-PGF estimators produce consistent estimates in that as the sample size increases, the estimates approach the true parameter values. The results are tabulated in Table 4.4 and presented in Figure 4.4.

Table 4.4: MSE and relative biases (in bracket) for estimators with samples from $NB(r = 2.0, p = 0.2)$, without contamination.

Sample sizes	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.5)		BHHJ-PGF(1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
20	907.0287	0.1192	860.0037	0.1192	102.7799	0.0701	23.7872	0.0432	88.7604	0.0099	582.4183	0.1225	1120.3530	0.1240
	(5.7016)	(0.9192)	(5.6339)	(0.9191)	(2.0146)	(0.6505)	(1.0460)	(0.4683)	(0.3589)	(0.1774)	(5.1604)	(0.9396)	(6.2105)	(0.9468)
50	52.4371	0.0334	60.0962	0.0334	7.6346	0.0194	3.1488	0.0128	0.4881	0.0026	46.1961	0.0347	66.0039	0.0349
	(0.9164)	(0.3235)	(0.9251)	(0.3237)	(0.4355)	(0.2318)	(0.2852)	(0.1707)	(0.0900)	(0.0629)	(0.9392)	(0.3304)	(1.0006)	(0.3329)
100	3.0943	0.0100	3.3493	0.0100	1.2440	0.0067	0.7377	0.0049	0.1664	0.0011	3.6364	0.0103	4.4934	0.0103
	(0.2181)	(0.1340)	(0.2207)	(0.1340)	(0.1511)	(0.1022)	(0.1138)	(0.0784)	(0.0405)	(0.0285)	(0.2271)	(0.1362)	(0.2314)	(0.1364)
200	0.4512	0.0035	0.4498	0.0035	0.3174	0.0026	0.2556	0.0021	0.0711	0.0005	0.4565	0.0035	0.4560	0.0035
	(0.0755)	(0.0567)	(0.0755)	(0.0568)	(0.0602)	(0.0450)	(0.0465)	(0.0329)	(0.0180)	(0.0125)	(0.0760)	(0.0573)	(0.0760)	(0.0575)
300	0.2260	0.0021	0.2259	0.0021	0.1780	0.0016	0.1564	0.0014	0.0463	0.0003	0.2266	0.0021	0.2264	0.0021
	(0.0495)	(0.0395)	(0.0490)	(0.0393)	(0.0410)	(0.0321)	(0.0314)	(0.0228)	(0.0125)	(0.0090)	(0.0500)	(0.0403)	(0.0500)	(0.0401)
500	0.1141	0.0011	0.1141	0.0011	0.0945	0.0009	0.0942	0.0009	0.0274	0.0002	0.1144	0.0011	0.1142	0.0011
	(0.0270)	(0.0220)	(0.0270)	(0.0217)	(0.0228)	(0.0176)	(0.0152)	(0.0098)	(0.0080)	(0.0055)	(0.0275)	(0.0222)	(0.0275)	(0.0222)
1000	0.0509	0.0005	0.0511	0.0005	0.0432	0.0004	0.0575	0.0006	0.0133	0.0001	0.0510	0.0005	0.0511	0.0005
	(0.0120)	(0.0095)	(0.0120)	(0.0095)	(0.0100)	(0.0077)	(0.0025)	(-0.0005)	(0.0035)	(0.0025)	(0.0120)	(0.0100)	(0.0120)	(0.0098)

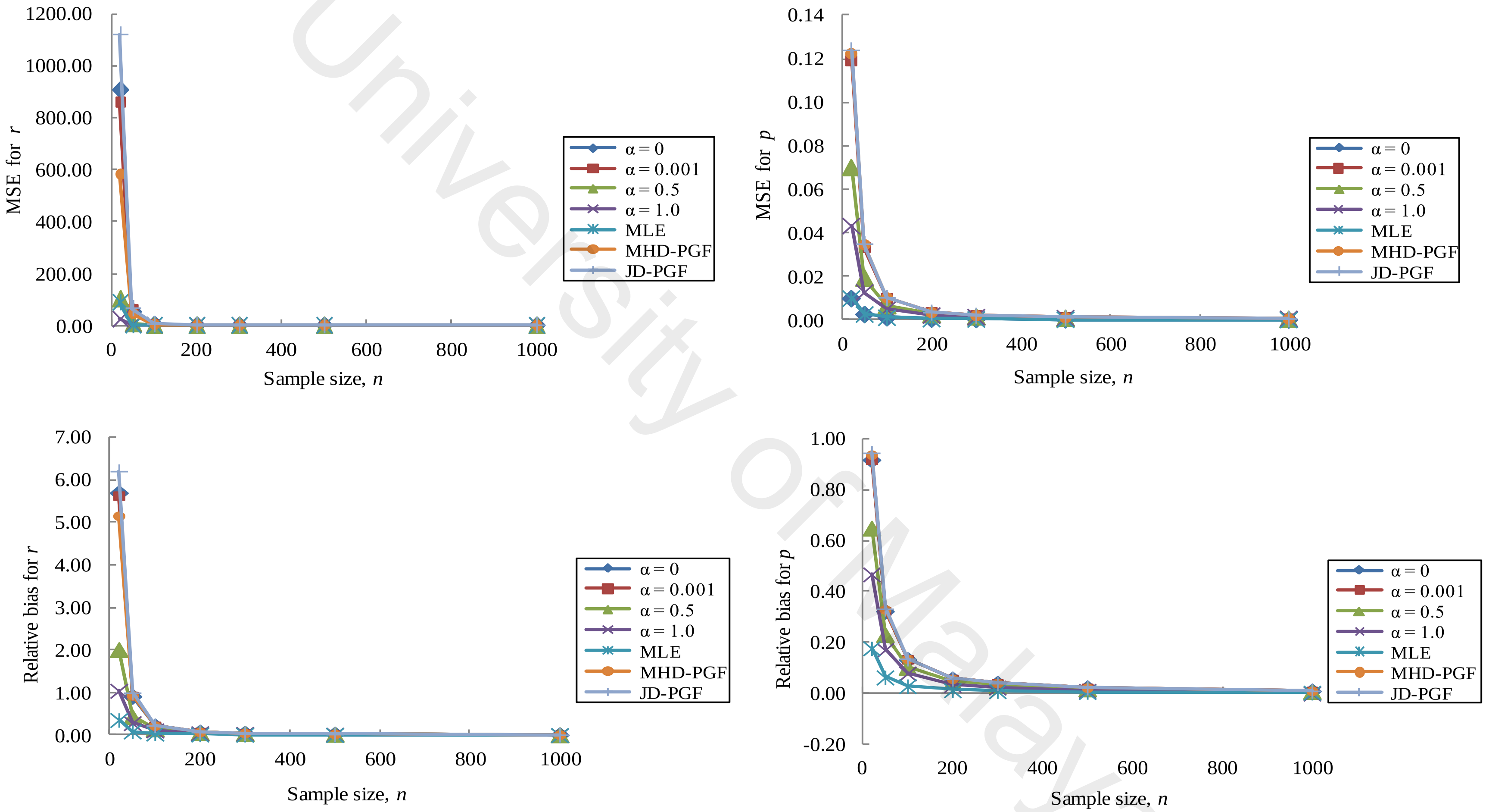

Figure 4.4: MSE and relative biases for estimators with samples from $NB(r = 2.0, p = 0.2)$, without contamination.

4.2.2.2 Sample data with contamination

Tables 4.5, 4.6, 4.7 and 4.8 as well as Figures 4.5, 4.6, 4.7 and 4.8 present the results for this section. When 5% contamination is introduced, MLE estimates are affected particularly in terms of its relative biases for the estimator of p. For small sample size, $n = 20$, an obvious fluctuation can be seen on the MSE and biases of MLE compared to the case of 0% contamination (Section 4.2.2.1). MSE is negatively biased throughout all sample sizes.

Among BHHJ-PGF estimators, it is found that large α value, that is BHHJ-PGF(1.0) performs better for smaller sample sizes ($n \leq 50$). As sample size increases from $n = 100$, better estimates are given by smaller α in terms of relative biases. Hence, BHHJ-PGF with large α is more robust for small sample size. Comparing all pgf-based estimators, BHHJ-PGF performs better for both parameters r and p when the sample size is small ($n < 100$) and comparably with, if not better than, MHD-PGF and JD-PGF as the sample size increases. This is true for the cases of 0%-30% contamination. BHHJ-PGF seems to be able to fill the gap of estimating parameter where the sample size is small with the presence of outliers.

Table 4.5: MSE and relative biases (in bracket) for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 5% contamination.

Sample sizes	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.5)		BHHJ-PGF(1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
20	562.0928	0.1023	611.4556	0.1023	51.8780	0.0536	12.0376	0.0303	0.5055	0.0036	411.2909	0.1061	835.7550	0.1082
	(4.3100)	(0.7886)	(4.4144)	(0.78815)	(1.4264)	(0.4884)	(0.7237)	(0.2972)	(-0.0820)	(-0.1775)	(4.1320)	(0.8131)	(5.0182)	(0.8235)
50	30.0010	0.0268	32.5328	0.0268	3.8504	0.0139	1.7299	0.0087	0.2499	0.0034	26.7930	0.0283	54.4970	0.0327
	(0.6549)	(0.1935)	(0.6593)	(0.1934)	(0.2767)	(0.0825)	(0.1496)	(0.0088)	(-0.1811)	(-0.2632)	(0.6829)	(0.2024)	(0.8964)	(0.2897)
100	2.1680	0.0081	2.1417	0.0080	0.8369	0.0051	0.5086	0.0037	0.1843	0.0027	2.4139	0.0083	3.0276	0.0094
	(0.1450)	(0.0307)	(0.1444)	(0.0305)	(0.0758)	(-0.0125)	(0.0317)	(-0.0490)	(-0.1753)	(-0.2415)	(0.1505)	(0.0320)	(0.1971)	(0.0945)
200	0.3488	0.0029	0.3474	0.0029	0.2441	0.0022	0.1986	0.0020	0.1672	0.0027	0.3540	0.0029	0.4065	0.0031
	(0.0218)	(-0.0422)	(0.0218)	(-0.0420)	(-0.0013)	(-0.0651)	(-0.0239)	(-0.0888)	(-0.1860)	(-0.2490)	(0.0223)	(-0.0418)	(0.0539)	(0.0170)
300	0.1832	0.0018	0.1831	0.0018	0.1421	0.0015	0.1268	0.0015	0.1613	0.0026	0.1846	0.0019	0.2053	0.0019
	(-0.0030)	(-0.0599)	(-0.0025)	(-0.0600)	(-0.0196)	(-0.0786)	(-0.0376)	(-0.0987)	(-0.1885)	(-0.2510)	(-0.0027)	(-0.0599)	(0.0279)	(-0.0010)
500	0.0959	0.0012	0.0958	0.0012	0.0804	0.0011	0.0802	0.0011	0.1565	0.0026	0.0963	0.0012	0.1037	0.0010
	(-0.0230)	(-0.0765)	(-0.0230)	(-0.0765)	(-0.0350)	(-0.0913)	(-0.0498)	(-0.1083)	(-0.1905)	(-0.2525)	(-0.0228)	(-0.0763)	(0.0067)	(-0.0182)
1000	0.0491	0.0008	0.0490	0.0008	0.0449	0.0008	0.0496	0.0009	0.1539	0.0026	0.0490	0.0008	0.0479	0.0005
	(-0.0365)	(-0.0879)	(-0.0365)	(-0.0880)	(-0.0467)	(-0.1007)	(-0.0580)	(-0.1145)	(-0.1925)	(-0.2535)	(-0.0365)	(-0.0875)	(-0.0075)	(-0.0300)

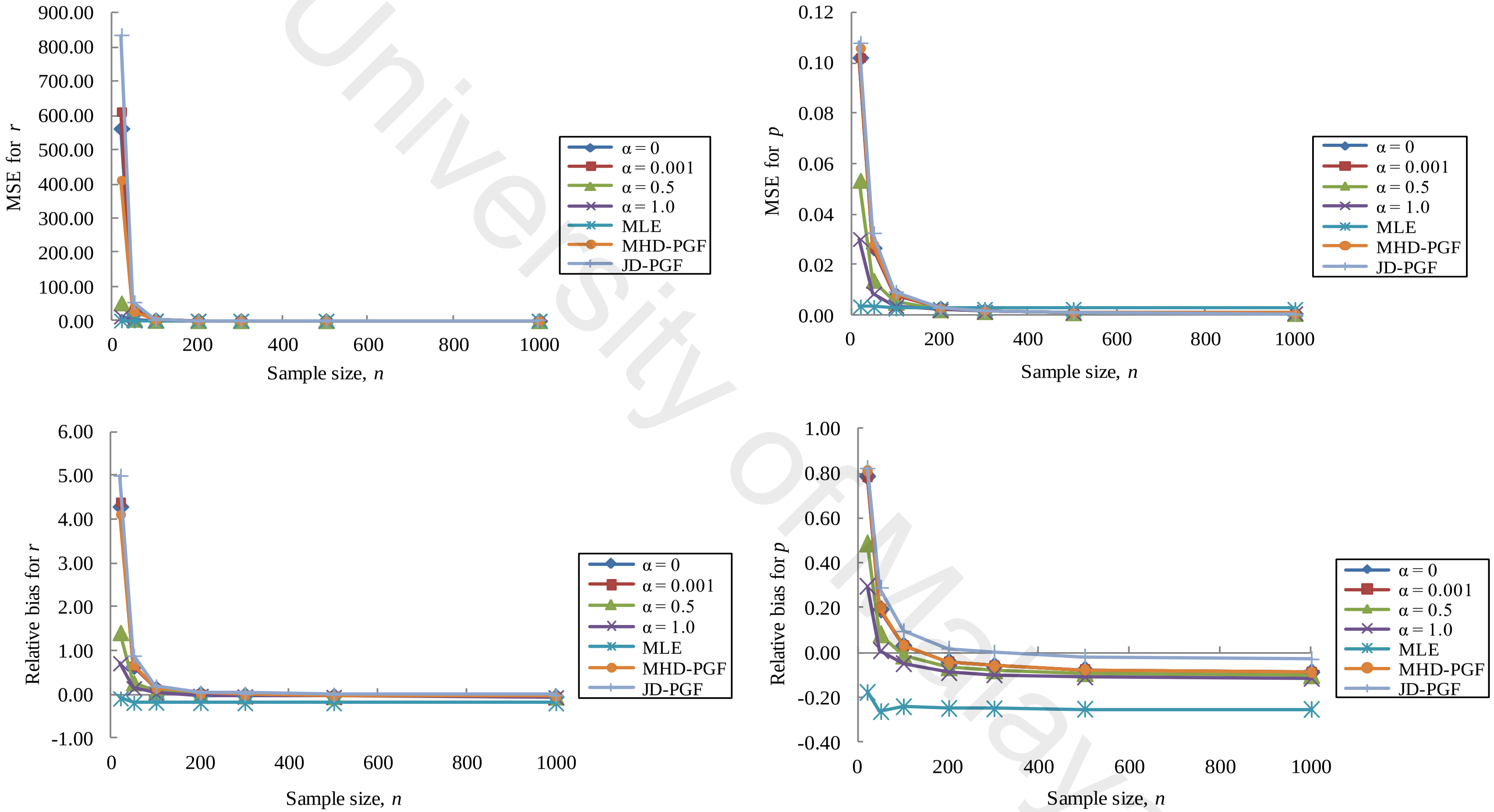

Figure 4.5: MSE and relative biases for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 5% contamination.

Table 4.6: MSE and relative biases (in bracket) for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 10% contamination.

Sample sizes	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.5)		BHHJ-PGF(1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
20	356.9423	0.0856	352.2816	0.0854	25.2626	0.0397	6.6144	0.0211	0.4551	0.0059	274.8914	0.0904	535.6973	0.0927
	(3.2861)	(0.6582)	(3.2615)	(0.6569)	(1.0139)	(0.3344)	(0.4950)	(0.1387)	(-0.2055)	(-0.3437)	(3.2681)	(0.6882)	(3.8487)	(0.7030)
50	20.1875	0.0229	19.2920	0.0229	2.8180	0.0116	1.2191	0.0072	0.3452	0.0060	17.5311	0.0244	21.3574	0.0248
	(0.5083)	(0.1110)	(0.5077)	(0.1110)	(0.1984)	(-0.0074)	(0.0756)	(-0.0885)	(-0.2550)	(-0.3755)	(0.5406)	(0.1215)	(0.5630)	(0.1250)
100	1.5030	0.0072	1.4927	0.0072	0.6337	0.0048	0.3964	0.0039	0.3283	0.0061	1.6575	0.0075	1.6868	0.0075
	(0.0835)	(-0.0638)	(0.0835)	(-0.0641)	(0.0127)	(-0.1170)	(-0.0347)	(-0.1604)	(-0.2688)	(-0.3848)	(0.0894)	(-0.0616)	(0.0910)	(-0.0610)
200	0.3011	0.0031	0.3009	0.0031	0.2153	0.0028	0.1832	0.0028	0.3235	0.0062	0.3067	0.0032	0.3062	0.0032
	(-0.0250)	(-0.1333)	(-0.0249)	(-0.1335)	(-0.0541)	(-0.1649)	(-0.0813)	(-0.1947)	(-0.2755)	(-0.3895)	(-0.0244)	(-0.1331)	(-0.0243)	(-0.1330)
300	0.1670	0.0024	0.1670	0.0024	0.1378	0.0023	0.1358	0.0025	0.3216	0.0062	0.1681	0.0024	0.1682	0.0024
	(-0.0490)	(-0.1516)	(-0.0490)	(-0.1518)	(-0.0718)	(-0.1783)	(-0.0955)	(-0.2055)	(-0.2780)	(-0.3910)	(-0.0487)	(-0.1513)	(-0.0490)	(-0.1516)
500	0.0997	0.0019	0.0997	0.0019	0.0931	0.0020	0.1061	0.0024	0.3200	0.0062	0.1015	0.0019	0.1000	0.0019
	(-0.0670)	(-0.1669)	(-0.0675)	(-0.1669)	(-0.0857)	(-0.1895)	(-0.1068)	(-0.2145)	(-0.2795)	(-0.3920)	(-0.0675)	(-0.1671)	(-0.0670)	(-0.1668)
1000	0.0644	0.0017	0.0645	0.0017	0.0669	0.0018	0.0893	0.0023	0.3199	0.0062	0.0640	0.0016	0.0652	0.0017
	(-0.0800)	(-0.1781)	(-0.0800)	(-0.1781)	(-0.0956)	(-0.1978)	(-0.1152)	(-0.2215)	(-0.2810)	(-0.3930)	(-0.0798)	(-0.1775)	(-0.0805)	(-0.1785)

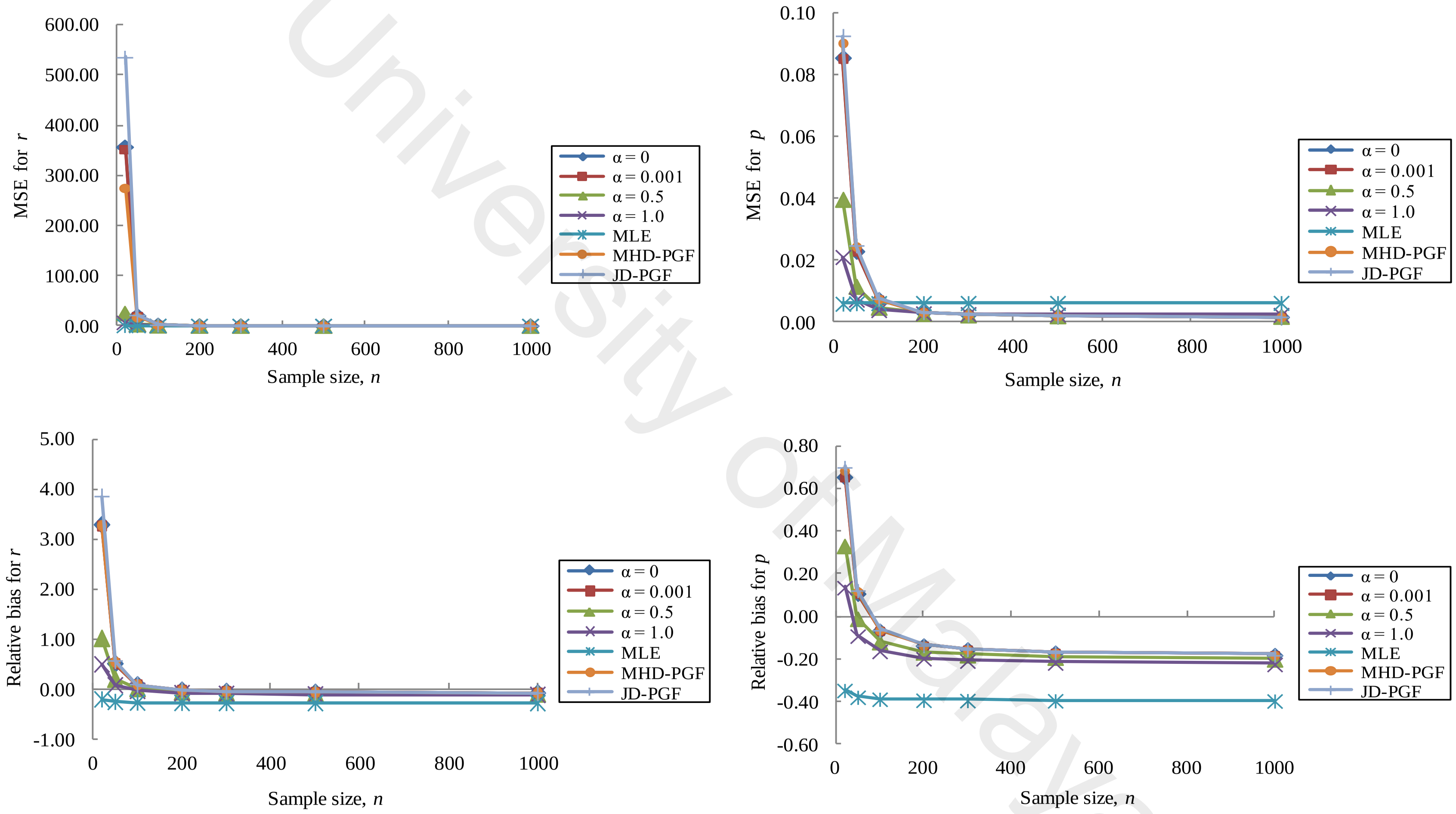

Figure 4.6: MSE and relative biases for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 10% contamination.

Table 4.7: MSE and relative biases (in bracket) for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 20% contamination.

Sample sizes	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.5)		BHHJ-PGF(1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
20	107.4564	0.0550	99.3802	0.0548	7.5956	0.0214	2.4628	0.0115	0.5305	0.0109	106.4352	0.0607	169.7528	0.0630
	(1.7543)	(0.3933)	(1.7275)	(0.3922)	(0.5372)	(0.0652)	(0.2056)	(-0.1246)	(-0.2930)	(-0.5073)	(1.9376)	(0.4361)	(2.1750)	(0.4514)
50	4.7784	0.0162	4.7478	0.0162	1.3607	0.0091	0.6921	0.0074	0.4858	0.0112	5.9416	0.0175	6.2128	0.0177
	(0.2558)	(-0.0803)	(0.2551)	(-0.0805)	(0.0446)	(-0.2084)	(-0.0598)	(-0.2919)	(-0.3263)	(-0.5255)	(0.2899)	(-0.0678)	(0.2979)	(-0.0646)
100	0.8706	0.0073	0.8691	0.0073	0.4420	0.0063	0.3311	0.0066	0.4850	0.0114	0.9499	0.0076	0.9641	0.0076
	(-0.0203)	(-0.2360)	(-0.0205)	(-0.2362)	(-0.0915)	(-0.2995)	(-0.1415)	(-0.3487)	(-0.3382)	(-0.5325)	(-0.0154)	(-0.2334)	(-0.0142)	(-0.2328)
200	0.2832	0.0053	0.2829	0.0053	0.2339	0.0057	0.2359	0.0064	0.4835	0.0115	0.2894	0.0054	0.2900	0.0053
	(-0.1045)	(-0.2962)	(-0.1046)	(-0.2963)	(-0.1426)	(-0.3381)	(-0.1759)	(-0.3745)	(-0.3427)	(-0.5349)	(-0.1042)	(-0.2960)	(-0.1037)	(-0.2956)
300	0.1886	0.0050	0.1889	0.0050	0.1890	0.0056	0.2078	0.0064	0.4840	0.0115	0.1902	0.0050	0.1894	0.0050
	(-0.1270)	(-0.3139)	(-0.1272)	(-0.3141)	(-0.1579)	(-0.3500)	(-0.1861)	(-0.3825)	(-0.3445)	(-0.5360)	(-0.1270)	(-0.3138)	(-0.1267)	(-0.3136)
500	0.1460	0.0048	0.1459	0.0048	0.1631	0.0055	0.1934	0.0064	0.4834	0.0115	0.1461	0.0048	0.1460	0.0048
	(-0.1415)	(-0.3259)	(-0.1415)	(-0.3260)	(-0.1681)	(-0.3585)	(-0.1939)	(-0.3885)	(-0.3455)	(-0.5366)	(-0.1413)	(-0.3257)	(-0.1415)	(-0.3258)
1000	0.1239	0.0048	0.1238	0.0048	0.1490	0.0055	0.1831	0.0064	0.4837	0.0115	0.1244	0.0048	0.1242	0.0048
	(-0.1530)	(-0.3358)	(-0.1530)	(-0.3357)	(-0.1765)	(-0.3650)	(-0.1999)	(-0.3930)	(-0.3470)	(-0.5371)	(-0.1531)	(-0.3359)	(-0.1530)	(-0.3360)

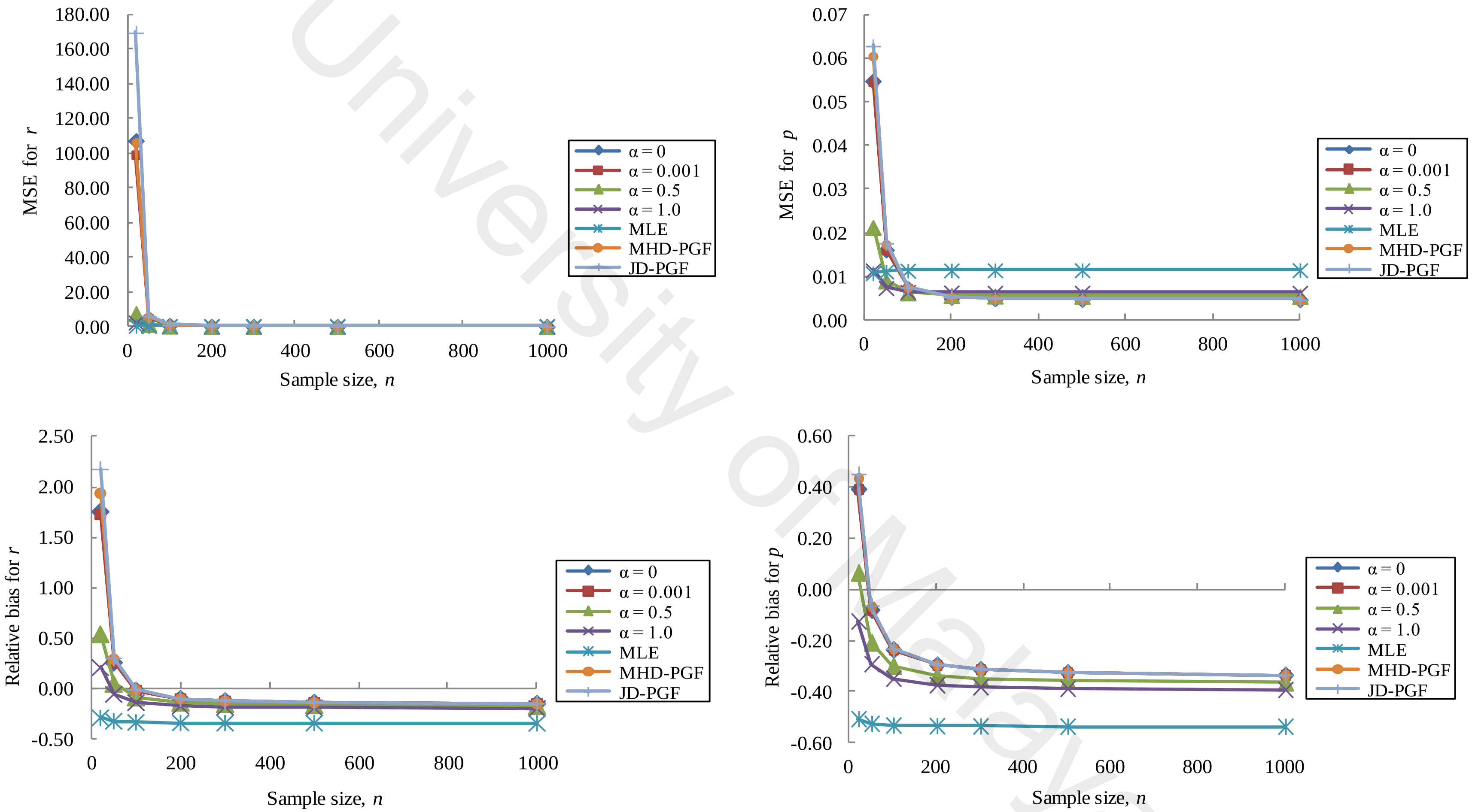

Figure 4.7: MSE and relative biases for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 20% contamination.

Table 4.8: MSE and relative biases (in bracket) for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 30% contamination.

Sample sizes	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.5)		BHHJ-PGF(1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
20	23.1500	0.0328	23.2093	0.0327	3.2966	0.0132	1.3064	0.0102	0.5282	0.0139	31.6176	0.0374	44.1124	0.0397
	(0.9629)	(0.1482)	(0.9605)	(0.1473)	(0.2804)	(-0.1616)	(0.0360)	(-0.3290)	(-0.2970)	(-0.5809)	(1.1230)	(0.1932)	(1.2386)	(0.2146)
50	2.5732	0.0134	2.5630	0.0134	0.8879	0.0101	0.5444	0.0104	0.4860	0.0143	3.0446	0.0143	3.2473	0.0146
	(0.1214)	(-0.2419)	(0.1206)	(-0.2424)	(-0.0560)	(-0.3691)	(-0.1505)	(-0.4485)	(-0.3268)	(-0.5947)	(0.1483)	(-0.2281)	(0.1568)	(-0.2242)
100	0.6630	0.0094	0.6620	0.0095	0.4101	0.0097	0.3745	0.0108	0.4845	0.0145	0.7187	0.0096	0.7269	0.0096
	(-0.0951)	(-0.3793)	(-0.0954)	(-0.3796)	(-0.1661)	(-0.4449)	(-0.2167)	(-0.4935)	(-0.3382)	(-0.6003)	(-0.0905)	(-0.3765)	(-0.0887)	(-0.3752)
200	0.3015	0.0088	0.3013	0.0088	0.2912	0.0098	0.3227	0.0110	0.4824	0.0145	0.3079	0.0088	0.3072	0.0088
	(-0.1657)	(-0.4312)	(-0.1659)	(-0.4314)	(-0.2073)	(-0.4755)	(-0.2421)	(-0.5115)	(-0.3424)	(-0.6024)	(-0.1654)	(-0.4310)	(-0.1647)	(-0.4304)
300	0.2424	0.0088	0.2425	0.0088	0.2659	0.0099	0.3082	0.0111	0.4827	0.0146	0.2445	0.0088	0.2432	0.0088
	(-0.1873)	(-0.4482)	(-0.1874)	(-0.4484)	(-0.2210)	(-0.4860)	(-0.2511)	(-0.5178)	(-0.3442)	(-0.6033)	(-0.1874)	(-0.4483)	(-0.1871)	(-0.4481)
500	0.2143	0.0088	0.2143	0.0088	0.2543	0.0100	0.3016	0.0111	0.4823	0.0146	0.2146	0.0088	0.2143	0.0088
	(-0.2000)	(-0.4586)	(-0.2000)	(-0.4586)	(-0.2303)	(-0.4930)	(-0.2573)	(-0.5222)	(-0.3454)	(-0.6037)	(-0.200)	(-0.4585)	(-0.200)	(-0.4585)
1000	0.2023	0.0089	0.2024	0.0089	0.2552	0.0102	0.2964	0.0112	0.4826	0.0146	0.2025	0.0089	0.2023	0.0089
	(-0.2105)	(-0.4674)	(-0.2105)	(-0.4675)	(-0.2393)	(-0.5002)	(-0.2622)	(-0.5257)	(-0.3464)	(-0.6042)	(-0.2106)	(-0.4674)	(-0.2105)	(-0.4674)

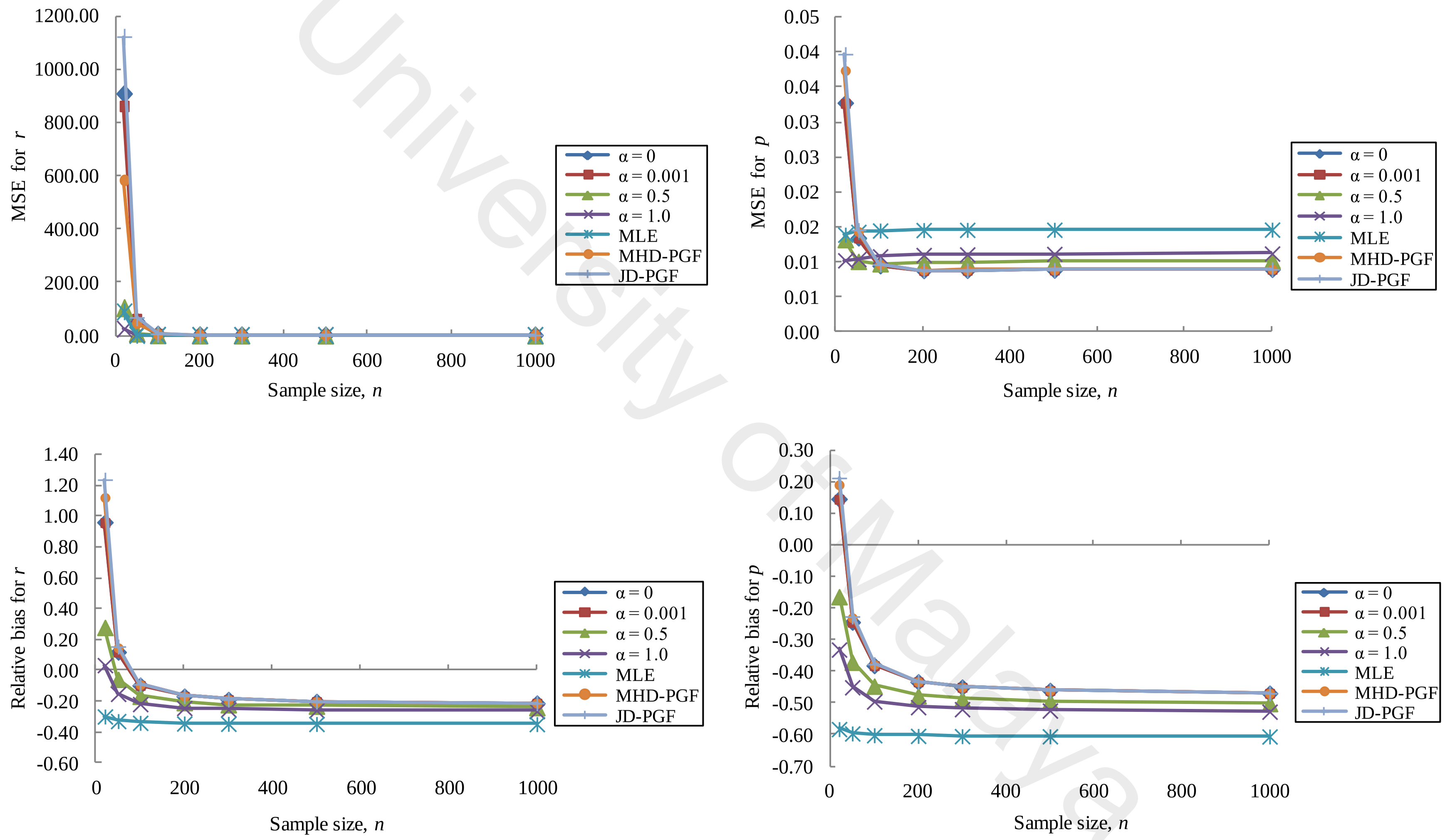

Figure 4.8: MSE and relative biases for estimators with samples from $NB(r = 2.0, p = 0.2)$, with 30% contamination.

4.2.3 Simulation with different parameter values for NB distribution

It is of interest to look into the performance of estimators for different parameter values for the NB(r, p) distribution. In addition to long-tailed $(p < 0.5)$ distribution NB(2,0.2) considered above, simulations using NB(32,0.8) and NB(8,0.5) to represent short-tailed $(p > 0.5)$ and moderate-tailed distribution, respectively, are performed for BHHJ-PGF(0.001), BHHJ-PGF(0.5), BHHJ-PGF(1.0), MHD-PGF, JD-PGF and also MLE. The sample size, n, is fixed at 500, and the number of simulation runs is set to be 8000. All three NB distributions are chosen to have the same mean, which is 8.

4.2.3.1 Sample data without contamination

At 0% contamination, all six estimators perform generally better for moderate- and long-tailed distributions compared to their performances in short-tailed distribution, NB(32,0.8) as shown in Table 4.9. Also, for short-tailed distribution, the values of MSEs and relative biases for all estimators involve a larger range. It can be observed that the estimators BHHJ-PGF(0.001), MHD-PGF and JD-PGF tend to produce resembling trend, with MHD-PGF performing better than its counterparts in short-tailed distribution. Comparing between MLE, BHHJ-PGF(0.5) and BHHJ-PGF(1.0), MLE has a generally better performance.

Among the proposed BHHJ-PGF estimator, the one with $\alpha = 1.0$ performs better than those with $\alpha = 0.5$ and $\alpha = 0.001$. This is true across moderate- and long-tailed cases, whereas in the short-tailed case, BHHJ-PGF with $\alpha = 0.5$ performs better.

Table 4.9: MSE and relative biases (in bracket) for estimation with samples of size $n = 500$ and different parameter values of $NB(r, p)$, without contamination.

Distributions	BHHJ-PGF (0.001)		BHHJ-PGF (0.5)		BHHJ-PGF (1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p
NB(32,0.8)	2193.9510	0.0193	249.3554	0.0256	260.2931	0.0346	413.5365	0.0027	803.7114	0.0188	1982.0780	0.0200
	(0.6263)	(-0.0179)	(-0.1695)	(-0.1036)	(-0.4570)	(-0.1870)	(0.1547)	(0.0045)	(0.2820)	(-0.0361)	(0.5158)	(-0.0305)
NB(8,0.5)	60.6740	0.0157	13.4630	0.0084	5.9231	0.0076	1.2426	0.0011	62.3583	0.0163	66.1363	0.0157
	(0.3239)	(0.0612)	(0.1255)	(0.0245)	(0.0085)	(-0.0240)	(0.0206)	(0.0058)	(0.3290)	(0.0597)	(0.3230)	(0.0584)
NB(2,0.2)	0.1141	0.0011	0.0945	0.0009	0.0942	0.0009	0.0274	0.0002	0.1144	0.0011	0.1142	0.0011
	(0.0270)	(0.0217)	(0.0228)	(0.0176)	(0.0152)	(0.0098)	(0.0080)	(0.0055)	(0.0275)	(0.0222)	(0.0275)	(0.0222)

4.2.3.2 Sample data with contamination

The performance of the proposed estimators in sample data with 1% and 5% outliers are investigated. Results of simulations with contaminated data are presented in Table 4.10 and Table 4.11 for 1% and 5% contamination, respectively.

In the case of 1% contamination, for the three different types of NB distributions, all pgf-based estimation methods behave similarly to the case without contamination. However, MLE is affected throughout short-, moderate- and long-tailed NB distribution when 1% of outlier is introduced to the sample data. MLE is more drastically impacted compared to pgf-based estimators, when the percentage of contamination is increased from 1% to 5%. This can be observed from the resulting estimates that give larger relative biases.

BHHJ-PGF(0.001) appears to be performing better than BHHJ-PGF(0.5) and BHHJ-PGF(1.0) when the percentage of contamination increases from 1% to 5%. It suggests that BHHJ-PGF with smaller α value is more robust when the percentage of contamination increases. This corresponds to the results in Table 4.3, where BHHJ-PGF with larger α values gives better performance when the percentage of contamination is small and as the contamination increase, better estimates are given by BHHJ-PGF with smaller α values.

Table 4.10: MSE and relative biases (in bracket) for estimation with samples of size $n = 500$ and different parameter values of $NB(r, p)$, with 1% contamination.

Distributions	BHHJ-PGF (0.001)		BHHJ-PGF (0.5)		BHHJ-PGF (1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p
NB(32,0.8)	1863.2400	0.0203	231.1296	0.0271	288.4844	0.0401	379.4143	0.0405	719.0645	0.0193	1751.7220	0.0202
	(0.5289)	(-0.0286)	(-0.2237)	(-0.1178)	(-0.4940)	(-0.2086)	(-0.6058)	(-0.2478)	(0.2327)	(-0.0440)	(0.4401)	(-0.0388)
NB(8,0.5)	51.6393	0.0148	10.9034	0.0081	4.9240	0.0072	4.0138	0.0064	51.5231	0.0151	53.5240	0.0151
	(0.2739)	(0.0424)	(0.0794)	(0.0003)	(-0.0374)	(-0.0510)	(-0.2378)	(-0.1516)	(0.2783)	(0.0422)	(0.2776)	(0.0412)
NB(2,0.2)	0.1082	0.0011	0.0895	0.0009	0.0880	0.0009	0.0306	0.0003	0.1092	0.0011	0.1087	0.0011
	(0.0165)	(0.0016)	(0.0105)	(-0.0050)	(0.0015)	(-0.0148)	(-0.0480)	(-0.0625)	(0.0170)	(0.0020)	(0.0175)	(0.0025)

Table 4.11: MSE and relative biases (in bracket) for estimation with samples of size $n = 500$ and different parameter values of $NB(r, p)$, with 5% contamination.

Distributions	BHHJ-PGF (0.001)		BHHJ-PGF (0.5)		BHHJ-PGF (1.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p
NB(32,0.8)	894.8013	0.0242	236.8618	0.0312	384.7989	0.0586	737.1935	0.2074	457.3844	0.0258	892.3161	0.0247
	(0.1553)	(-0.0766)	(-0.3976)	(-0.1732)	(-0.6009)	(-0.2793)	(-0.8484)	(-0.5689)	(0.0224)	(-0.0873)	(0.1162)	(-0.0838)
NB(8,0.5)	22.9342	0.0127	6.0146	0.0088	4.7639	0.0111	20.1399	0.0502	26.4287	0.0138	26.1407	0.0132
	(0.0974)	(-0.0308)	(-0.0717)	(-0.0904)	(-0.1849)	(-0.1547)	(-0.5601)	(-0.4472)	(0.1092)	(-0.0309)	(0.1075)	(-0.0302)
NB(2,0.2)	0.0958	0.0012	0.0804	0.0011	0.0802	0.0011	0.1565	0.0026	0.0963	0.0012	0.0963	0.0012
	(-0.0230)	(-0.0765)	(-0.0350)	(-0.0913)	(-0.0498)	(-0.1083)	(-0.1905)	(-0.2525)	(-0.0228)	(-0.0763)	(-0.0230)	(-0.0764)

4.3 Efficiency of BHHJ-PGF against MLE

Findings from previous sections in this research suggest that in cases without contamination, BHHJ-PGF with greater α values perform better whereas MLE has the best performance among the estimators. It is well-known that MLE is an efficient estimator.

In this section, further analysis is conducted to compare the efficiency of BHHJ-PGF relative to that of MLE by crudely comparing the ratio of MSEs of MLE to BHHJ-PGF method for a deemed large enough samples of size 500. If the ratio is less than 1, it indicates that MLE is more efficient than the other estimator that it was being compared to. If the ratio is greater than 1, then the other estimator can be considered as more efficient. Only selected results are organized in Table 4.12 and shown in Figure 4.9 due to space constraint. Complete results are presented in Appendix D.

With the absence of outlier, BHHJ-PGF is relatively less efficient than MLE. Among different α values, BHHJ-PGF with greater α values are more efficient by having smaller MSEs. At increasing percentage of outliers, the proposed estimators especially that with smaller α values show higher efficiency relative to MLE for both parameters r and p. With 20% contamination to the sample data, the MSEs of BHHJ-PGF(0.001) is at least twice smaller than that of MLE. This implies that besides robustness, BHHJ-PGF with smaller α values ($\alpha = 0, \alpha = 0.001$) appears more efficient than the MLE.

Table 4.12: Relative MSEs of MLE to BHHJ-PGF estimation.

Contamination	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.5)		BHHJ-PGF(1.0)		BHHJ-PGF(2.0)		MLE	
	r	p	r	p	r	p	r	p	r	p	r	p
0%	0.2404	0.1824	0.2404	0.1822	0.2901	0.2288	0.2912	0.2247	0.3719	0.2972	1.0000	1.0000
1%	0.2818	0.2851	0.2825	0.2921	0.3417	0.3636	0.3473	0.3613	0.4238	0.4393	1.0000	1.0000
5%	1.6321	2.2388	1.6330	2.2425	1.9462	2.4866	1.9511	2.3149	1.6823	1.8181	1.0000	1.0000
10%	3.2109	3.2395	3.2085	3.2374	3.4389	3.0518	3.0161	2.5864	2.1520	1.8897	1.0000	1.0000
20%	3.3120	2.3887	3.3141	2.3889	2.9647	2.0814	2.4991	1.8119	1.7603	1.4340	1.0000	1.0000
30%	2.2504	1.6540	2.2504	1.6539	1.8965	1.4583	1.5992	1.3114	1.0834	1.0837	1.0000	1.0000
40%	1.3033	1.2189	1.3028	1.2186	1.1265	1.1118	0.9529	1.0273	0.6587	0.8952	1.0000	1.0000
50%	1.3929	0.8400	1.3926	0.8399	1.2268	0.7889	1.0535	0.7486	0.7781	0.6854	1.0000	1.0000

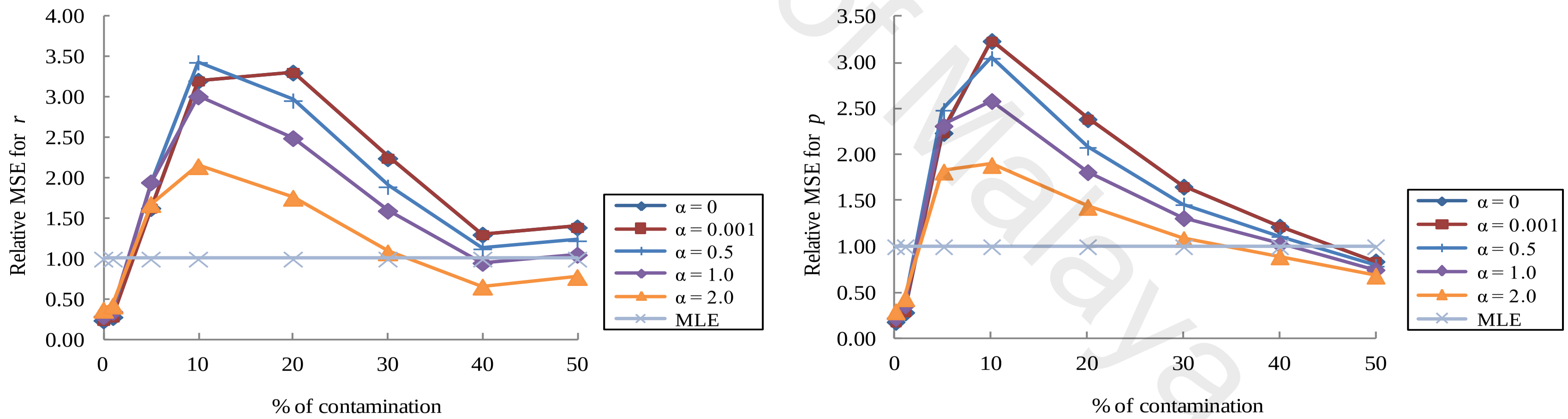

Figure 4.9: Relative MSEs of MLE to BHHJ-PGF estimation.

CHAPTER 5: APPLICATION TO REAL DATA

In this section, the proposed estimator BHHJ-PGF(α) is applied to real life dataset, along with MHD-PGF, JD-PGF and MLE for comparison purposes. The α-values considered here are 0.001, 0.5 and 1.0. Four sets of real life data (a) *Drosophila* data (Simpson, 1987), (b) European red mites on apple leaves (Bliss & Fisher, 1953), (c) the number of ticks counted on each of 82 sheep (Ross & Preece, 1985), and (d) the number of thunderstorm events at Cape Kennedy, Florida in June throughout 11-year period from 1957 to 1967 (Falls et al., 1971) are considered, where the first will be fitted with Poisson distribution and subsequent datasets are fitted with NB distribution. Note that in order to calculate χ^2 goodness of fit, values for classes with expected frequencies less than 5 are grouped and indicated by brackets in the tables.

5.1 Data set 1: *Drosophila*

Drosophila data, which is also considered in Basu et al. (1998), is one of the results taken from an experiment involving the exposure of chemical to male *drosophila* fruit flies, which are then mated with unexposed females. The observed frequency represents the number of male flies with x daughter flies carrying a recessive lethal mutation on the X chromosome. The estimation results are summarized in Table 5.1.

Among all estimations, $\hat{\lambda} = 0.4184$ resulted from BHHJ-PGF(1.0) give the best fit with the smallest χ^2 value. MLE, however, does not fit the data well. This could be due to the data point $x = 91$, which is very likely an extreme outlier. Estimation is performed again by using MLE with the possible outlier removed, leaving the counts of $x = 0{,}1$ and 2 only. The estimated parameter is $\hat{\lambda} = 0.3939$ with the χ^2 value of 0.9094 and 1 degree of freedom.

It is noted that this dataset has a sample size of $n = 34$, and contain only a small percentage of contamination. Therefore, while the other more robust pgf-based methods appear to disregard this point, it can be seen from the expected frequencies in Table 5.1 that MLE, which is sensitive to the presence of contamination, regards the possible outlier as a valid data point to be included as part of the fitted distribution and hence, yields a much larger mean parameter estimate of $\hat{\lambda} = 3.0588$. Table 5.1 also reveals that BHHJ-PGF estimators with greater α values perform slightly better by giving a better fit with smaller χ^2 value. This is consistent with the simulation results in Table 4.1, which shows that BHHJ-PGF with greater α value results in better performance in the absence of contamination in the Poisson data.

Table 5.1: Fit of Po(λ) distribution to *Drosophila* data (Simpson, 1987).

***Drosophila* counts, x**	**Observed frequency**	**Expected frequency**					
		BHHJ-PGF (0.001)	**BHHJ-PGF (0.5)**	**BHHJ-PGF (1.0)**	**MHD-PGF**	**JD-PGF**	**MLE**
0	23	22.48	22.43	22.37	22.48	22.48	1.60
1	7	9.30	9.33	9.36	9.30	9.30	4.88
2 }	3	1.92	1.94	1.96	1.92	1.92	7.47
3 }	0	0.27	0.27	0.27	0.27	0.27	7.61
4 }	0	0.03	0.03	0.03	0.03	0.03	5.82
5 }	0	0.00	0.00	0.00	0.00	0.00	3.56
6 }	0	0.00	0.00	0.00	0.00	0.00	1.82
7 }	0	0.00	0.00	0.00	0.00	0.00	0.79
8 }	0	0.00	0.00	0.00	0.00	0.00	0.30
9 }	0	0.00	0.00	0.00	0.00	0.00	0.10
10 }	0	0.00	0.00	0.00	0.00	0.00	0.03
11 }	0	0.00	0.00	0.00	0.00	0.00	0.01
12 - 90 }	0	0.00	0.00	0.00	0.00	0.00	0.00
91 }	1	0.00	0.00	0.00	0.00	0.00	0.00
N	34	34	34	34	34	34	34
Total χ^2		2.0121	1.9771	1.9465	2.0119	2.0119	308.0596
Estimated parameter	$\hat{\lambda}$	0.4136	0.4161	0.4184	0.4136	0.4136	3.0588

Note: The $\chi^2_{0.05}$ at 1 degree of freedom is 3.841. The curly bracket "}" indicates that the corresponding expected frequencies were grouped for these x values to calculate the χ^2 statistic.

5.2 Data set 2: European red mites

The second data set involving European red mites on apple leaves is also utilized by (Sim & Ong, 2010), where T_2 in their work corresponds to MHD-PGF mentioned here. There is no apparent outlier in this data set. The results for the considered estimators of the NB distribution parameters are shown in Table 5.2. All pgf-based methods resulted in comparable parameter estimates and gave slightly better fit than MLE.

Table 5.2: Fit of NB(r, p) distribution to European red mites on apple leaves data (Bliss & Fisher, 1953).

Number of red mites per leave, x	Observed frequency	Expected frequency					
		BHHJ-PGF (0.001)	BHHJ-PGF (0.5)	BHHJ-PGF (1.0)	MHD-PGF	JD-PGF	MLE
0	70	70.09	70.07	70.07	70.09	70.09	69.50
1	38	36.74	36.79	36.79	36.74	36.74	37.59
2	17	19.73	19.75	19.75	19.73	19.73	20.10
3	10	10.68	10.68	10.67	10.68	10.68	10.70
4	9	5.80	5.79	5.79	5.80	5.80	5.69
5 }	3	3.16	3.15	3.15	3.16	3.16	3.02
6 }	2	1.72	1.72	1.71	1.72	1.72	1.60
7 }	1	0.94	0.94	0.93	0.94	0.94	0.85
8+ }	0	1.14	1.12	1.12	1.14	1.14	0.96
N	150	150	150	150	150	150	150
Total χ^2		2.36	2.36	2.36	2.36	2.36	2.49
Estimated parameter	$\hat{r}$	0.9533	0.9572	0.9575	0.9533	0.9533	1.0237
	$\hat{p}$	0.4502	0.4515	0.4517	0.4502	0.4502	0.4717

Note: The $\chi^2_{0.05}$ at 3 degrees of freedom is 7.82. The curly bracket "}" indicates that the corresponding expected frequencies were grouped for these x values to calculate the χ^2 statistic.

5.3 Data set 3: Ticks on sheep

The data involving number of ticks counted on each of 82 sheep (Ross & Preece, 1985) is fitted with NB distribution using BHHJ-PGF (α=0.001,0.5,1.0), MHD-PGF, JD-PGF and ML estimators. The result is presented in Table 5.3.

It can be seen that all pgf-based estimators result in similar estimated parameters values, but MLE estimates are rather different. Among BHHJ-PGF methods, greater α

gives estimates that fit the data better; the results from BHHJ-PGF (1.0) have the smallest χ^2-value. BHHJ-PGF (0.001) has χ^2-value similar to that of JD-PGF as both are KL-type divergences. On the other hand, MLE does not fit this data set, which spreads over a wide range of tick counts, appearing to have a long right tail. There may also be the possibility of outliers or data contamination for the data set. The results from the analysis of this data set appears to be consistent with the simulation results in the case of long-tailed NB distribution, where the BHHJ-PGF with larger α values give better estimates when there is a small percentage of contamination (Table 4.10).

Table 5.3: Fit of NB(r, p) distribution to the number of ticks on 82 sheep (Ross & Preece, 1985).

Tick counts, x	Observed frequency	Expected frequency					
		BHHJ-PGF (0.001)	BHHJ-PGF (0.5)	BHHJ-PGF (1.0)	MHD-PGF	JD-PGF	MLE
0 }	4	3.86	3.87	3.89	3.86	3.86	5.26
1 }	5	6.82	6.81	6.80	6.81	6.82	7.35
2	11	8.44	8.40	8.35	8.41	8.44	8.03
3	10	8.93	8.88	8.82	8.90	8.93	7.96
4	9	8.66	8.61	8.55	8.63	8.66	7.48
5	11	7.94	7.90	7.85	7.92	7.94	6.80
6	3	6.99	6.96	6.93	6.98	6.99	6.04
7	5	5.98	5.97	5.95	5.98	5.98	5.28
8 }	3	5.01	5.00	5.00	5.01	5.01	4.56
9 }	2	4.11	4.12	4.13	4.12	4.11	3.90
10 }	2	3.33	3.35	3.36	3.34	3.33	3.31
11 }	5	2.67	2.69	2.71	2.68	2.67	2.78
12+	12	9.25	9.44	9.65	9.36	9.25	13.25
N	82	82	82	82	82	82	82
Total χ^2		7.65	7.58	7.52	7.61	7.65	8.77
Estimated parameter	$\hat{r}$	2.5157	2.4878	2.4583	2.5013	2.5157	1.7775
	$\hat{p}$	0.2966	0.2932	0.2894	0.2948	0.2966	0.2132

Note: The $\chi^2_{0.05}$ at 7 degrees of freedom is 14.07. The curly bracket "}" indicates that the corresponding expected frequencies were grouped for these x values to calculate the χ^2 statistic.

5.4 Data set 4: Thunderstorms

There is no apparent outlying data point for this last data set on the number of thunderstorm events at Cape Kennedy, Florida shown in Table 5.4. Hence, MLE is expected to perform as one of the best parameter estimation methods to fit the data with a NB distribution. This is confirmed in Table 5.4 that shows MLE estimates giving the smallest χ^2-value and hence, the best fit with its estimated parameters. Pgf-based methods perform similarly in terms of estimated parameters values.

It is observed that the BHHJ-PGF with increasing α=0.001 to α=1.0 produces estimation result increasingly closer to that of MLE. This echoes the simulation results in Table 4.9; in the case of moderate-tailed NB distribution without contamination, MLE performs the best whereas BHHJ-PGF (1.0) gives the best performance among the pgf-based methods. By tuning the α parameter to as large as 3.0, we find that BHHJ-PGF produces estimated values, $\hat{r}$= 1.0255, $\hat{p}$= 0.5729, with χ^2=1.62 similar to that of MLE. Comparing to MLE, BHHJ-PGF (3.0) gives a better fit with closer expected frequencies for the first 80% of the data. On the other hand, MLE method produces estimates that better represent the tail. Tuning α to even larger values produce estimates similar to BHHJ-PGF (3.0) but with larger χ^2-values.

Table 5.4: Fit of NB(r, p) distribution to the number of thunderstorm events at Cape Kennedy, Florida, in June from 1957 to 1967 (Falls et al., 1971).

Thunderstorm counts, x	Observed frequency	Expected frequency						
		BHHJ-PGF (0.001)	BHHJ-PGF (0.5)	BHHJ-PGF (1.0)	BHHJ-PGF (3.0)	MHD-PGF	JD-PGF	MLE
0	187	186.65	186.60	186.55	186.38	186.65	186.65	184.64
1	77	80.71	80.91	81.08	81.64	80.71	80.71	84.56
2	40	35.21	35.23	35.25	35.31	35.21	35.21	35.88
3	17	15.40	15.36	15.33	15.21	15.40	15.40	14.82
4 }	6	6.75	6.70	6.66	6.54	6.75	6.75	6.04
5 }	2	2.96	2.93	2.90	2.81	2.96	2.96	2.44
6 }	1	2.31	2.27	2.23	2.11	2.31	2.31	1.63
N	330	330	330	330	330	330	330	330
Total χ^2		1.75	1.72	1.69	1.62	1.75	1.75	1.62
Estimated parameter	$\hat{r}$	0.9828	0.9918	0.9992	1.0255	0.9828	0.9828	1.1724
	$\hat{p}$	0.5600	0.5628	0.5651	0.5729	0.5600	0.5600	0.6094

Note: The $\chi^2_{0.05}$ at 2 degrees of freedom is 5.99.

CHAPTER 6: CONCLUSION

This research intends to find a new robust approach to parameter estimation by modifying an existing density based divergence through the introduction of the probability generating function to replace the probability density function.

Theoretical proofs are given to show that the proposed divergence, BHHJ-PGF serves as a consistent estimator which is also asymptotically normal. BHHJ proposed by (Basu et al., 1998) provides a bridge between robustness and efficiency through its tuning parameter, α. With that in mind, simulation study is conducted to look into the effect of the values of α and the performance of BHHJ-PGF in various percentages of contamination.

It can be concluded that in cases where there are small percentages of outliers present, among BHHJ-PGF estimators, those with larger α values performs better. As the contamination increases, BHHJ-PGF with smaller α values give smaller MSEs and relative biases. Hence, BHHJ-PGF with smaller α provides more robust estimates for larger contamination. In small sample sizes, BHHJ-PGF with large α is more robust. Here, the tuning parameter α serves to mitigate the impact of outliers.

The real data applications in Chapter 5 mainly involve biological and environmental data. The negative binomial distribution, which is the choice distribution for simulation study, can represent these data well. These applications clearly demonstrated the usefulness of the newly proposed method to provide comparable, robust estimates.

The contributions of this research could be extended to include other discrete distributions as well as to multivariate distributions for future work. Other transforms such as moment generating function and characteristic function, in place of the probability generating function, may also be considered for further investigations.

REFERENCES

Anscombe, F. J., & Guttman, I. (1960). Rejection of Outliers. *Technometrics, 2*(2), 123-147.

Arbous, A. G., & Kerrich, J. E. (1951). Accident statistics and the concept of accident-proneness. *Biometrics, 7*(4), 340-432.

Banik, S., & Kibria, B. M. G. (2009). On some discrete distributions and their applications with real life data. *Journal of Modern Applied Statistical Methods, 8*(2), 423-447.

Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika, 85*(3), 549-559.

Basu, A., Shioya, H., & Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*. New York: CRC Press.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics, 5*(3), 445-463.

Bliss, C. I., & Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics, 9*(2), 176-200.

Cressie, N., & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological), 46*(3), 440-464.

Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramer-von Mises Tests. *The Annals of Mathematical Statistics, 28*(4), 823-838.

Davies, P. L., & Gather, U. (2005). Breakdown and groups. *The Annals of Statistics, 33*(3), 977-988.

Dowling, M. M., & Nakamura, M. (1997). Estimating parameters for discrete distributions via the empirical probability generating function. *Communications in Statistics - Simulation and Computation, 26*(1), 301-313.

Falls, L. W., Williford, W. O., & Carter, M. C. (1971). Probability distributions for thunderstorm activity at Cape Kennedy, Florida. *Journal of Applied Meteorology, 10*(1), 97-104.

Feuerverger, A., & McDunnough, P. (1984). On statistical transform methods and their efficiency. *The Canadian Journal of Statistics, 12*(4), 303-317.

Field, C., & Smith, B. (1994). Robust estimation: A weighted maximum likelihood approach. *International Statistical Review, 62*(3), 405-424.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A, 222*(594-604), 309-368.

Ghosh, A., & Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electronic Journal of Statistics*, *7*(2013), 2420-2456.

Ghosh, A., & Basu, A. (2016). Robust estimation in generalized linear models: the density power divergence approach. *TEST, 25*(2), 269-290.

Gurland, J. (1959). Some applications of the negative binomial and other contagious distributions. *American Journal of Public Health*, *49*(10), 1388-1399.

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics, 42*(6), 1887-1896.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*(346), 383-393.

Hong, C., & Kim, Y. (2001). Automatic selection of the tuning parameter in the minimum density power divergence estimation. *Journal of the Korean Statistical Association*, *30*, 453-465.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics, 35*(1), 73-101.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 186*(1007), 453-461.

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate Discrete Distributions, Third Edition*. New Jersey: John Wiley & Sons, Inc.

Kang, J., & Lee, S. (2014). Minimum density power divergence estimator for Poisson autoregressive models. *Computational Statistics & Data Analysis*, *80*(2014), 44-56.

Karlis, D., & Xekalaki, E. (1998). Minimum Hellinger distance estimation for Poisson mixtures. *Computational Statistics and Data Analysis*(29), 81-103.

Kemp, C. D., & Kemp, A. W. (1988). Rapid estimation for discrete distributions. *The Statistician, 37*, 243-255.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79-86.

Ng, C. M., Ong, S. H., & Srivastava, H. M. (2013). Parameter estimation by Hellinger type distance for multivariate distributions based upon probability generating functions. *Applied Mathematical Modelling*, *37*(12–13), 7374-7385.

Pak, R. J. (2014). The minimum density power divergence estimation for the lognormal density. *Communications in Statistics - Theory and Methods*, *43*(21), 4582-4588.

Patra, S., Maji, A., & Basu, A. (2013). The power divergence and the density power divergence families: the mathematical connection. *Sankhya: The Indian Journal of Statistics, 75-B*(Part 1), 16-28.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society A, 185*, 71-110.

Pearson, K. (1902). On the sytematic fitting of curves to observations and measurements. *Biometrika, 1*(3), 265-303.

Rade, L. (1972). On the use of generating functions and laplace transforms in applied probability theory. *International Journal of Mathematical Education in Science and Technology, 3*(1), 25-33.

Ross, G. J. S., & Preece, D. A. (1985). The negative binomial distribution. *Journal of the Royal Statistical Society. Series D (The Statistician), 34*(3), 323-335.

Rueda, R., & O' Reilly, F. (1999). Tests of fit for discrete distributions based on the probability generating function. *Communications in Statistics - Simulation and Computation, 28*(1), 259-274.

Rueda, R., Pérez-Abreu, V., & O'Reilly, F. (1991). Goodness of fit for the Poisson distribution based on the probability generating function. *Communications in Statistics - Theory and Methods, 20*(10), 3093-3110.

Sharifdoust, M., Ng, C. M., & Ong, S. H. (2016). Probability generating function based Jeffrey's divergence for statistical inference. *Communications in Statistics - Simulation and Computation, 45*(7), 2445-2458.

Sim, S. Z., & Ong, S. H. (2010). Parameter estimation for discrete distributions by generalized Hellinger-type divergence based on probability generating function. *Communications in Statistics - Simulation and Computation, 39*(2), 305-314.

Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association, 82*(399), 802-807.

Stefanski, L. A., & Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician, 56*(1), 29-38.

Warwick, J. (2005). A data-based method for selecting tuning parameters in minimum distance estimators. *Computational Statistics & Data Analysis, 48*(3), 571-585.

Warwick, J., & Jones, M. C. (2005). Choosing a robustness tuning parameter. *Journal of Statistical Computation and Simulation, 75*(7), 581-588.

Zhao, C., Hong, H.-s., Bao, W.-m., & Zhang, L.-p. (2008). Robust recursive estimation of auto-regressive updating model parameters for real-time flood forecasting. *Journal of Hydrology, 349*(3-4), 376-382.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

1. **Tay, S.Y.**, Ng, C.M. & Ong, S.H. (2016). Parameter estimation using probability generating function based minimum power divergence. AIP Conference Proceedings, Vol. 1750, Issue 1.

2. **Tay, S.Y.**, Ng, C.M. & Ong, S.H. (2018). Parameter estimation by minimizing a probability generating function-based power divergence. Communication in Statistics – Simulation and Computation. (Published online. doi: https://doi.org/10.1080/03610918.2018.1468462

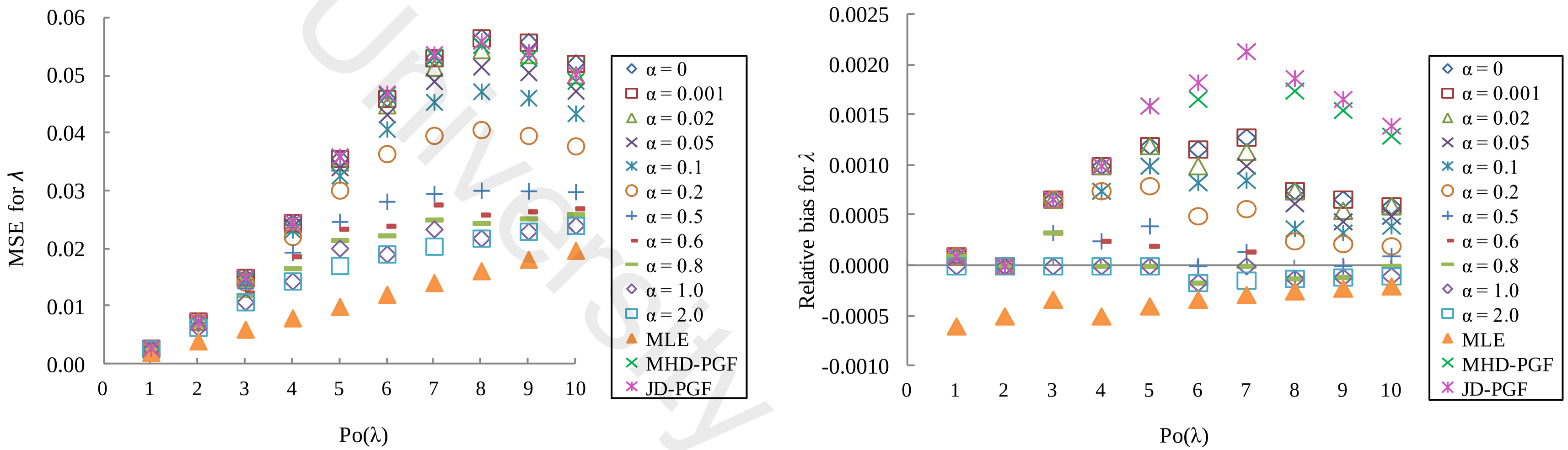

Figure B1: (Complete of Figure 4.1) MSE and relative biases for estimators with samples of size $n = 500$, from $Po(\lambda)$, without contamination.

Table B2: MSE and relative biases (in bracket) for estimators with samples of size $n = 500$, from $Po(\lambda)$, with 5% contamination.

Po(λ)	BHHJ-PGF(α)											MLE	MHD-PGF	JD-PGF
	α = 0	α = 0.001	α = 0.02	α = 0.05	α = 0.1	α = 0.2	α = 0.5	α = 0.6	α = 0.8	α = 1.0	α = 2.0			
Po(1)	0.0068	0.0068	0.0069	0.0069	0.0069	0.0069	0.0070	0.0070	0.0071	0.0072	0.0074	0.0246	0.0068	0.0068
	(0.0626)	(0.0626)	(0.0626)	(0.0627)	(0.0629)	(0.0632)	(0.0640)	(0.0645)	(0.0651)	(0.0656)	(0.0678)	(0.1490)	(0.0626)	(0.0626)
Po(2)	0.0140	0.0140	0.0140	0.0141	0.0141	0.0142	0.0144	0.0145	0.0147	0.0149	0.0158	0.0940	0.0140	0.0140
	(0.0390)	(0.0390)	(0.0390)	(0.0390)	(0.0395)	(0.0400)	(0.0415)	(0.0420)	(0.0430)	(0.0440)	(0.0475)	(0.1495)	(0.0390)	(0.0390)
Po(3)	0.0246	0.0246	0.0246	0.0246	0.0246	0.0246	0.0247	0.0249	0.0252	0.0255	0.0277	0.2087	0.0247	0.0247
	(0.0307)	(0.0307)	(0.0310)	(0.0310)	(0.0313)	(0.0320)	(0.0340)	(0.0347)	(0.0360)	(0.0373)	(0.0420)	(0.1497)	(0.0307)	(0.0307)
Po(4)	0.0381	0.0381	0.0380	0.0378	0.0375	0.0371	0.0368	0.0369	0.0374	0.0382	0.0428	0.3678	0.0382	0.0382
	(0.0270)	(0.0270)	(0.0270)	(0.0273)	(0.0277)	(0.0285)	(0.0313)	(0.0320)	(0.0337)	(0.0353)	(0.0410)	(0.1498)	(0.0270)	(0.0270)
Po(5)	0.0550	0.0550	0.0546	0.0540	0.0531	0.0518	0.0506	0.0508	0.0519	0.0536	0.0625	0.5727	0.0555	0.0556
	(0.0254)	(0.0254)	(0.0256)	(0.0258)	(0.0264)	(0.0274)	(0.0306)	(0.0316)	(0.0336)	(0.0354)	(0.0416)	(0.1498)	(0.0258)	(0.0258)
Po(6)	0.0721	0.0721	0.0712	0.0700	0.0683	0.0660	0.0647	0.0655	0.0681	0.0713	0.0874	0.8207	0.0732	0.0735
	(0.0245)	(0.0245)	(0.0247)	(0.0250)	(0.0257)	(0.0268)	(0.0305)	(0.0317)	(0.0340)	(0.0360)	(0.0428)	(0.1497)	(0.0252)	(0.0253)
Po(7)	0.0878	0.0878	0.0865	0.0848	0.0825	0.0796	0.0803	0.0823	0.0875	0.0933	0.1197	1.1153	0.0891	0.0896
	(0.0244)	(0.0244)	(0.0247)	(0.0251)	(0.0257)	(0.0270)	(0.0311)	(0.0324)	(0.0349)	(0.0370)	(0.0444)	(0.1497)	(0.0254)	(0.0254)
Po(8)	0.0992	0.0992	0.0978	0.0959	0.0936	0.0915	0.0970	0.1009	0.1096	0.1189	0.1589	1.4538	0.0996	0.1005
	(0.0241)	(0.0241)	(0.0244)	(0.0249)	(0.0255)	(0.0270)	(0.0316)	(0.0330)	(0.0355)	(0.0379)	(0.0458)	(0.1498)	(0.0251)	(0.0251)
Po(9)	0.1089	0.1088	0.1076	0.1061	0.1046	0.1045	0.1172	0.1234	0.1368	0.1505	0.2075	1.8378	0.1081	0.1094
	(0.0242)	(0.0242)	(0.0246)	(0.0250)	(0.0258)	(0.0274)	(0.0322)	(0.0337)	(0.0363)	(0.0389)	(0.0473)	(0.1498)	(0.0252)	(0.0252)
Po(10)	0.1163	0.1163	0.1157	0.1152	0.1154	0.1186	0.1406	0.1497	0.1689	0.1884	0.2654	2.2659	0.1154	0.1168
	(0.0244)	(0.0244)	(0.0247)	(0.0252)	(0.0261)	(0.0278)	(0.0328)	(0.0344)	(0.0373)	(0.0399)	(0.0488)	(0.1500)	(0.0251)	(0.0251)

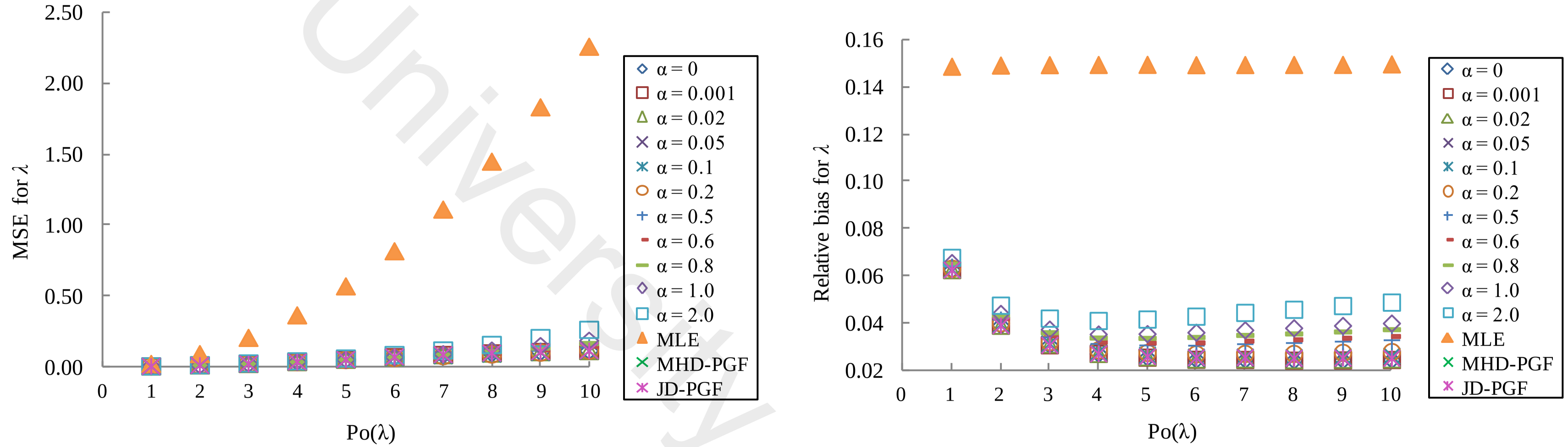

Figure B2: (Complete of Figure 4.2) MSE and relative biases for estimators with samples of size $n = 500$, from $Po(\lambda)$, with 5% contamination.

APPENDIX C

Table C: (Complete of Table 4.3) MSE and relative biases (in bracket) for estimators with samples of size $n = 500$, from $NB(r = 2.0, p = 0.2)$ for varying α and percentages of contamination.

Contamination	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.02)		BHHJ-PGF(0.05)		BHHJ-PGF(0.1)		BHHJ-PGF(0.2)		BHHJ-PGF(0.5)	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
0%	0.1141	0.0011	0.1141	0.0011	0.1132	0.0011	0.1117	0.0011	0.1097	0.0011	0.1066	0.0010	0.0945	0.0009
	(0.0270)	(0.0220)	(0.0270)	(0.0217)	(0.027)	(0.0218)	(0.0265)	(0.0215)	(0.026)	(0.0208)	(0.0248)	(0.0197)	(0.0228)	(0.0176)
1%	0.1085	0.0011	0.1082	0.0011	0.1075	0.0011	0.1063	0.0011	0.1044	0.0010	0.1020	0.0010	0.0895	0.0009
	(0.0165)	(0.0015)	(0.0165)	(0.0016)	(0.0165)	(0.0015)	(0.0160)	(0.0011)	(0.0155)	(0.0000)	(0.0137)	(-0.0016)	(0.0105)	(-0.0050)
5%	0.0959	0.0012	0.0958	0.0012	0.0951	0.0012	0.0940	0.0012	0.0923	0.0011	0.0923	0.0011	0.0804	0.0011
	(-0.0230)	(-0.0765)	(-0.0230)	(-0.0765)	(-0.0235)	(-0.0772)	(-0.024)	(-0.0779)	(-0.0255)	(-0.0796)	(-0.0282)	(-0.0828)	(-0.0350)	(-0.0913)
10%	0.0997	0.0019	0.0997	0.0019	0.0994	0.0019	0.0986	0.0019	0.0981	0.0019	0.0960	0.0019	0.0931	0.0020
	(-0.0670)	(-0.1669)	(-0.0675)	(-0.1669)	(-0.0680)	(-0.1678)	(-0.0690)	(-0.1691)	(-0.0710)	(-0.1719)	(-0.0747)	(-0.1760)	(-0.0857)	(-0.1895)
20%	0.1460	0.0048	0.1459	0.0048	0.1462	0.0049	0.1467	0.0049	0.1480	0.0050	0.1504	0.0051	0.1631	0.0055
	(-0.1415)	(-0.3259)	(-0.1415)	(-0.3260)	(-0.1425)	(-0.3272)	(-0.1440)	(-0.3291)	(-0.1469)	(-0.3325)	(-0.1520)	(-0.3388)	(-0.1681)	(-0.3585)
30%	0.2143	0.0088	0.2143	0.0088	0.2154	0.0089	0.2154	0.0089	0.2264	0.0091	0.2264	0.0093	0.2543	0.0100
	(-0.2000)	(-0.4586)	(-0.2000)	(-0.4586)	(-0.2015)	(-0.4599)	(-0.2015)	(-0.4599)	(-0.2080)	(-0.4670)	(-0.2119)	(-0.4724)	(-0.2303)	(-0.4930)
40%	0.2854	0.0132	0.2855	0.0132	0.2871	0.0132	0.2895	0.0133	0.3305	0.0138	0.3020	0.0137	0.3302	0.0144
	(-0.2446)	(-0.5673)	(-0.2447)	(-0.5674)	(-0.2458)	(-0.5687)	(-0.2476)	(-0.5706)	(-0.2590)	(-0.5797)	(-0.2562)	(-0.5799)	(-0.2726)	(-0.5969)
50%	0.3462	0.0174	0.3463	0.0174	0.3478	0.0174	0.3501	0.0175	0.3879	0.0179	0.3626	0.0178	0.3931	0.0185
	(-0.2764)	(-0.6553)	(-0.2764)	(-0.6553)	(-0.2773)	(-0.6563)	(-0.2789)	(-0.6578)	(-0.2883)	(-0.6642)	(-0.2863)	(-0.6649)	(-0.3007)	(-0.6778)

Table C continued

Contamination	BHHJ-PGF(0.6)		BHHJ-PGF(0.8)		BHHJ-PGF(1.0)		BHHJ-PGF(1.1)		BHHJ-PGF(2.0)		MLE		MHD-PGF		JD-PGF	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p	r	p
0%	0.0950	0.0009	0.0889	0.0009	0.0942	0.0009	0.0904	0.0009	0.0738	0.0007	0.0274	0.0002	0.1144	0.0011	0.1142	0.0011
	(0.0206)	(0.0154)	(0.0191)	(0.0139)	(0.0152)	(0.0098)	(0.0149)	(0.0095)	(0.0092)	(0.0043)	(0.0080)	(0.0055)	(0.0275)	(0.0222)	(0.0275)	(0.0222)
1%	0.0870	0.0008	0.0820	0.0008	0.0880	0.0009	0.0890	0.0009	0.0722	0.0007	0.030578	0.000314	0.1092	0.0011	0.1087	0.0011
	(0.0092)	(-0.0063)	(0.0068)	(-0.0090)	(0.0015)	(-0.0148)	(-0.0008)	(-0.0170)	(-0.0102)	(-0.0271)	(-0.0480)	(-0.0625)	(0.0170)	(0.0020)	(0.0175)	(0.0025)
5%	0.0818	0.0011	0.0754	0.0010	0.0802	0.0011	0.0821	0.0012	0.0930	0.0014	0.1565	0.0026	0.0963	0.0012	0.0963	0.0012
	(-0.0385)	(-0.0952)	(-0.0429)	(-0.1003)	(-0.0498)	(-0.1083)	(-0.0531)	(-0.1120)	(-0.0774)	(-0.1399)	(-0.1905)	(-0.2525)	(-0.0228)	(-0.0763)	(-0.0230)	(-0.0764)
10%	0.0957	0.0021	0.0933	0.0021	0.1061	0.0024	0.1083	0.0025	0.1487	0.0033	0.3200	0.0062	0.1015	0.0019	0.1000	0.0019
	(-0.0900)	(-0.1947)	(-0.0966)	(-0.2030)	(-0.1068)	(-0.2145)	(-0.1106)	(-0.2190)	(-0.1461)	(-0.2600)	(-0.2795)	(-0.3920)	(-0.0675)	(-0.1671)	(-0.0670)	(-0.1668)
20%	0.1667	0.0057	0.1727	0.0059	0.1934	0.0064	0.2017	0.0065	0.2746	0.0080	0.4834	0.0115	0.1461	0.0048	0.1460	0.0048
	(-0.1728)	(-0.3639)	(-0.1819)	(-0.3750)	(-0.1939)	(-0.3885)	(-0.1995)	(-0.3945)	(-0.2397)	(-0.4395)	(-0.3455)	(-0.5366)	(-0.1413)	(-0.3257)	(-0.1415)	(-0.3258)
30%	0.2584	0.0102	0.2729	0.0106	0.3016	0.0111	0.3117	0.0113	0.4452	0.0135	0.4823	0.0146	0.2146	0.0088	0.2143	0.0088
	(-0.2350)	(-0.4985)	(-0.2450)	(-0.5098)	(-0.2573)	(-0.5222)	(-0.2623)	(-0.5275)	(-0.3100)	(-0.5730)	(-0.3454)	(-0.6037)	(-0.2000)	(-0.4310)	(-0.2000)	(-0.4585)
40%	0.3384	0.0146	0.3671	0.0152	0.3904	0.0156	0.4088	0.0159	0.5647	0.0179	0.3720	0.0160	0.2881	0.0132	0.2853	0.0132
	(-0.2774)	(-0.6018)	(-0.2891)	(-0.6127)	(-0.2988)	(-0.6000)	(-0.3048)	(-0.6269)	(-0.3508)	(-0.6638)	(-0.3025)	(-0.6330)	(-0.2451)	(-0.5677)	(-0.2444)	(-0.5671)
50%	0.4141	0.0188	0.4377	0.0192	0.4578	0.0195	0.4543	0.0195	0.6198	0.0213	0.1919	0.0162	0.3470	0.0174	0.3459	0.0174
	(-0.3074)	(-0.6829)	(-0.3162)	(-0.6899)	(-0.3236)	(-0.6958)	(-0.3241)	(-0.6968)	(-0.3681)	(-0.7256)	(-0.2140)	(-0.6359)	(-0.2763)	(-0.6553)	(-0.2760)	(-0.6550)

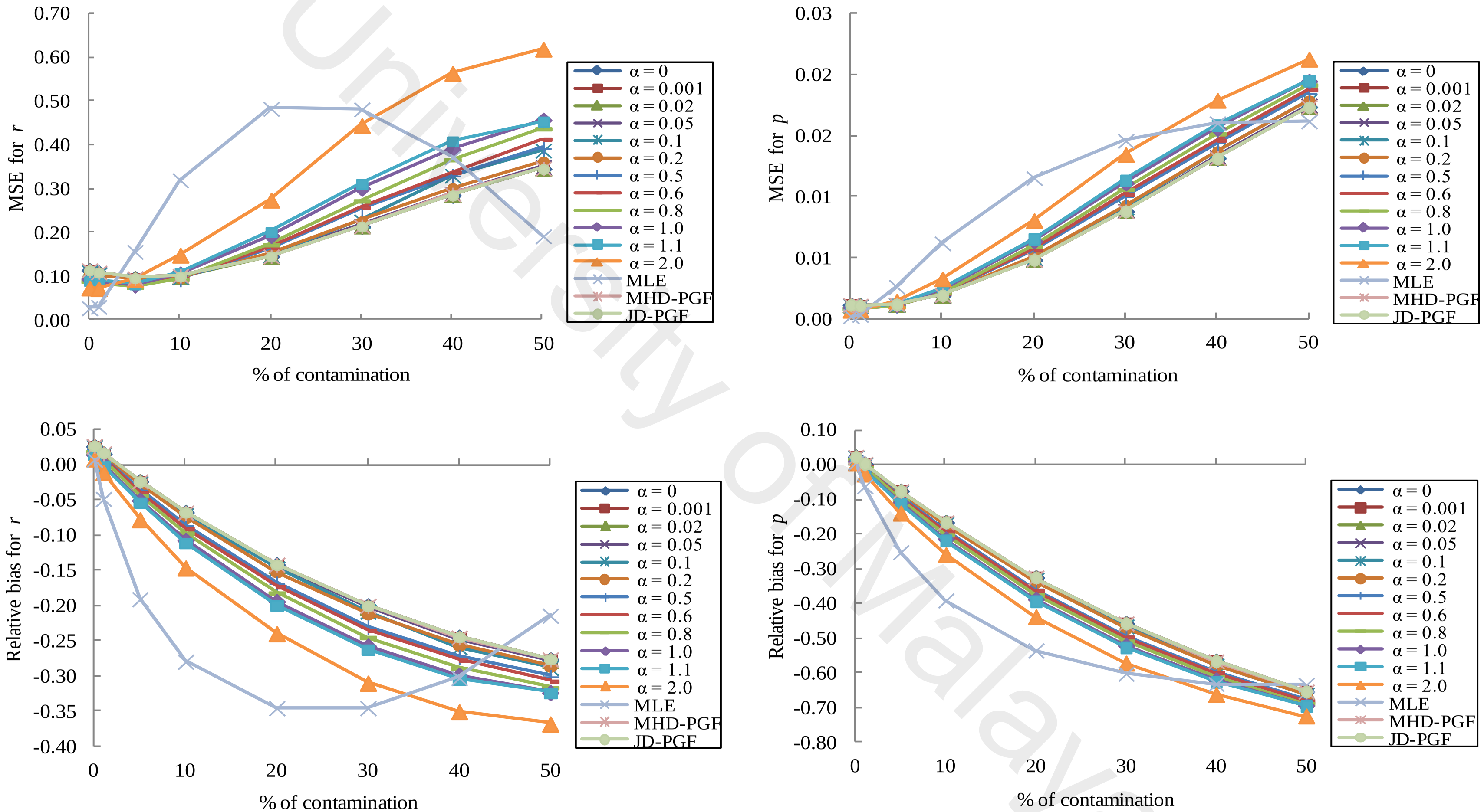

Figure C: (Complete of Figure 4.3) MSE and relative biases for estimators with samples of size $n = 500$, from $NB(r = 2.0, p = 0.2)$ for varying α and percentages of contamination.

APPENDIX D

Table D: (Complete of Table 4.12) Relative MSEs of MLE to BHHJ-PGF estimation.

Contamination	BHHJ-PGF(0)		BHHJ-PGF(0.001)		BHHJ-PGF(0.02)		BHHJ-PGF(0.05)		BHHJ-PGF(0.1)		BHHJ-PGF(0.2)		BHHJ-PGF(0.5)		BHHJ-PGF(0.6)	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p	r	p
0%	0.2404	0.1824	0.2404	0.1822	0.2423	0.1844	0.2456	0.1873	0.2501	0.1911	0.2573	0.1978	0.2901	0.2288	0.2887	0.2257
1%	0.2818	0.2851	0.2825	0.2921	0.2845	0.2948	0.2877	0.2985	0.2928	0.3036	0.2998	0.3119	0.3417	0.3636	0.3514	0.3769
5%	1.6321	2.2388	1.6330	2.2425	1.6447	2.2496	1.6642	2.2701	1.6960	2.2957	1.6960	2.2957	1.9462	2.4866	1.9139	2.4039
10%	3.2109	3.2395	3.2085	3.2374	3.2208	3.2329	3.2470	3.2309	3.2608	3.1958	3.3322	3.1827	3.4389	3.0518	3.3423	2.9429
20%	3.3120	2.3887	3.3141	2.3889	3.3065	2.3776	3.2958	2.3600	3.2657	2.3265	3.2143	2.2681	2.9647	2.0814	2.8995	2.0321
30%	2.2504	1.6540	2.2504	1.6539	2.2395	1.6462	2.2395	1.6462	2.1304	1.5996	2.1306	1.5750	1.8965	1.4583	1.8667	1.4330
40%	1.3033	1.2189	1.3028	1.2186	1.2959	1.2138	1.2850	1.2066	1.1256	1.1634	1.2319	1.1725	1.1265	1.1118	1.0992	1.0952
50%	1.3929	0.8400	1.3926	0.8399	1.3868	0.8378	1.3775	0.8343	1.2433	0.8162	1.3300	0.8179	1.2268	0.7889	1.1646	0.7766

Table D continued

Contamination	BHHJ-PGF(0.8)		BHHJ-PGF(1.0)		BHHJ-PGF(1.1)		BHHJ-PGF(2.0)		MLE	
	r	p	r	p	r	p	r	p	r	p
0%	0.3085	0.2428	0.2912	0.2247	0.3034	0.2356	0.3719	0.2972	1.0000	1.0000
1%	0.3729	0.4002	0.3473	0.3613	0.3436	0.3528	0.4238	0.4393	1.0000	1.0000
5%	2.0752	2.5185	1.9511	2.3149	1.9050	2.2339	1.6823	1.8181	1.0000	1.0000
10%	3.4292	2.8781	3.0161	2.5864	2.9555	2.5205	2.1520	1.8897	1.0000	1.0000
20%	2.7990	1.9436	2.4991	1.8119	2.3964	1.7616	1.7603	1.4340	1.0000	1.0000
30%	1.7672	1.3761	1.5992	1.3114	1.5471	1.2857	1.0834	1.0837	1.0000	1.0000
40%	1.0133	1.0570	0.9529	1.0273	0.9099	1.0098	0.6587	0.8952	1.0000	1.0000
50%	1.1018	0.7612	1.0535	0.7486	1.0615	0.7471	0.7781	0.6854	1.0000	1.0000

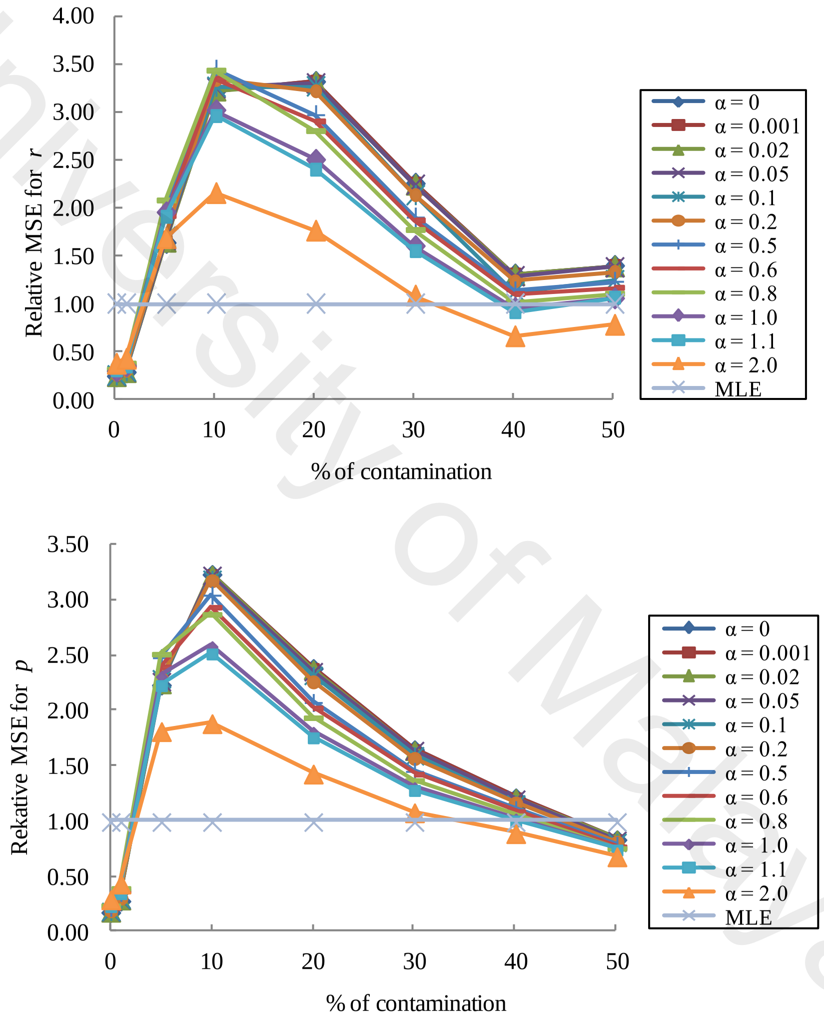

Figure D: (Complete of Figure 4.9) Relative MSEs of MLE to BHHJ-PGF estimation.