T.C.

AKSARAY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

DEPARTMENT OF ELECTRICAL ELECTRONIC AND COMPUTER ENGINEERING

STATISTICAL ANALYSIS OF REFRACTIVE INDEX STRUCTURE PARAMETER

MASTER OF SCIENCE THESIS

İbrahim YAZAR

SUPERVISOR

Asst. Prof. Dr. Yusuf Erkan YENİCE

AKSARAY, 2019

İbrahim YAZAR, M.Sc. with thesis student No. 162335001 of Graduate School of Natural and Applied Sciences, successfully presented the M. Sc. Thesis titled **"STATISTICAL ANALYSIS OF REFRACTIVE INDEX STRUCTURE PARAMETER"** prepared after satisfying all the requirements identified in the concerned bylaws, before the jury whose signatures are affixed below.

Thesis Advisor: Asst. Prof. Dr. Yusuf Erkan YENİCE
Aksaray University
I approve that this thesis is a Master Thesis in scope and quality ....

Jury Member: Asst. Prof. Dr. Filiz SARI
Aksaray University
I approve that this thesis is a Master Thesis in scope and quality ....

Jury Member: Asst. Prof. Dr. Serdar ÇİÇEK
Nevşehir Hacı Bektaş Veli University
I approve that this thesis is a Master Thesis in scope and quality ....

Date of Defence: 04/07/2019

I agree that this thesis, which is accepted by the jury, has fulfilled the requirements for being a Master Thesis.

....

Assoc. Prof. Dr. M. Ali HINIS

DECLARATION

I declare that I write this work as a master thesis without any help and a way that would be contrary to academic rules and scientific ethics, morals, and traditions.

I declare that I will endure all moral and legal consequences that will arise if any situation is found to be contrary to my statement, regardless of a certain time period by the Institute.

İbrahim YAZAR

ACKNOWLEDGEMENT

First of all, I would like to express my sincere gratitude to my advisor Asst. Prof. Dr. Yusuf Erkan Yenice. Also, I would like to thank Asst. Prof. Dr. Filiz Sarı, who was my previous advisor, all my instructors and colleagues. Additionally, I would like to thank my family and my friends who were there for me throughout my life.

Finally, I would like to thank Dr. Nurdan Bacı and her team at TUBITAK, who share their research data with me and made it possible to finish this thesis.

İbrahim YAZAR
AKSARAY, 2019

TABLE OF CONTENTS

YÜKSEK LİSANS TEZİ

KIRILMA İNDİSİ YAPI PARAMETRESİNİN İSTATİSTİKSEL ANALİZİ

İbrahim YAZAR

Aksaray Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Yusuf Erkan YENİCE

ÖZET

Kablosuz optik haberleşme sistemlerinin tasarımında karşılaşılan en büyük sorunlardan birisi olan türbülansı sadece kırılma indisi yapı parametresiyle modellemek, diğer parametreleri uygun bir şekilde tahmin ederek mümkündür. Bu yüzden kırılma indisi yapı parametresinin olasılık yoğunluk fonksiyonunun bilinmesi çok önemlidir. Bu konuda farklı modeller ortaya atılmış olsa da hiçbiri teorik olarak ispatlanamamıştır.

Bu çalışmanın ana konusu kırılma indisi yapı parametresinin olasılık dağılımının tek taraftan mı yoksa iki taraftan mı sınırlı olduğunun incelenmesidir. İlk etapta Pearson ve Johnson dağılım sistemleri ve uç değer kuramı kullanılmıştır ve dağılımın iki taraftan sınırlı dağılımlarla daha iyi modellenebileceği sonucuna varılmıştır. Ayrıca kırılma indisi yapı parametresinin modeli olarak tek taraftan sınırlı dağılımlar içerisinde en çok kullanılan lognormal dağılımı ile iki taraftan sınırlı dağılımlar içerisinde en çok kullanılan beta dağılımı karşılaştırılmış ve beta dağılımının daha iyi sonuçlar verdiği görülmüştür. Son olarak, beta dağılımı parametrelerini tahmin etme yöntemlerinden olan momentler metodunun yeterli olmadığı, momentler yöntemiyle yapılan tahminlerden sonra en küçük kareler yöntemiyle yapılan iyileştirmenin gerekli olduğu gösterilmiştir.

Anahtar Kelimeler: Kablosuz optik haberleşme, Optik türbülans, Kırılma indisi yapı parametresi, Pearson dağılım sistemi, Johnson dağılım sistemi, Uç değer kuramı, Ortalama aşım fonksiyonu, QQ çizimi, Momentler yöntemi, En küçük kareler yöntemi, Ki-kare testi, K-S testi.

Temmuz, 2019; 69 sayfa

M.Sc. THESIS

STATISTICAL ANALYSIS OF REFRACTIVE INDEX STRUCTURE PARAMETER

İbrahim YAZAR

Aksaray University
Graduate School of Natural and Applied Sciences
Department of Electrical Electronic and Computer Engineering

Supervisor: Asst. Prof. Dr. Yusuf Erkan YENİCE

ABSTRACT

It is possible to model turbulence, which is one of the biggest problems encountered when designing an optical wireless communication system, with only the refractive index structure parameter and good estimation of the rest of the parameters. Which is why knowing the probability distribution function of the refractive index structure parameter is very important. Although different models have been proposed in this regard, none have been proved theoretically.

The main topic of this study is to examine whether the probability distribution of the refractive index structure parameter is bounded or semi-infinite. At the start, Pearson and Johnson distribution systems and Extreme Value Theory were applied and it was concluded that the refractive index structure parameter could be better modelled with bounded distributions. After that, beta distribution and lognormal distribution, which belong to bounded and semi-infinite distributions families respectively and the more known and used distributions as a model for refractive index structure parameter, are compared with each other and it was found out that beta distribution gives better results. Finally, it had been shown that the method of moments, which is one of the methods for estimating the beta distribution parameters, is not enough and after the estimations are made by the method of moments, it is necessary to improve them with the least squares method.

Keywords: Wireless optical communication, Optical turbulence, Refractive index structure parameter, Pearson distribution system, Johnson distribution system, Extreme Value Theory, Mean excess function, QQ plot, Method of moments, Least squares method, Chi-Square test, K-S test.

July, 2019; 69 pages

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

Cdf	Cumulative Distribution Function
C_n^2	Refractive Index Structure Parameter
C_T^2	Temperature Structure Parameter
C_V^2	Velocity Structure Parameter
DARPA	The Defense Advanced Research Projects Agency
D_n	Structure Function
DWLS	Diagonally Weighted Least Squares
GEV	Generalized Extreme Value Distribution
GPD	Generalized Pareto Distribution
K	Kurtosis
k	Shape Parameter of Generalized Pareto Distribution
K-S test	Kolmogorov-Smirnov Test
l	Length
L_0	Outer Scale
l_0	Inner Scale
MoM	Method of Moments
n	Refractive Index's Deviation from the Mean
OLS	Ordinary Least Squares
PWM	Probability Weighted Moments
QQ Plot	Quantile-Quantile Plot
R_e	Reynolds Number
S	Skewness
S-K plane	Skewness-Kurtosis Plane
V	Velocity
v	Kinematic Viscosity
x_i	i^{th} Sample of the Data Set
x_{max}	Maximum Value of the Data Set
x_{min}	Minimum Value of the Data Set
$\overline{x}$	Mean of the Data Set
ε	Dissipation Rate
η	Order of the higher order L-moment
μ^l	Mean of the Logarithmic Values
ξ	Location Parameter of Generalized Pareto Distribution
σ	Scale Parameter of Generalized Pareto Distribution
σ^2	Variance of the Data Set
σ^l	Standard Deviation of the Logarithmic Values

1. INTRODUCTION

Since the atmosphere is made of gases, it is considered as a viscous fluid and thus classical turbulence (mechanical turbulence) theories are also applicable to it. Any viscous fluid considered to have two states of motion which are laminar and turbulent. If a fluid has a uniform velocity flow characteristics or its velocity flow characteristic change in regular fashion, it is considered to have a laminar motion whereas a fluid which lost its uniform characteristic and consists of random subflows has a turbulent motion. Laminar motion becomes turbulent after it exceeds critical Reynolds number value, which is defined as;

$$R_e = \frac{Vl}{v} \tag{1.1}$$

Where V is velocity (m/s), l is length (m) and v is kinematic viscosity (m^2/s).

Turbulence breaks down the air into small unstable pockets of air which are called eddies. Characteristics such as pressure, density, temperature and refractive index are constant inside these eddies. When one of these eddies' altitude change, assuming there's no heat transfer between the eddy and its environment, its temperature changes with adiabatic lapse rate, which is 9.8 K/km. If this adiabatic lapse rate is different than the lapse rate of that part of the atmosphere, that eddy will have a different temperature which means it will have a different refractive index than its surroundings. This will result in optical turbulence since mechanical turbulent motion alone doesn't suffice to have optical turbulence, refraction index fluctuations are also needed.

To model the turbulence, Kolmogorov came up with a cascade theory of turbulence. As shown in Figure 1.1, larger eddies break up into smaller ones due to inertial forces and the energy is transferred from macroscale L_0 to microscale l_0. L_0 is called outer scale and l_0 is called inner scale. Energy is dissipated as heat for lower scales than the inner scale (Andrews and Phillips, 2005).

Since the effects of turbulence depend on the difference between locations rather than the absolute values, structure-function is used to define it. Kolmogorov showed that wind velocity's structure-function satisfies 2/3 power law;

$$D_V(\vec{r}) = \langle [V(\vec{r_1}) - V(\vec{r_2})]^2 \rangle = C_V^2 r^{2/3}, \qquad l_0 < r < L_0 \tag{1.2a}$$

$$\vec{r} = \vec{r_1} - \vec{r_2} \tag{1.2b}$$

Where $\langle \quad \rangle$ is assemble average, C_V^2 is velocity structure parameter and $\vec{r_i}$ is the location.

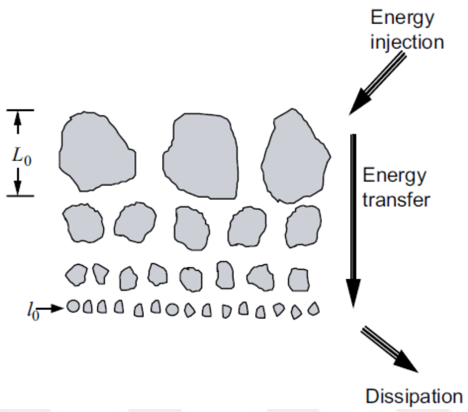

Figure 1.1. Kolmogorov cascade theory of turbulence (Andrews and Phillips, 2005).

The range in which this structure function is applicable is called inertial subrange and it is bounded by l_0 and L_0 as stated in (1.2a). C_V^2 is defined as;

$$C_V^2 = 2\varepsilon^{2/3} \tag{1.3}$$

Where ε is the average energy dissipation rate.

Kolmogorov's theory is also applicable to both potential temperature and refractive index structure fluctuations and they are also governed by the 2/3 law;

$$D_T(\vec{r}) = \langle [T(\vec{r_1}) - T(\vec{r_2})]^2 \rangle = C_T^2 r^{2/3}, \qquad l_0 < r < L_0 \tag{1.4a}$$

$$D_n(\vec{r}) = \langle [n(\vec{r_1}) - n(\vec{r_2})]^2 \rangle = C_n^2 r^{2/3}, \qquad l_0 < r < L_0 \tag{1.4b}$$

Where T is potential temperature, C_T^2 is temperature structure parameter, n is refractive index and C_n^2 is refractive index structure parameter. Since subtraction eliminates big fluctuations that affect both measuring points, structure-function acts like a high-pass filter and offers more stable results than the correlation function.

The relationship between C_T^2 and C_n^2 is given as (Owens, 1967; Andrews and Phillips, 2005);

$$C_n^2 = \left(79 \times 10^{-6} \frac{P}{T^2}\right) C_T^2 \tag{1.5}$$

Where P is pressure.

There are many possible methods to measure C_n^2. Scintillometer is an instrument which measures C_n^2 over a short path. For another method, the potential temperature of two points can be obtained via two fine wire thermometer and from (1.4a) temperature structure function can be found and by using (1.5) C_n^2 can be obtained. Also, it is possible to estimate C_n^2 by meteorological data.

1.1 Damaging Effects of Turbulence

There are five different damaging effects of turbulence and they happen, depending on the eddy size that laser beam encounters;

- Beam wander,
- Angle of arrival fluctuations,
- Scintillation,
- Beam spreading,
- Phase fluctuations.

Beam wander and angle of arrival fluctuations occur when laser beams veer off as a whole because of the eddies larger than them. Eddies with the size close to the beam diameter act as a diverging or converging lens on some part of or the whole beam which result in constructive or destructive interference in beam cross-section leading to fluctuations in amplitude which is called scintillation. When laser beam encounters smaller eddies than their diameter, beamlets are scattered and refracted independently of each other resulting in beam spreading and phase fluctuations.

1.2 Aim of the Thesis

Damaging effects are some of the biggest problem encountered by system engineers while designing outdoor optical wireless communication systems and thus modelling optical turbulence correctly is very crucial. To simplify the model, l_0 is assumed to be zero and L_0 is assumed to be infinity which makes the structure-function applicable everywhere not just in inertial subrange. This makes modelling refractive index fluctuations possible with just one parameter, C_n^2.

As will be explained in detail, later in the literature review chapter, work on finding the best probability distribution function (pdf) to fit C_n^2 started more than half-century ago and still, there are different views. Lognormal distribution which is a semi-infinite (bounded on only one side) distribution and beta distribution which is a bounded distribution can be used as a model.

Since one in a millionth or even billionth errors affects the performance of the optical wireless communication systems, extreme values of modelled pdfs are very important. So the first aim of this thesis was finding out whether C_n^2 has an upper bound, like beta distribution, or doesn't have one, like lognormal distribution. To do so, two different approaches are used. The first one is calculating the third-moment skewness (S) and fourth-moment kurtosis (K) and using Skewness-Kurtosis plane to decide the boundedness via Pearson and Johnson distribution systems. Although this approach is used in different fields application, as far as known it hasn't been applied to the analysis of C_n^2. The second approach is applying the Extreme Value Theory (EVT) to examine the extreme values of C_n^2 and this approach also hasn't been applied to the analysis of C_n^2 yet. After that, some of the known pdf parameter estimation methods are implemented and their results are compared. Since the results in the first part suggest boundedness, the focus on parameter estimation methods are on beta distribution.

2. LITERATURE REVIEW

During this thesis work, a lot of different theories and approaches are encountered concerning both turbulence and statistical analysis of a given data set. The ones, used during the thesis are given in this chapter.

2.1 Probability Distribution of Turbulence Parameters

With the possibility of modelling optical turbulence with only refractive index structure parameter, research on the best fit of probability density function (pdf) for C_n^2 began immediately. At first, lognormal distribution was suggested as the best model but later on, the advantages of using beta distribution are found.

Obukhov (1962) proposed that for large Reynolds number, average dissipation value between points r_1 and r_2 has lognormal distribution.

Gurvich and Yaglom, (1967) by using Kolmogorov's cascade theory of turbulence and central limit theorem, theoretically showed that logarithm of dissipation is normally distributed. Given the big eddies breakdown into smaller eddies under the effect of turbulence as stated in the Kolmogorov's cascade theory of turbulence, for ε is the dissipation rate and α_i is the ratio of the average of the dissipation rate in the ith breakdown of an eddy;

$$\overline{\varepsilon_n} = \varepsilon_0 \alpha_1 \alpha_2 \alpha_3 \dots \alpha_n \tag{2.1a}$$

$$\log \overline{\varepsilon_n} = \log \varepsilon_0 + \sum_{i=1}^{n} \log \alpha_i \tag{2.1b}$$

The central limit theorem states that for a set that's consisting of independent random variables, as the sample size gets larger, sampling distribution approaches to the normal distribution, assuming α_i is independent. So for big enough n, $\log \overline{\varepsilon_n}$ is normally distributed which means dissipation is lognormal distributed.

Stewart et al., (1970) by examining the data of velocity fluctuations in the boundary layer, showed that although lognormal distribution gives reasonably good results in general for describing the derivative of velocity, significant deviations occur for very small and very large values.

Chen (1971) showed that although the probability density function of velocity derivative approaches infinity for smaller values and thus cannot be lognormal, the average of squared velocity derivative shows lognormality between the bounds of inner and outer scale.

Chadwick and Moran, (1980) analysed year long C_n^2 data and concluded that it can be modelled as a lognormal random variable.

Andreas (1989) showed that although beta distribution underestimates the mode (peak of the histogram), it predicts the location of mode better than lognormal distribution.

Andreas et al., (2003) reached a conclusion that for modelling C_n^2 and l_0, the beta distribution is applicable. Although they couldn't theoretically prove why beta distribution is better as a fit than lognormal distribution, the explanation given was broad usability of beta distribution as it can approximate different kinds of distributions, including lognormal distribution (Harr, 1987).

2.2 Selecting a Suitable Probability Distribution Function

Huyse and Thacker, (2004) studied on selecting a proper pdf to model a given data set. They stated that, especially with insufficient data, choosing pdf is not an exact science. Chosen pdfs in the literature and judgment of the researcher play a big role in the selection yet there are methods to ensure to limit the possible suitable pdfs. Pearson and Johnson distribution systems match every possible Skewness-Kurtosis combination to a unique distribution family. Another possible approach is examining the behaviour of the tails to determine the behaviour of the parent distribution using Extreme Value Theory (EVT).

2.2.1 Pearson and Johnson distribution systems and their applications

Pearson (1895, 1901) first came up with the idea of using Skewness-Kurtosis plane to classify distributions and defined curve equations that separate distribution types.

Johnson (1949) proposed another set of curves for classification of distribution types derived by the method of translation.

DeBrota et al., (1988) used Johnson distribution system to model univariate populations as an input for simulation experiments.

While working on turbulent dispersion, Mole and Clarke, (1995) showed that when skewness is plotted against kurtosis, data collapsed onto a single curve.

Chatwin et al., (1995) used skewness-kurtosis plane and Pearson distribution system to choose the probability distribution that models turbulent dispersion and showed that beta distribution is a good fit.

Podladchikova et al., (2003) used Pearson distribution system to classify coronal heating probability densities.

George (2007) used Johnson distribution system to analyse microarray data.

Graf et al., (2010) suggested using the skewness-kurtosis plane to decide the boundedness of turbulent temperature using Pearson distribution system.

2.2.2 Extreme Value Theory and its applications

Fisher and Tippett, (1928) suggested that block maxima of a set of random variables follow one of the three types of distributions (Frechet, Weibull or Gumbel type). Gnedenko (1943) generalized these three distribution types into a single distribution, in which shape parameter c determines which type of distribution (Figure 2.1), leading to Generalized Extreme Value (GEV) distribution and Fisher-Tippett-Gnedenko theorem, also known as the Fisher-Tippett theorem or first theorem in Extreme Value Theory. Balkema and De Haan, (1974) and Pickands III (1975) showed that for a set of random variables, all values that exceed some certain threshold follows generalized Pareto distribution (GPD) which leads to Pickands-Balkema-De Haan theorem, also known as the second theorem in Extreme Value Theory. Shape parameter k of GPD determines the behaviour of the tail (Figure 2.2). When k is equal to 0, GPD becomes exponential distribution. The upper tail of the distribution is bounded for positive values of k and unbounded for negative values of k.

Smith (1989) found out that although the overall data of ozone in Texas showed no particular trend, its extreme values showed a downward trend. Küchenhoff and

Thamerus (1995) also worked on air pollution using Extreme Value Theory. They showed that both GEV and GPD are applicable and useful.

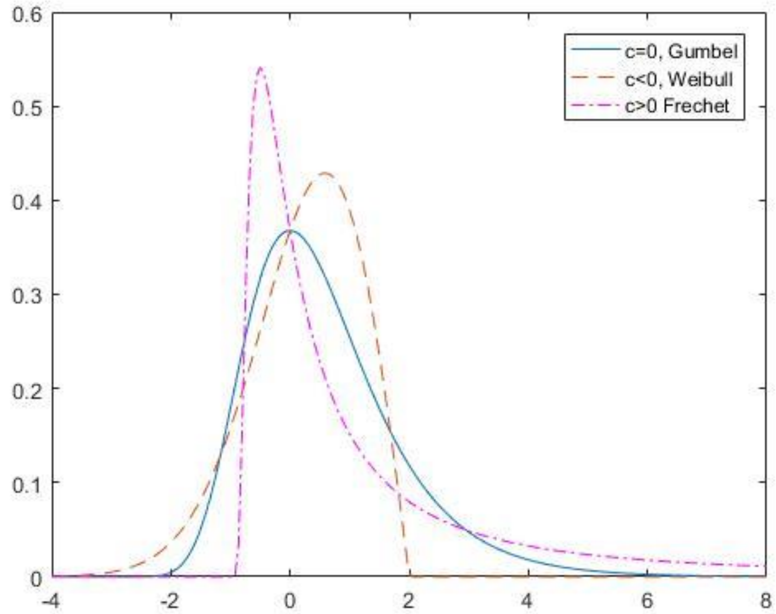

Figure 2.1. PDF of GEV Distribution.

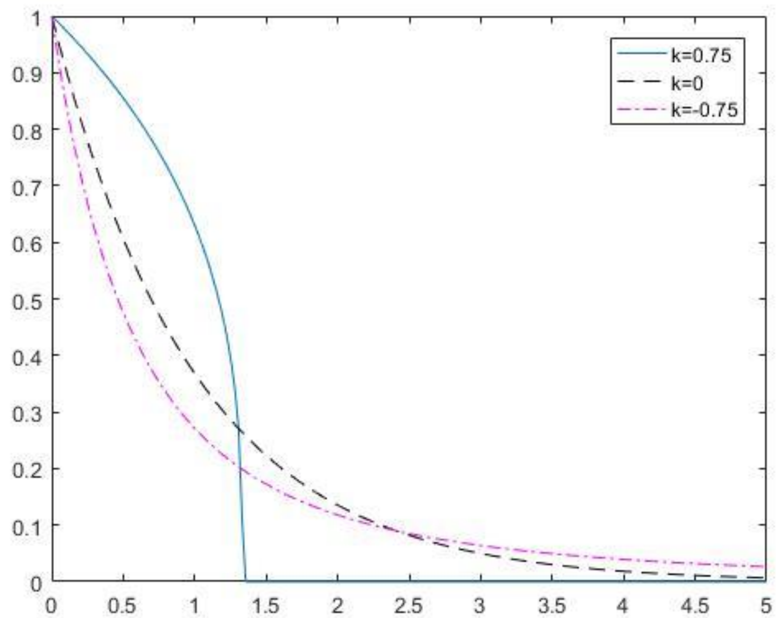

Figure 2.2. PDF of GPD.

De Haan (1990) applied Extreme Value Theory to sea level data in the Netherlands to calculate the safe height of sea dikes to prevent floods.

McNeil (1997) applied Extreme Value Theory for pricing of high-excess loss layers in insurance using large fire insurance losses in Denmark and found out that GPD provides a good estimation for the tails of loss severity distributions which are the main concern of insurance cases.

Gilli and Kellezi, (2006) applied Extreme Value Theory to financial portfolios where predicting extreme events is very important. They applied the first and the second theorems of extreme value theory and talked about the advantages and disadvantages of both theorems. In their application, the second theorem of Extreme Value Theory gave better results.

Acero et al., (2018) used Pickands-Balkema-De Haan theorem to show that during the telescopic period possible minimum and possible maximum solar activity is already recorded and solar activity won't exceed these limits in the near future.

To decide which cryptocurrencies are the least risky and which are the riskiest, Gkillas and Katsiampa, (2018) applied Extreme Value Theory and used GPD to model the behaviour of their distributions' tails.

Weng et al., (2018) worked on safety of the neural network systems, which is very important for some of their applications such as malware detection and autonomous driving systems, using Extreme Value Theory for evaluation.

2.2.2.1 Exploratory inspection methods for Extreme Value Theory

While applying the second theorem in Extreme Value Theory choosing a correct threshold becomes the key to find boundedness of the data. Mean excess function plot of the data is examined to choose the proper threshold. In addition to that, examining the general behaviour of the mean excess function plot gives hints about boundedness. QQ plots, although they are not used in choosing threshold, are also another inspection method to decide boundedness.

Embrechts et al., (1997) showed how to use exploratory methods such as QQ plots and mean excess function to model extremal events.

Coles (2001) also explained exploratory methods and how to use mean excess function plots to select the threshold for Pickands-Balkema-De Haan theorem, considering different aspects such as breakpoints of the trend in the plot or sample size remained for GPD.

2.2.2.2 GPD shape parameter estimation

Since shape parameter k of GPD determines the boundedness of the distribution, estimating it correctly becomes very important. In literature, there's no proven way to estimate it but there are different methods such as probability weighted moments (PWM), method of moments (MoM), L-moments and higher order L-moments (LH-moments) that give good estimations.

Greenwood et al., (1979) suggested using PWM for estimating parameters of different distributions and showed that it is more useful in some cases where conventional moments can't be used.

Landwehr et al., (1979) showed that PWM produces lesser biased results compared to the maximum likelihood method. Hosking et al., (1985) also showed that PWM is better to estimate shape parameters of GEV than maximum likelihood method since it doesn't have a severe bias.

Hosking and Wallis, (1987) on their study of how to estimate GPD shape parameters, showed that using MoM or PWM will give better results than the maximum likelihood method.

Hosking (1990) examined L-moments which are linear combinations of the PWM and their advantage over other estimation methods, which is less independence to the sample size. He showed that with smaller sample sizes, L-moments provide as good a result as the maximum likelihood method, with less computation.

Wang (1997) used LH-moments which are linear combinations of higher order statistics and showed that LH- moments are even better than L-moments at predicting extreme events in small sample sizes.

Meshgi and Khalili, (2009) developed formulae to predict the shape parameter of GPD using LH-moments.

Bermudez and Kotz, (2010) put different methods to estimate GPD parameters together and explained their advantages and disadvantages and which applications and fields they are best to be used.

2.3 Beta Distribution Parameters Estimation

Since the results of this research mostly indicated boundedness for C_n^2, methods of estimating beta distribution parameters are also examined.

Johnson et al., (1995) gave formulae to predict beta distribution shape parameters using MoM.

AbouRizk (1990) and AbouRizk et al., (1994) came up with an improvement of MoM where after computing the initial estimation of shape parameters of beta distribution, by using ordinary least square (OLS) and diagonally weighted least square (DWLS) methods, shape parameters are optimized to give the best fit for cdf of sample data.

3. MATERIAL AND METHODS

3.1 Material

Two independent data sets, eleven subsets -six coming from the first, five coming from the second data set- is used for this research. Brief description of these data sets will be given and for more information about the measurements, respective reports can be referred.

3.1.1 Channel characterization for free-space optical communications project data

This project is funded by The Defense Advanced Research Projects Agency (DARPA) and the final report "Channel Characterization for Free-Space Optical Communication" is open for public access (Andrews et al., 2012). Six figures that contain C_n^2 measurements are chosen considering time and location proximity of the measurements and also the number of samples in the set. Measurements are from; between Hollister and Fremont at 07/06, between Hollister and Fremont at 09/06, Hollister at 07/06, Hollister at 08/06 (between 13:23-15:00), Hollister at 08/06 (between 16:30-18:15), Hollister at 09/06 which are taken from figures 6, 8, 9, 10, 11 and 12 respectively and to prevent clutter and make them easy to recognise, from now on they will be referred to as Hollister – Fremont (1), Hollister – Fremont (2), Hollister (1), Hollister (2), Hollister (3) and Hollister (4) respectively. These figures are imported to an image processing program by using a screenshot application and coordinates of each data point are precisely determined. First four moments of each data subset are calculated and given in Table 3.1.

3.1.2 TUBITAK data

Raw data of measurements from BLS-900 Scintillometer, which is used by researchers at TUBITAK (The Scientific and Technological Research Council of Turkey) in their research of modelling C_n^2 by the artificial neural network, Bacı et al., (2018) are taken. Except for the day that's missing some data, five out of six consecutive days of measurements are chosen between 03/11/2017 and 08/11/2017, leaving out 06/11/2017 with the missing data, and measurements of every day is taken as a subset and called TUBITAK (1), TUBITAK (2), TUBITAK (3), TUBITAK (4) and

TUBITAK (5) in daily order. First four moments of each data subset are calculated and given in Table 3.2.

Table 3.1. Channel characterization for free-space optical communications data.

Data	Sample Size	Mean	Variance	Skewness	Kurtosis
Hollister – Fremont (1)	82	$2.23x10^{-15}$	$6.31x10^{-31}$	1.0543	3.5492
Hollister – Fremont (2)	89	$1.97x10^{-15}$	$3.99x10^{-31}$	0.9457	3.8545
Hollister (1)	101	$1.22x10^{-13}$	$9.18x10^{-27}$	0.9037	2.3588
Hollister (2)	203	$1.64x10^{-12}$	$2.40x10^{-25}$	0.3650	2.3892
Hollister (3)	131	$6.41x10^{-13}$	$6.48x10^{-26}$	0.5848	2.8655
Hollister (4)	157	$1.96x10^{-12}$	$1.94x10^{-25}$	0.0817	2.3929

Table 3.2. TUBITAK data.

Data	Sample Size	Mean	Variance	Skewness	Kurtosis
TUBITAK (1)	2880	$2.83x10^{-13}$	$6.59x10^{-26}$	0.7825	2.4249
TUBITAK (2)	2880	$3.19x10^{-14}$	$1.66x10^{-27}$	1.9500	5.9337
TUBITAK (3)	2880	$1.02x10^{-14}$	$8.24x10^{-29}$	3.6075	27.0376
TUBITAK (4)	2880	$1.10x10^{-13}$	$9.87x10^{-27}$	1.0780	3.4151
TUBITAK (5)	2880	$9.53x10^{-14}$	$4.43x10^{-27}$	0.7243	2.7779

3.2 Methods

Since there's no consensus on which distribution is best to model C_n^2, narrowing the possibilities is the first approach, by using Pearson and Johnson distribution systems and checking the boundedness of C_n^2. Another way of checking boundedness is by using Extreme Value Theory and inspecting the upper part of the measurements since the real interest is showing if C_n^2 has an upper bound or not. Instead of Fisher-Tippett-Gnedenko theorem which relies on block maxima of the data, Pickands-Balkema-De

Haan theorem which uses samples that exceeds some certain threshold is used. Reason of choosing that theorem was because some data sets used in this research have as low as 82 samples in them, so dividing them into equal-length subsets would have left those subsets with only couple of data points which could disrupt the result. After picking the most suitable threshold for each data set, subsets which consist of the samples exceeding that threshold is created. The resulting subsets are inspected for boundedness via shape parameter estimation of GPD. Mean excess function plots are used as a tool for deciding the thresholds of data sets and also as an exploratory inspection of boundedness by checking the trend of them. QQ plots are also used in Extreme Value Theory to check the boundedness. After sorting them in ascending order, samples are plotted against quantiles of the exponential distribution. Convexity of the end part of the graph means the data set is bounded while concave result means it is unbounded. The exponential distribution is regarded as the limits of bounded distributions. Convex behaviour means sample values are smaller than exponential distribution estimates which means the data set is bounded.

In the second part of the research most known and used distributions of bounded and semi-infinite distribution families, beta and lognormal distributions respectively, are tested with the goodness of fit tests Chi-Square test and K-S test and also they are compared with each other.

3.2.1 Pearson distribution system

Although there are seven types of families in Pearson distribution system, the focus of this research is whether the distribution is bounded or not. Skewness-Kurtosis (S-K) plane is divided into four regions and three curves separating them are used. After that, the skewness (S) and kurtosis (K) pair of each data set are scattered in S-K plane. The first curve is the one separating bounded distributions and impossible area;

$$K = S^2 + 1 \tag{3.1}$$

The second curve is separating bounded and semi-infinite distributions;

$$K = \frac{3S^2}{2} + 3 \tag{3.2}$$

The last curve is separating semi-infinite and unbounded distributions;

$$S^2(K+3)^2 = 4(4K-3S^2)(2K-3S^2-6) \tag{3.3}$$

3.2.2 Johnson distribution system

The curve separating bounded distributions and impossible area for Johnson distribution system is the same as the one for Pearson distribution system but the rest of the regions are different. Johnson distribution system don't have semi-infinite distributions region like Pearson distribution system but has a curve separating bounded and unbounded distributions (Lognormal system) which S and K is defined by a parameter ω;

$$S = (\omega+2)\sqrt{w-1} \tag{3.4a}$$

$$K = \omega^4 + \omega^3 + \omega^2 - 3 \tag{3.4b}$$

3.2.3 Extreme Value Theory

Since the second theorem of Extreme Value Theory is used as mentioned earlier, GPD is used as a distribution whose cumulative distribution function (cdf) is defined as;

$$F(x|k,\xi,\sigma) = \begin{cases} 1-\left(1-k\dfrac{x-\xi}{\sigma}\right)^{1/k}, k \neq 0, \\ 1-\exp\left(-\dfrac{x-\xi}{\sigma}\right), \quad k=0. \end{cases} \tag{3.5}$$

Where k is the shape parameter, ξ is the location parameter and σ is the scale parameter. GPD has a finite end point for the positive values of k.

3.2.3.1 Mean excess function

Mean excess function is plotted after calculating;

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}(x_i-u)\right): u < x_{max}\right\} \tag{3.6}$$

For x_i is the sample that exceeds the threshold u, n_u is the number of samples in the subset and x_{max} is the maximum value of the data set.

A known method for choosing u is dividing u axis into semi-equal length between x_{max} and the minimum value of the data set x_{min};

$$u_i = x_{min} + i\left(\frac{x_{max} - x_{min}}{100}\right) : i = 1,2,3 \dots ,100 \tag{3.7}$$

Another method is proposed in this research where u axis is not plotted linearly but each data sample is used as a threshold after the data set is sorted in ascending order;

$$u_i = x_i : i = 1,2,3, \dots , n \tag{3.8}$$

Where n is the number of samples in the data set.

Exploratory inspection of mean excess function is done by checking the trend of the graph. If it shows a downward trend after threshold, this means that the data set is bounded while an upward trend means the data set is unbounded. Mean excess function of lognormal and beta distributions in Figures 3.1 and 3.2 respectively show the behaviour of the distributions. Figure 3.1 also shows that due to the nature of the traditional mean excess function, there's a downward trend between the samples. This happens because although threshold u in (3.6) is increasing, the number of samples that exceed the threshold and their sum doesn't change until the threshold value reaches to the next sample. With the proposed method there's no such issue because thresholds are picked as sample values.

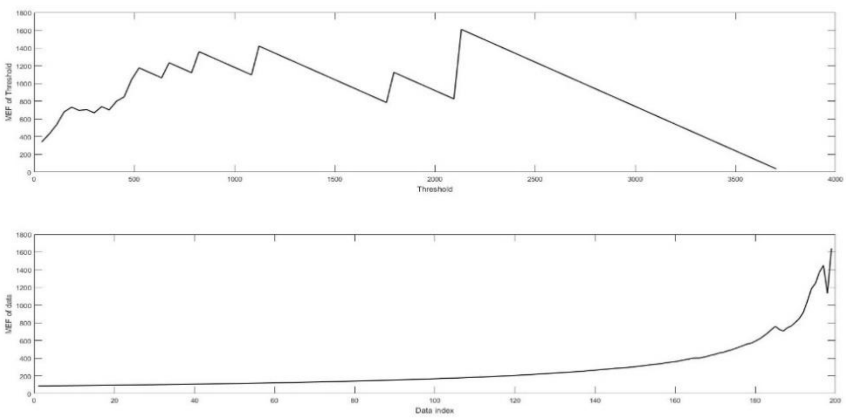

Figure 3.1. Mean excess function of the lognormal distribution.

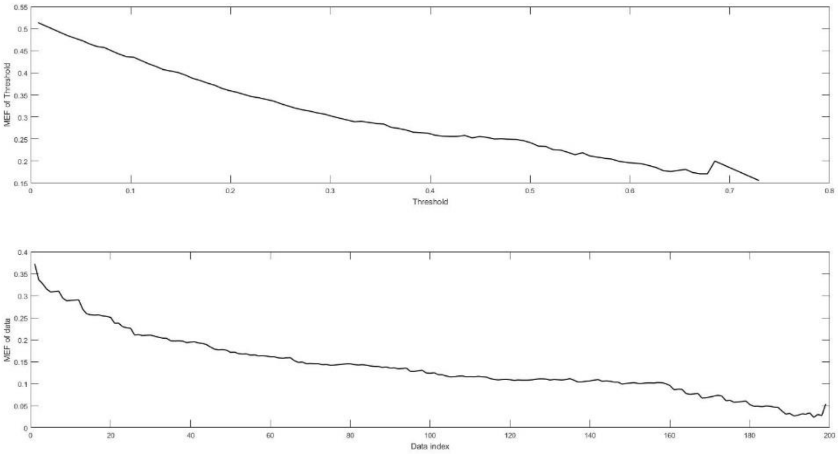

Figure 3.2. Mean excess function of the beta distribution.

3.2.3.2 Estimation of shape parameter values of GPD

There are lots of different methods proposed for estimating shape parameter (k) but in this research three of them, method of moments (MoM), probability weighted moments (PWM) and higher order L-moments (LH-moments) are used.

For MoM, k is estimated as;

$$k = \frac{1}{2}\left(\frac{\bar{x}}{\sigma^2} - 1\right) \tag{3.9}$$

Where $\bar{x}$ is the mean and σ^2 is the variance.

For PWM, k is estimated as;

$$k = \frac{\bar{x}}{\bar{x} - 2\alpha_1} - 2 \tag{3.10}$$

Where there are two different ways to calculate α_1;

$$\alpha_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \left(1 - \frac{i - 0.35}{n}\right) \tag{3.11a}$$

$$\alpha_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{n - i}{n - 1} x_i \tag{3.11b}$$

For zeroth order meaning L-moments, LH-moments method for estimating k is;

$$k = \frac{(\eta+3)\left(1-3\bar{\tau}_3^{\eta}\right)}{3\bar{\tau}_3^{\eta}+\eta+3} \tag{3.12}$$

Where

$$\bar{\tau}_3^{\eta} = \frac{\lambda_3^{\eta}}{\lambda_2^{\eta}} \tag{3.13}$$

Given

$$\lambda_2^{\eta} = \frac{1}{2}\frac{1}{\binom{n}{\eta+2}}\sum_{i=1}^{n}\left[\binom{i-1}{\eta+1}-\binom{i-1}{\eta}\binom{n-i}{1}\right]x_i \qquad (3.14\ a)$$

$$\lambda_3^{\eta} = \frac{1}{3}\frac{1}{\binom{n}{\eta+3}}\sum_{i=1}^{n}\left[\binom{i-1}{\eta+2}-2\binom{i-1}{\eta+1}\binom{n-i}{1}+\binom{i-1}{\eta}\binom{n-i}{2}\right]x_i \quad (3.14\ b)$$

Where λ is the L-moment and η is the order.

Since MATLAB has a limit for how big a number can be, only the first three order are calculated. Thresholds for GPD are also chosen with this limit in mind without compromising the result.

3.2.3.3 QQ plot

Quantile function of the exponential distribution is;

$$F^{-1}\left(\frac{i}{n+1}\right) = \frac{1}{\lambda}\left(\frac{n+1}{n+1-i}\right) \tag{3.15}$$

Where λ is the rate parameter. QQ plot is plotted by scattering data points;

$$\left\{\left(x_i, F^{-1}\left(\frac{i}{n+1}\right)\right) : i = 1,2,3,\dots,n\right\} \tag{3.16}$$

In Figures 3.3 and 3.4 lognormal distribution and beta distribution QQ plots against exponential distribution are shown respectively. As it can be seen, while exponential distribution underestimates the sample values when it is plotted against lognormal which results in concave behaviour, against beta distribution it overestimates and the plot shows convex behaviour. This means that compared to the exponential distribution, lognormal distribution has a heavier tail and beta distribution has a lighter tail.

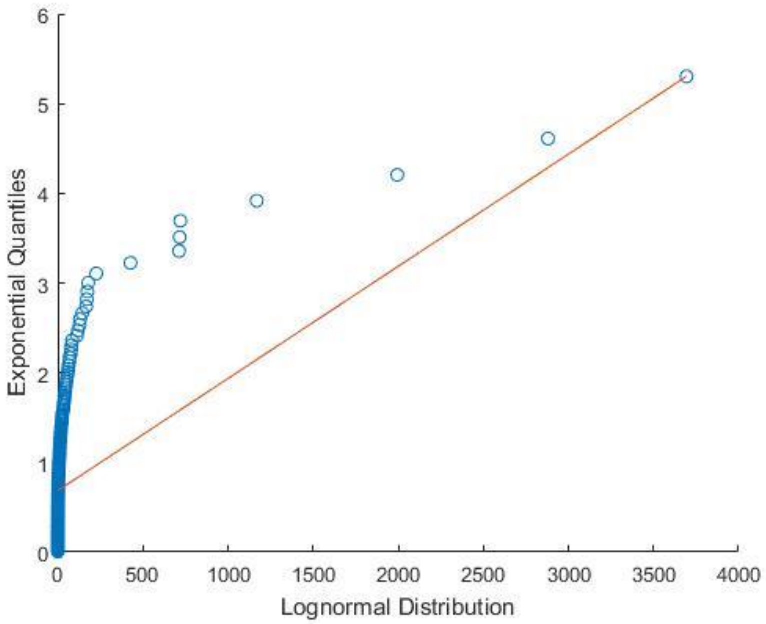

Figure 3.3. QQ plot of the lognormal distribution.

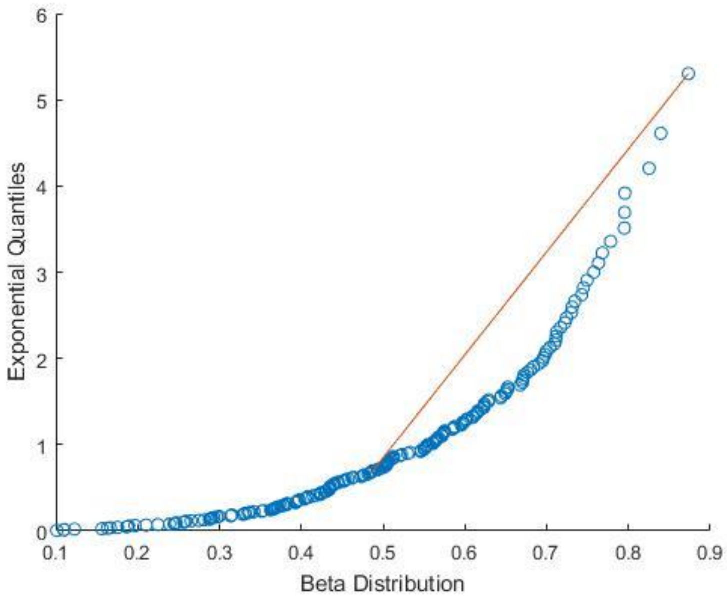

Figure 3.4. QQ plot of the beta distribution.

3.2.4 Beta distribution parameter estimation

Just like the estimation of GPD distribution parameters, beta distribution also has different methods that can be applied. Out of them, two methods were selected. First one is the method of moments (MoM) which uses skewness and kurtosis values of the set to estimate shape parameters p and q;

$$p, q = \frac{r}{2}\left[1 \underline{+} (r+2)\sqrt{\frac{S^2}{(r+2)S^2 + 16(r+1)}}\right] \begin{cases} q < p \; for \; S < 0 \\ p < q \; for \; S > 0 \end{cases} \quad (3.17)$$

Given

$$r = 6\frac{K - S^2 - 1}{6 + 3S^2 - 2K} \quad (3.18)$$

The second method is an improvement over MoM. The least square error between cdf of the data set and beta distribution with shape parameters estimated by MoM is tried to be minimized. The minimization problem is;

$$Minimize \sum_{i=1}^{n} w_i \left[F(x_i|p, q) - \frac{i}{n+1}\right]^2 \quad (3.19)$$

Where w_i is the weight of each point and $F(x_i|p, q)$ is cdf of beta distribution with parameters p and q.

For picking w_i, there are two approaches that can be used. In ordinary least squares (OLS) all points have the same weight so w_i is equal to 1. In diagonally weighted least squares (DWLS) w_i is given as;

$$w_i = \frac{(n+2)(n+1)}{i\left(1 - \frac{i}{n+1}\right)} \quad (3.20)$$

3.2.5 Lognormal distribution shape parameters

The lognormal distribution is defined by two parameters $\mu^l$and σ^l whose formulae are;

$$\mu^l = ln\left(\frac{\bar{x}^2}{\sqrt{\bar{x}^2 + \sigma^2}}\right) \qquad (3.21a)$$

$$\sigma^l = \sqrt{ln\left(1 + \frac{\sigma^2}{\bar{x}^2}\right)} \qquad (3.21b)$$

3.2.6 Goodness of fit tests

Although there are various goodness of fit tests, K-S test for continuous data and Chi-Square test for discrete data are one of the most known ones. K-S test is applied by measuring the deviation of the suggested distribution’s cdf from the cdf of sample data. Since the data is continuous, to apply Chi-Square test histograms of the sample data and suggested distribution are compared. There are different formulae to decide the bin size for histograms (Sturges, 1926; Doane, 1976; Freedman and Diaconis, 1981). After thorough research of these formulae, bin sizes selected for this research are, 8 for data set with samples less than 200, 10 for the data set with 203 samples and 16 for data sets with 2880 samples.

4. RESEARCH FINDINGS

In this section results of every method mentioned earlier will be given for each data set separately and discussed briefly at first and then all results will be combined together to inspect the general behaviour of C_n^2.

4.1 Hollister – Fremont (1) Data Results

This data set is inside the region of the bounded distributions for both Pearson and Johnson distribution systems (Figure 4.1).

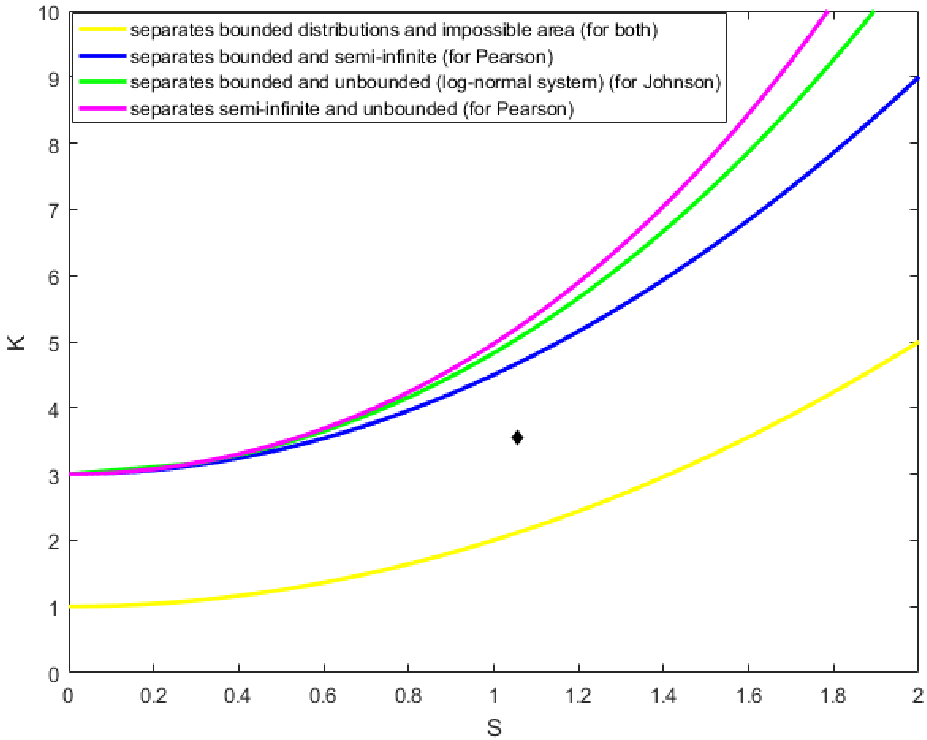

Figure 4.1. Pearson and Johnson distribution systems result of Hollister – Fremont (1) data.

Mean excess function of the data set shows a general downward trend although it stays at the same level for a while in the middle (Figure 4.2). The threshold for the Extreme Value Theory is chosen as 45% which leaves 16 samples for GPD. Resulting k values estimations, given in Table 4.1 are all positive. The upper end of the QQ plot, given in Figure 4.3, shows convexity. All of the examinations show boundedness.

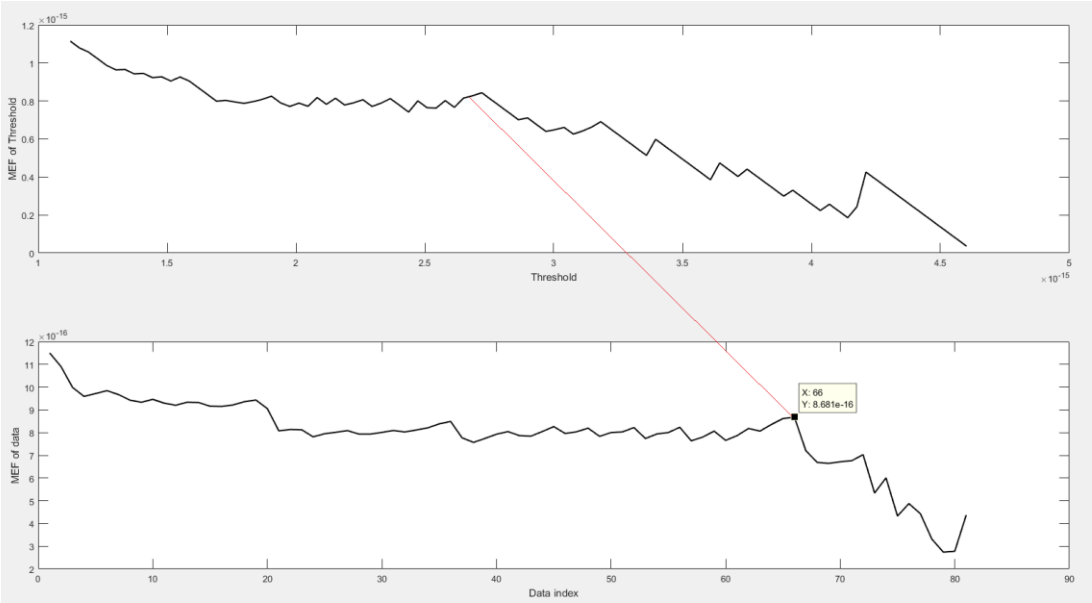

Figure 4.2. Mean excess function of Hollister – Fremont (1) data.

Table 4.1. k value estimation of Hollister – Fremont (1) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
24.91673	8.210283	9.838544	0.503861	0.572082	0.561317	0.477596

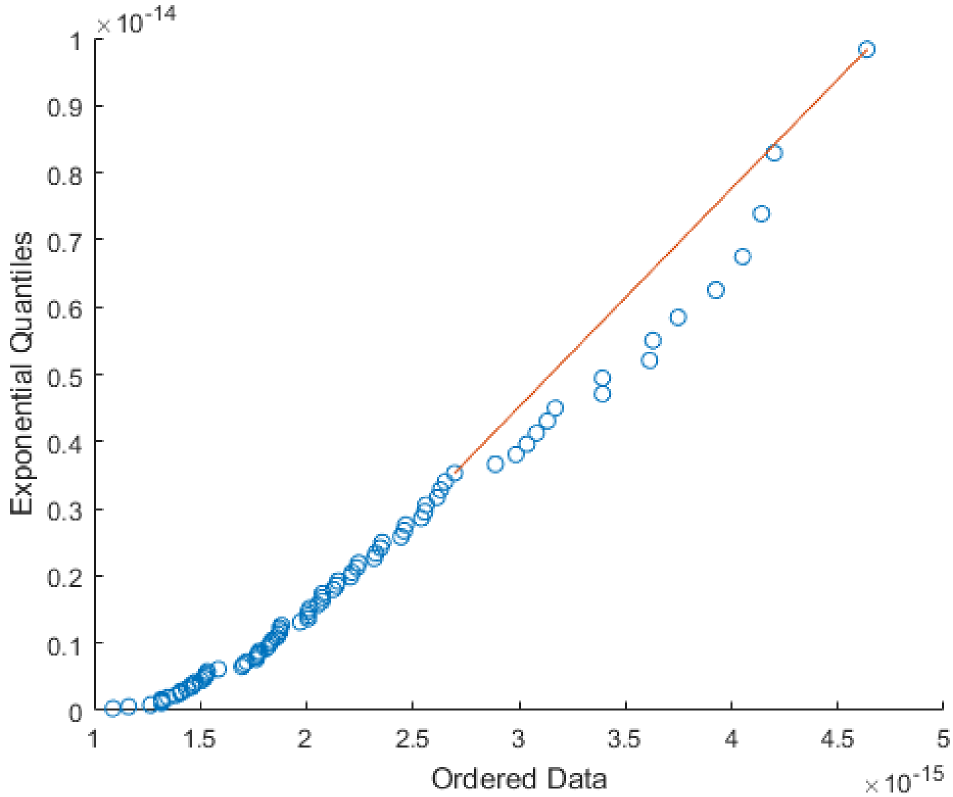

Figure 4.3. QQ plot of Hollister – Fremont (1) data.

Inspection of the cdf graph given in Figure 4.4 shows that beta distribution with parameters calculated via MoM is a very bad fit compared to other methods. Error graph in Figure 4.5 shows that OLS and DWLS results are pretty close to each other and for some parts, they are a better fit for the empirical cdf and for some other parts lognormal is better.

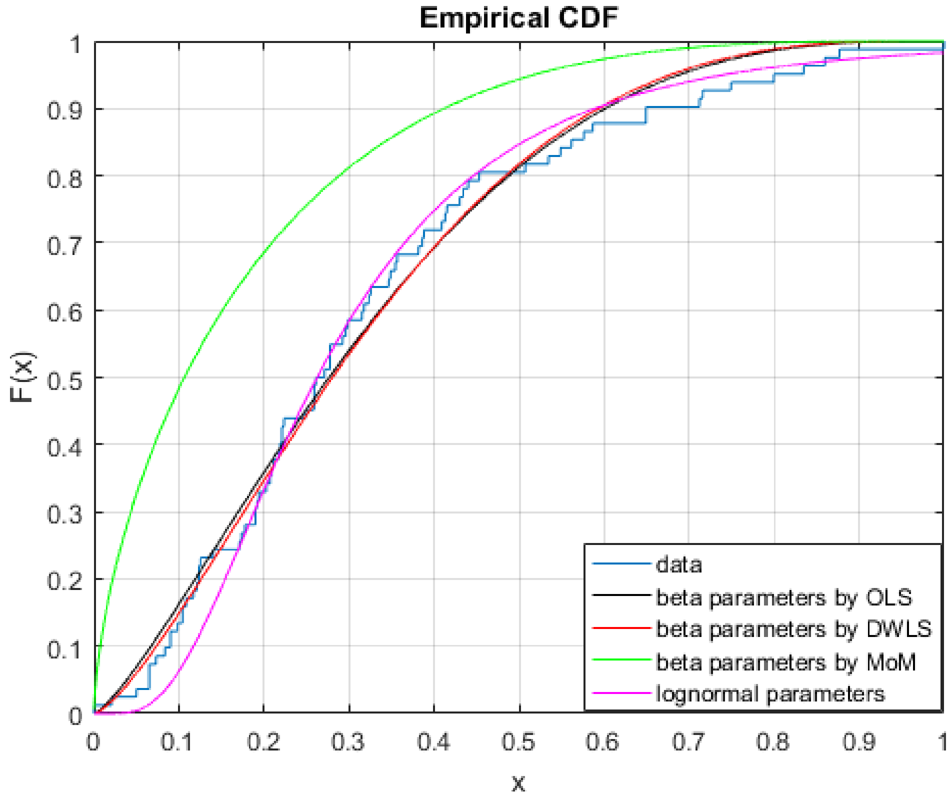

Figure 4.4. Cdf fitting of Hollister – Fremont (1) data.

For this data set and almost all others, including MoM results in error comparison graphs makes it impossible to read because they have a lot bigger error compared to others so in the error comparison graphs only results of OLS, DWLS methods are compared to lognormal distribution results. It should be noted that error in error comparison graph is actually error squared for each data point for two reasons, one is to make them positive and two implement them directly to least square methods.

Chi-Square and K-S test results are given in Table 4.2. Except for the MoM, the other three passed both tests while MoM fails both. As they are very similar to each other, OLS and DWLS methods' results were very close and they are better than lognormal. Results of OLS are a little bit better than DWLS and although it passes lognormal results are very close to critical values for both tests.

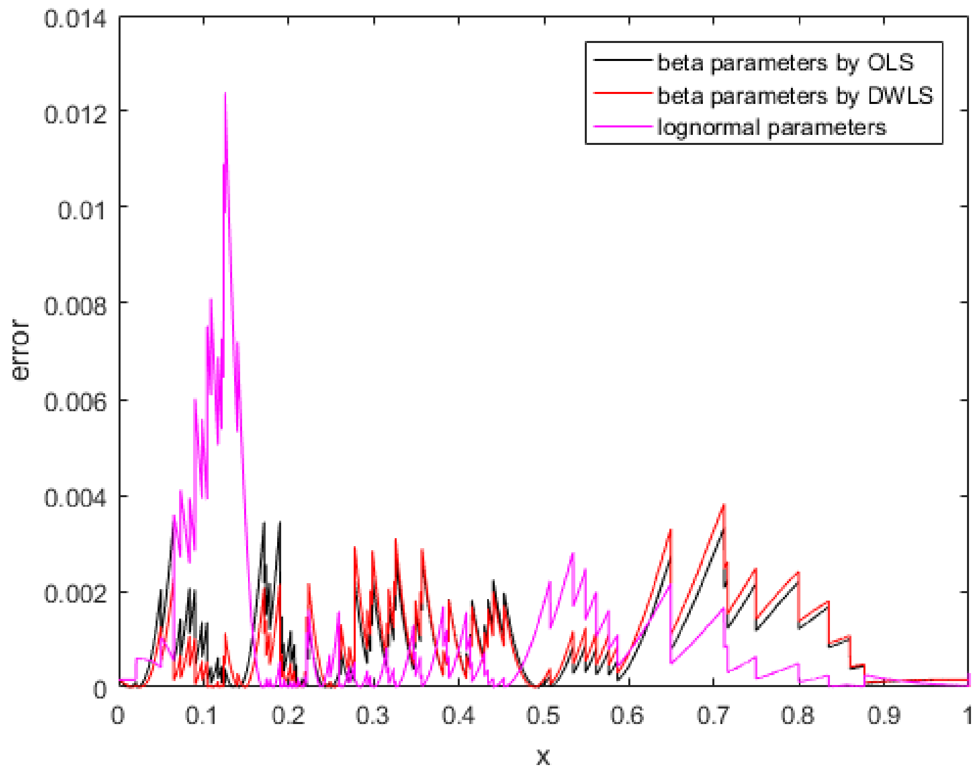

Figure 4.5. Error comparison for Hollister – Fremont (1) data.

Table 4.2. Hollister – Fremont (1) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	0	0	**4.0152**	**0.0590**
DWLS	0	0	4.5854	0.0618
MoM	1	1	51.8292	0.3935
Lognormal	0	0	12.5491	0.1114

4.2 Hollister – Fremont (2) Data Results

Although it is close to the curve separating bounded and semi-infinite distributions, data set is inside the region of the bounded distributions for both Pearson and Johnson distribution systems (Figure 4.6).

Mean excess function in Figure 4.7 starts with a downward trend then shows an upward trend in the middle and finish with a downward trend so it shows signs of boundedness. The threshold is chosen 68% and that leaves only 6 samples for GPD distribution. Table 4.3 shows the shape parameter estimation. For the chosen subset QQ plot in Figure 4.8 shows convexity albeit only 6 samples considered. General inspection for QQ plot also shows convexity. Although the remaining subset for GPD was very small, considering the result of Pearson and Johnson distribution systems and general trend in both mean excess function and QQ plot, it is safe to assume that this data set is bounded.

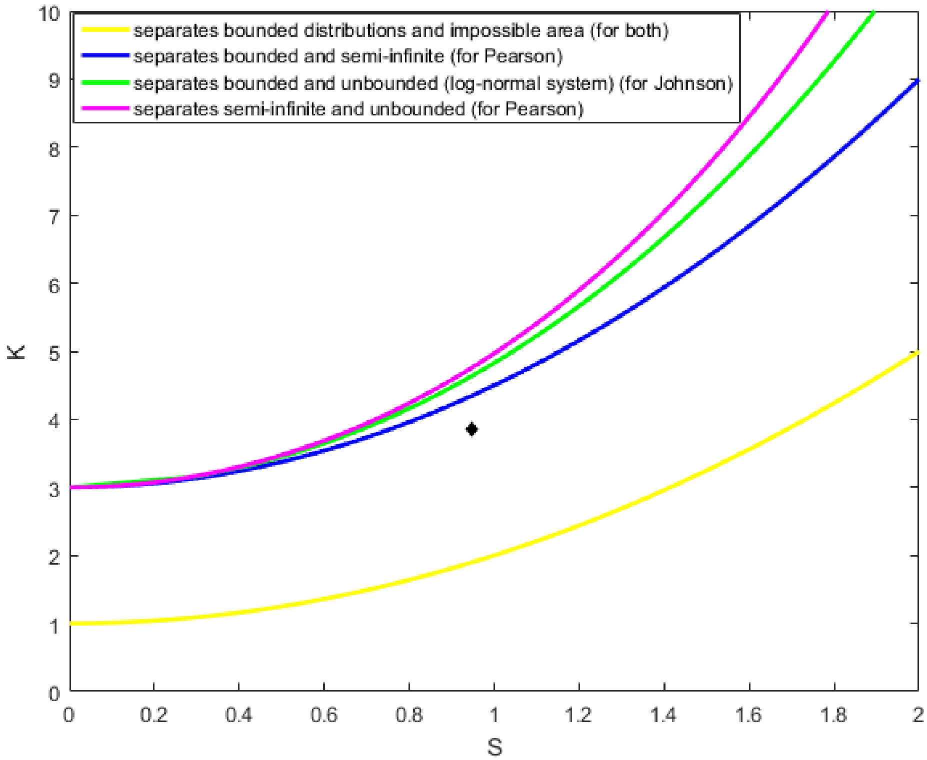

Figure 4.6. Pearson and Johnson distribution systems result of Hollister – Fremont (2) data.

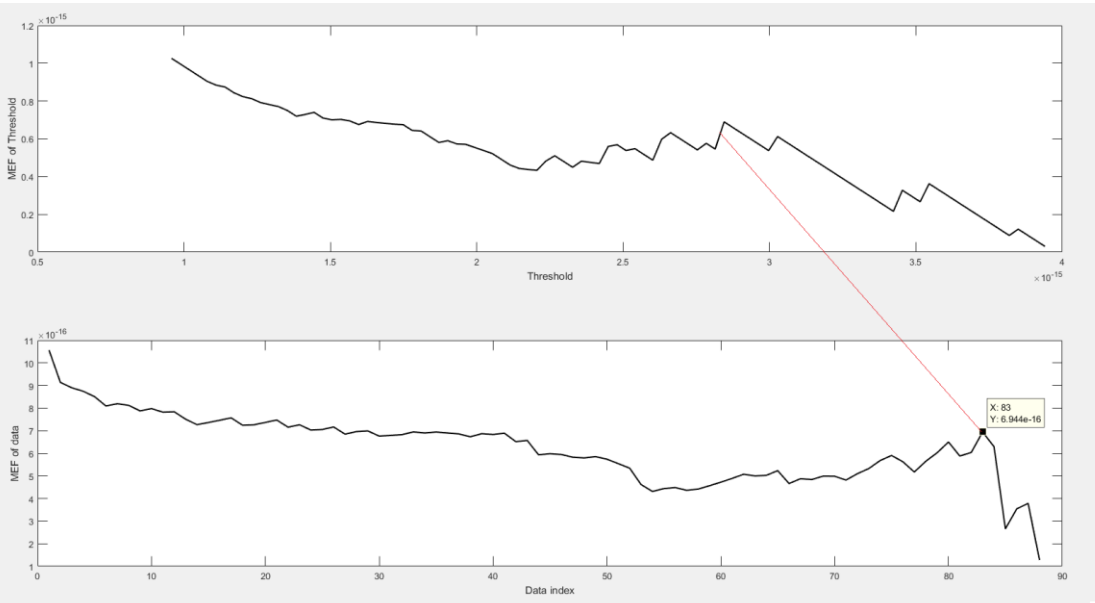

Figure 4.7. Mean excess function of Hollister – Fremont (2) data.

Table 4.3. k value estimation of Hollister – Fremont (2) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
63.65583	8.16438	15.22376	1.125356	0.369295	1.469799	3.933333

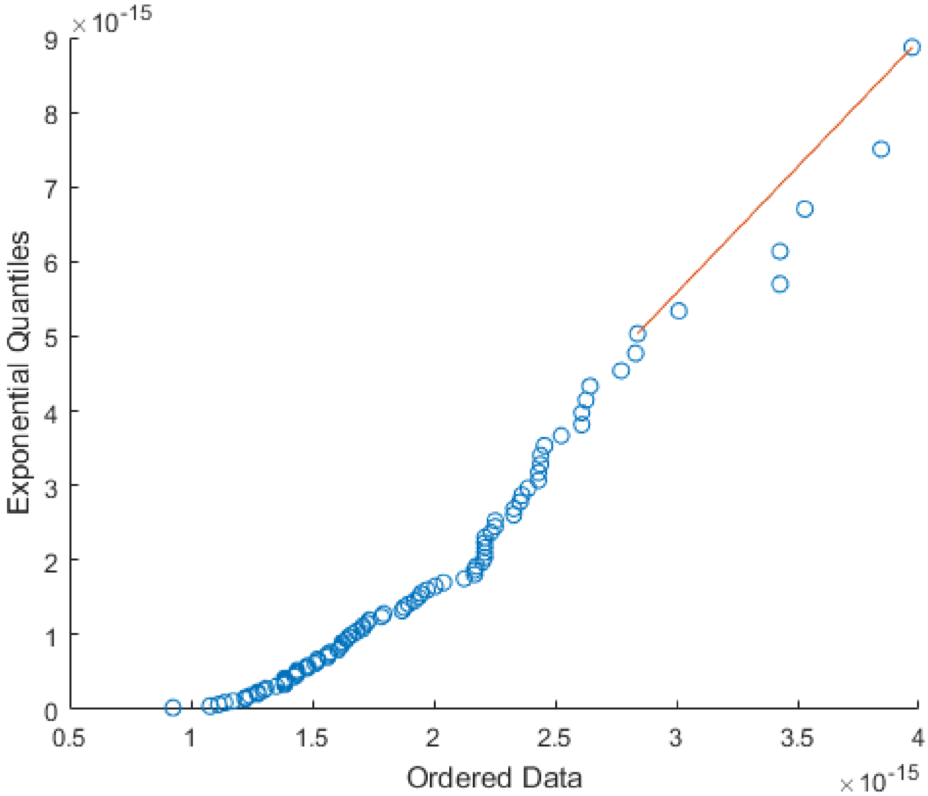

Figure 4.8. QQ plot of Hollister – Fremont (2) data.

Cdf graph inspection in Figure 4.9 again shows MoM is very bad for estimating beta distribution parameters while OLS and DWLS method results are very similar. Error comparison in Figure 4.10 shows except for a part close to the upper tail, OLS and DWLS show better results than lognormal.

All except MoM again pass the K-S test while only OLS and DWLS pass Chi-Square test (Table 4.4). OLS gives the best result for the Chi-Square test and although it is very close to OLS result, DWLS' is the best for the K-S test.

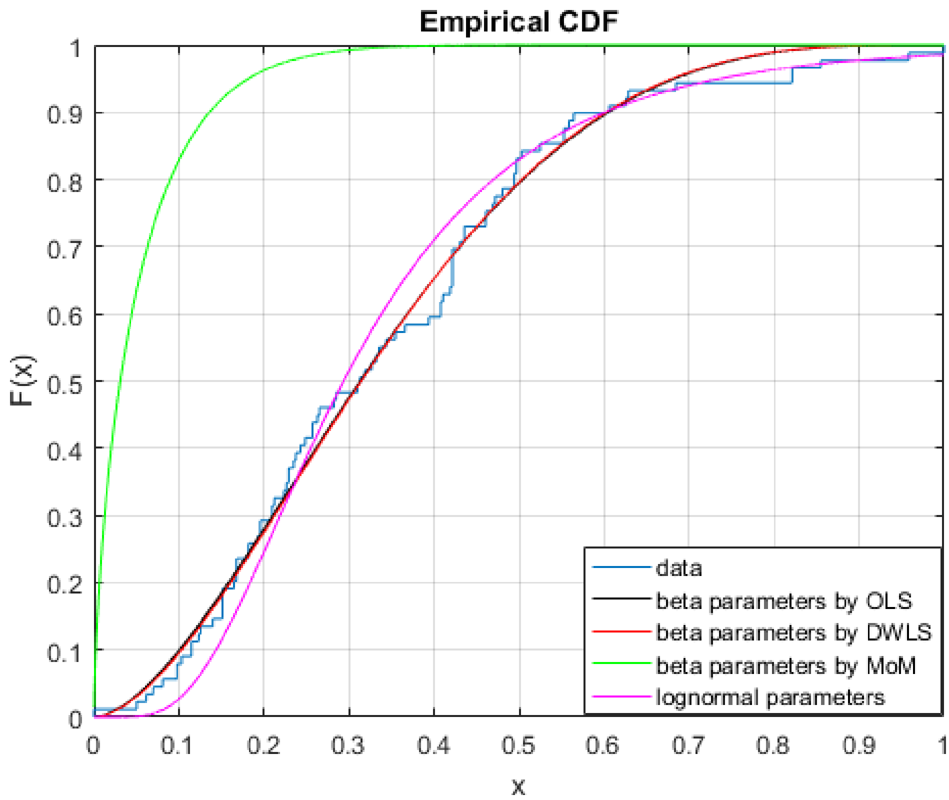

Figure 4.9. Cdf fitting of Hollister – Fremont (2) data.

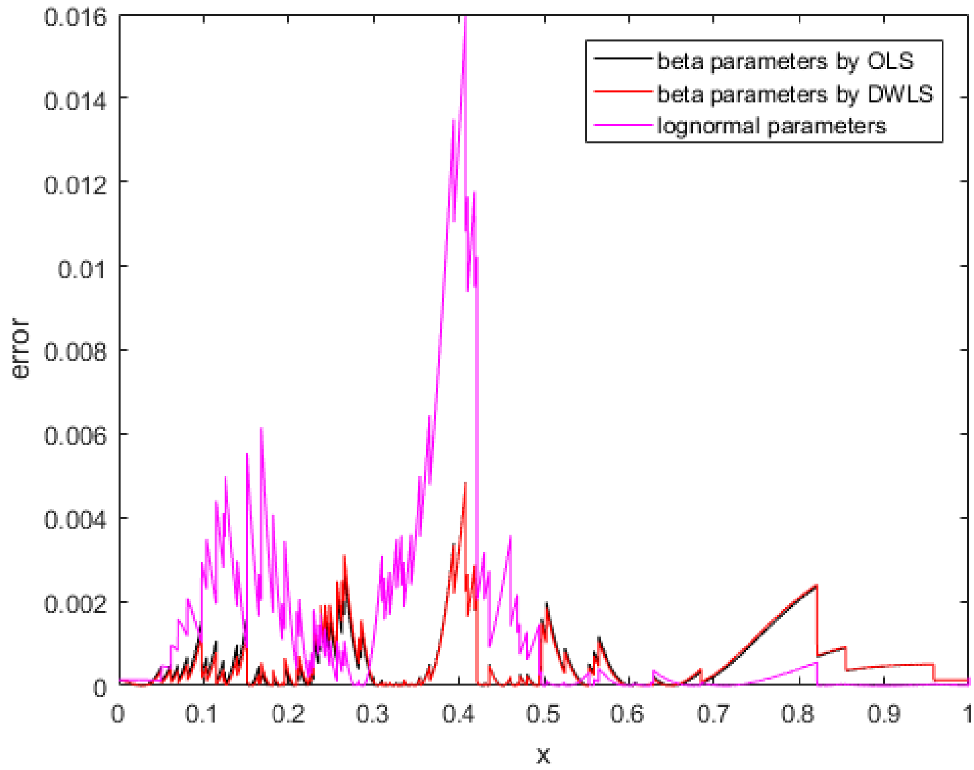

Figure 4.10. Error comparison for Hollister – Fremont (2) data.

Table 4.4. Hollister – Fremont (2) goodness of fit test results.

	MATLAB Default Check		Calculation Results	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	0	0	**9.5412**	0.069635
DWLS	0	0	12.2079	**0.069567**
MoM	1	1	467.7730	0.774251
Lognormal	1	0	14.1708	0.126304

4.3 Hollister (1) Data Results

S-K graph in Figure 4.11 shows that the result of Pearson and Johnson distribution systems is well inside the region of the bounded distributions.

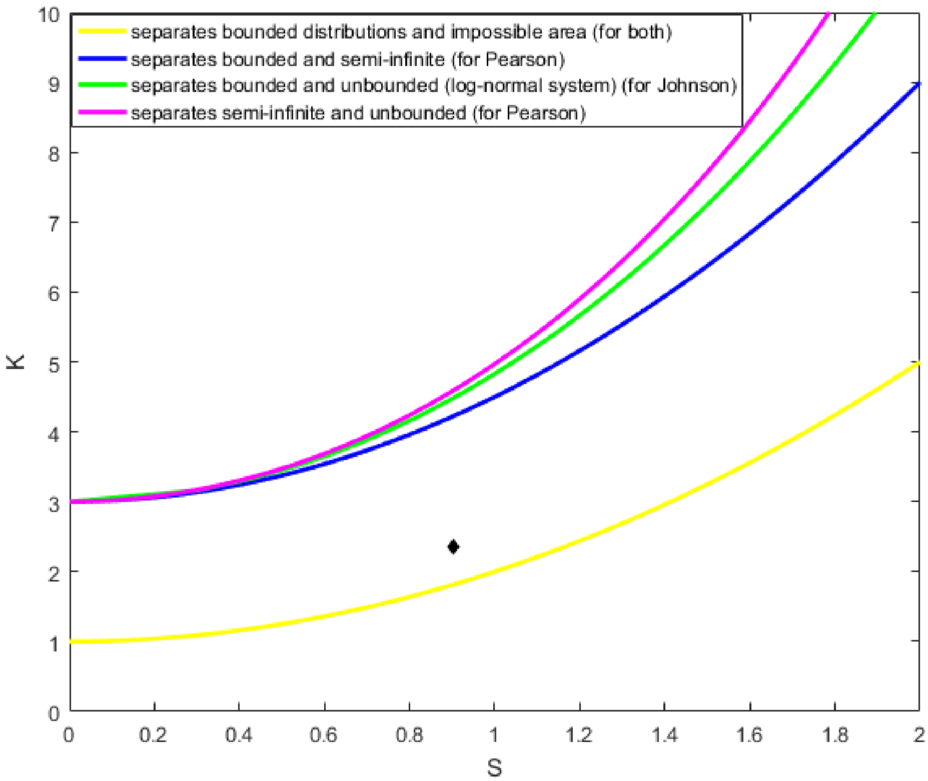

Figure 4.11. Pearson and Johnson distribution systems result of Hollister (1) data.

Mean excess function starts with an upward trend but after the threshold at 21%, it shows a downward trend. There are 38 samples left for GPD and all k value estimation methods give positive results for k, given in Table 4.5. QQ plot in Figure 4.13 shows clear convexity and all inspections point to clear boundedness of this data set.

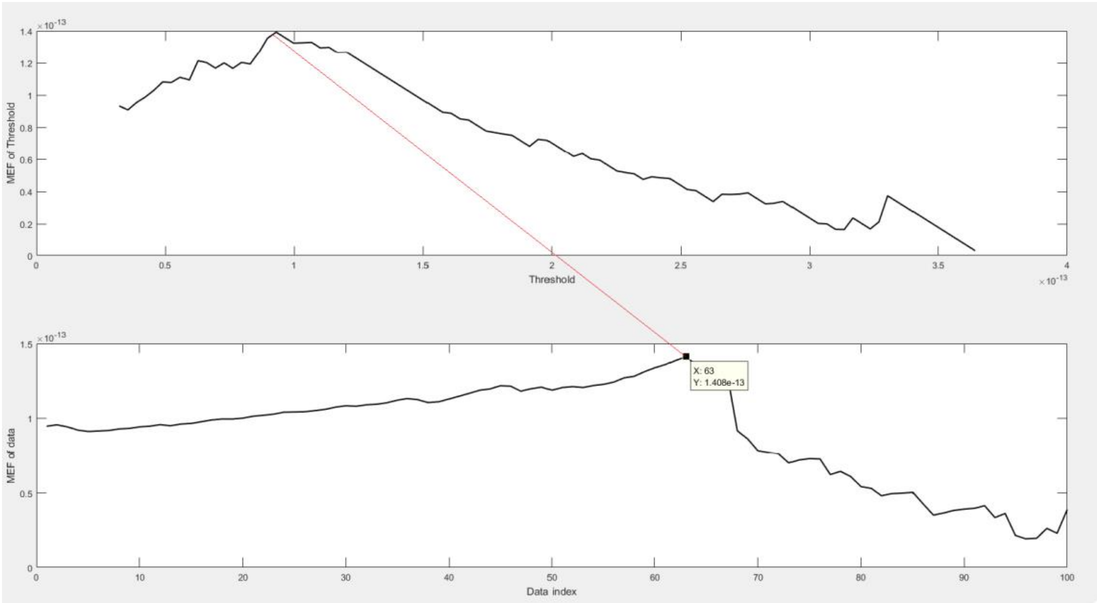

Figure 4.12. Mean excess function of Hollister (1) data.

Table 4.5. k value estimation of Hollister (1) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
5.828068	3.977278	4.108222	1.215129	0.818358	0.653861	0.56801

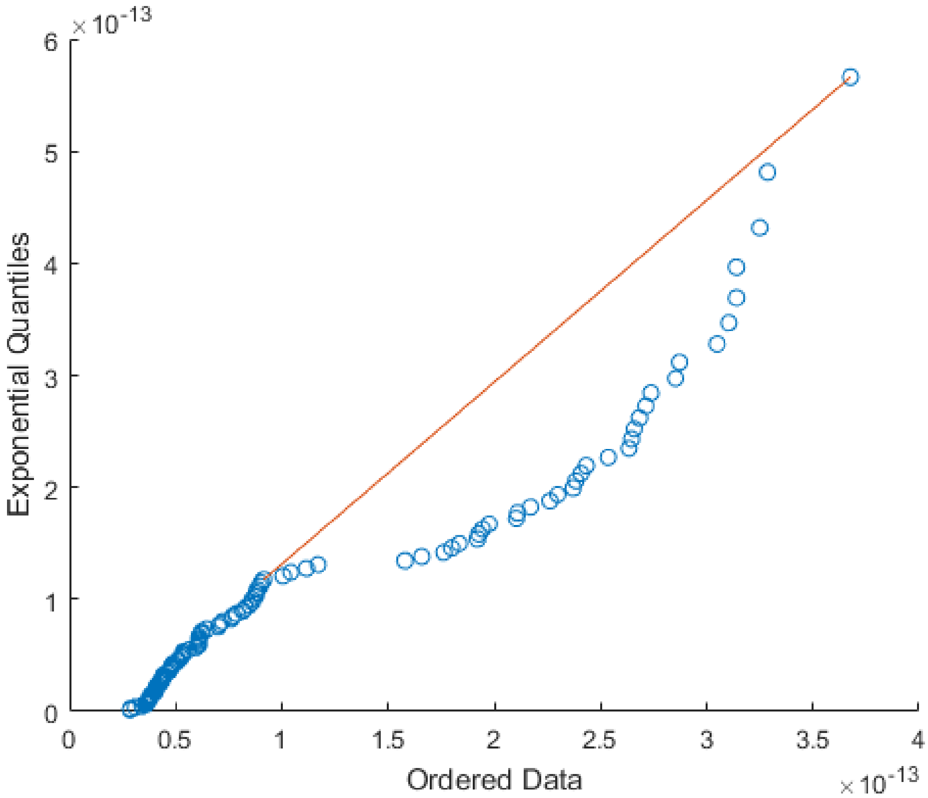

Figure 4.13. QQ plot of Hollister (1) data.

OLS and DWLS cdf graphs in Figure 4.14 differ from each other and error graph in Figure 4.15 shows OLS was a better fit for the middle parts although it is bad at predicting lower values of the set. They both do well compared to lognormal.

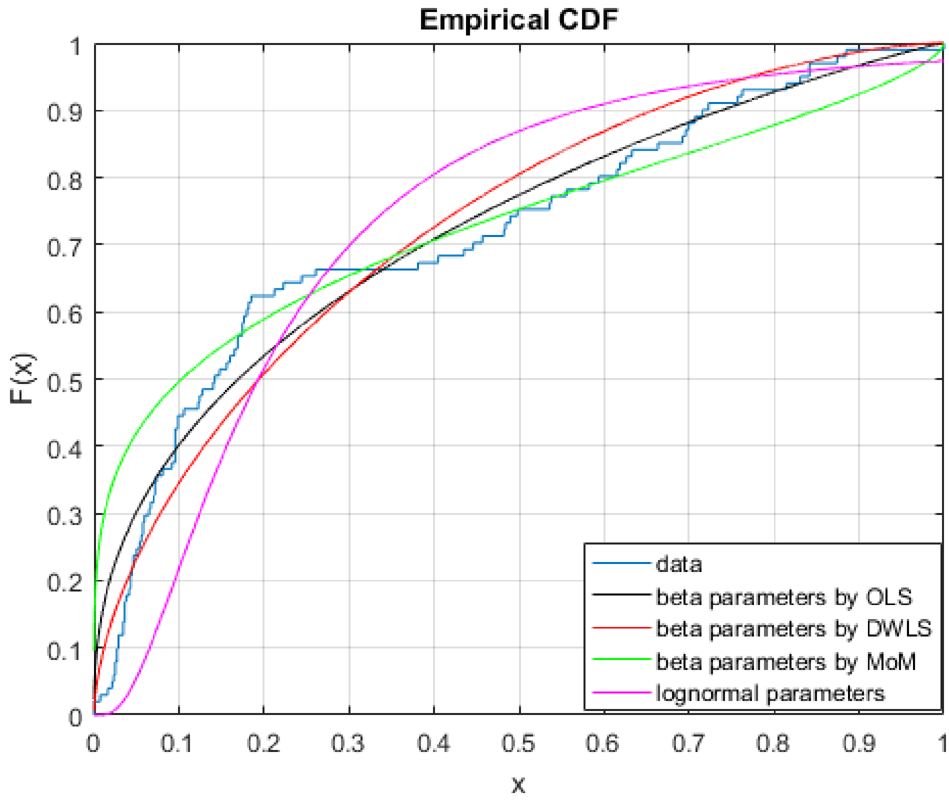

Figure 4.14. Cdf fitting of Hollister (1) data.

There's only 1 sample at the third bin of the histogram which considering the Chi-Square test formula, makes it hard to pass the test. The only pass of the tests comes from MATLAB's Chi-Square test for OLS and the possible reason for that is a different selection of bin numbers. DWLS result is very close to the critical value for the K-S test but it also fails (Table 4.6).

Table 4.6. Hollister (1) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	0	1	87.9048	0.1728
DWLS	1	1	173.2272	**0.1364**
MoM	1	1	**80.8810**	0.3032
Lognormal	1	1	282.3929	0.2314

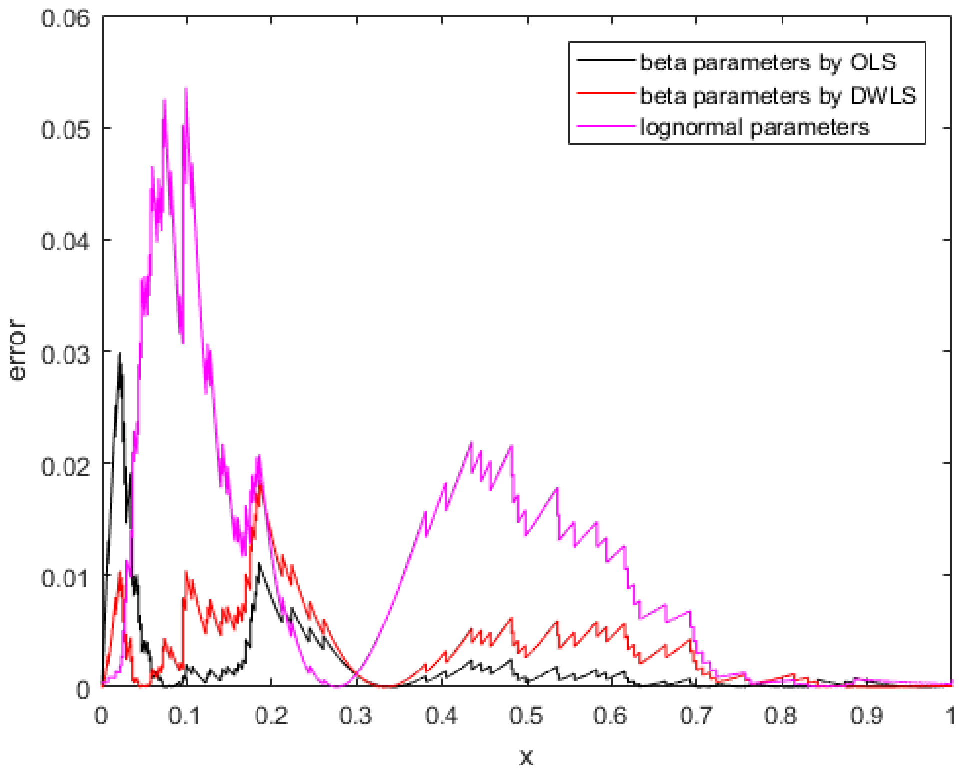

Figure 4.15. Error comparison for Hollister (1) data.

4.4 Hollister (2) Data Results

Figure 4.16 shows that S-K value of the data set is inside the region of the bounded distributions.

Mean excess function shows a downward trend throughout and it is hard to decide the threshold from the conventional method so proposed method with matching thresholds with sample values is used to decide on the threshold (Figure 4.17). The threshold is at 61% and the number of samples left is 53. K value estimation results are given in Table 4.7. QQ plot inspection in Figure 4.18 shows convexity for both in general and for samples above threshold so every inspection shows signs toward boundedness.

Table 4.7. k value estimation of Hollister (2) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
52.81051	14.95851	16.4053	0.296228	0.269277	0.270298	0.272595

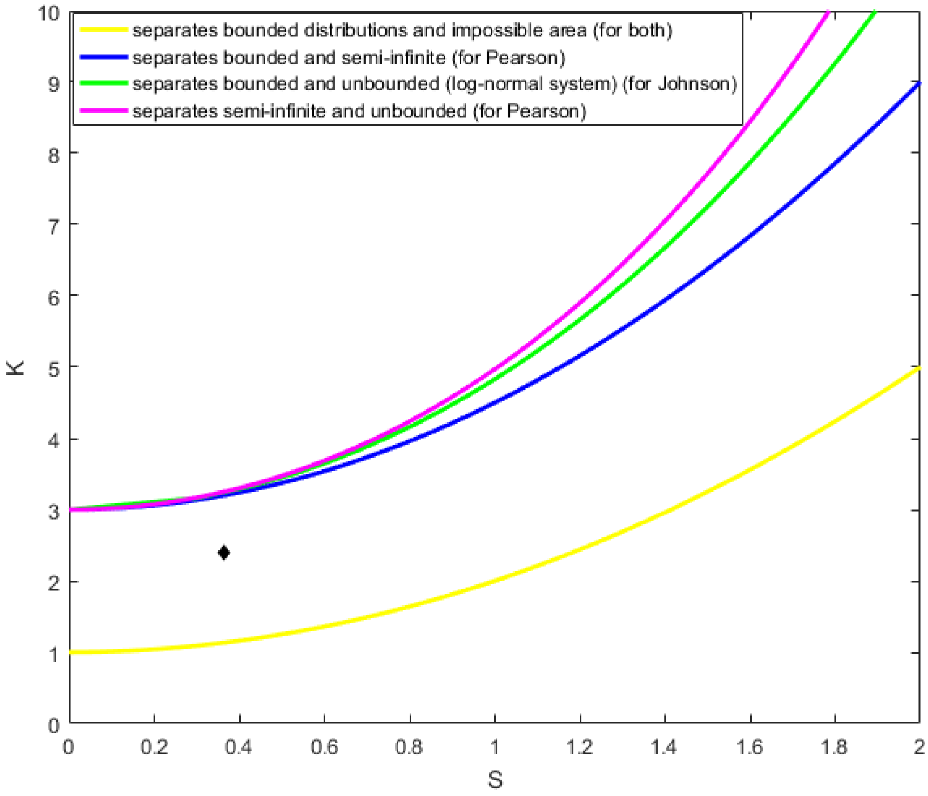

Figure 4.16. Pearson and Johnson distribution systems result of Hollister (2) data.

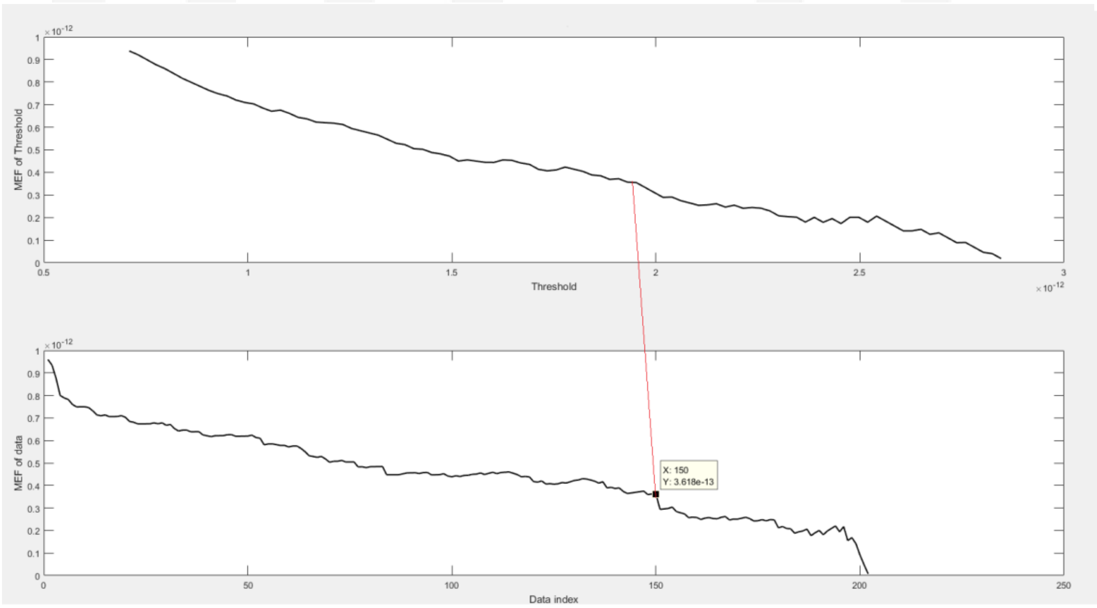

Figure 4.17. Mean excess function of Hollister (2) data.

Figure 4.19 shows that OLS and DWLS results are again almost similar. Figure 4.20 shows they are better than lognormal and MoM estimation is the worst fit.

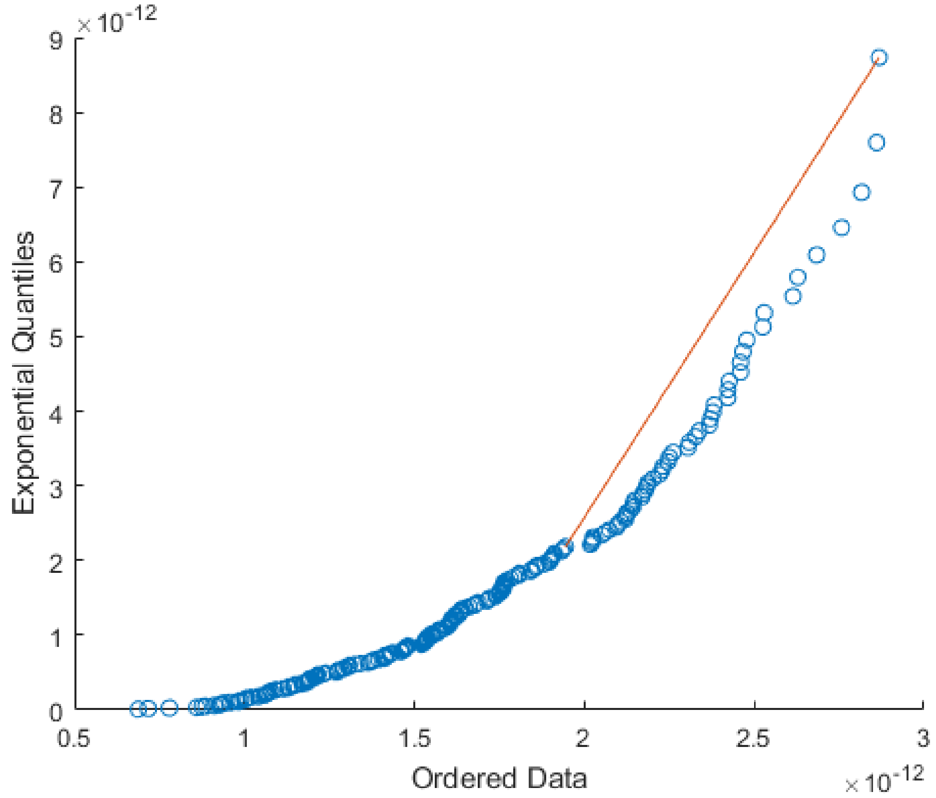

Figure 4.18. QQ plot of Hollister (2) data.

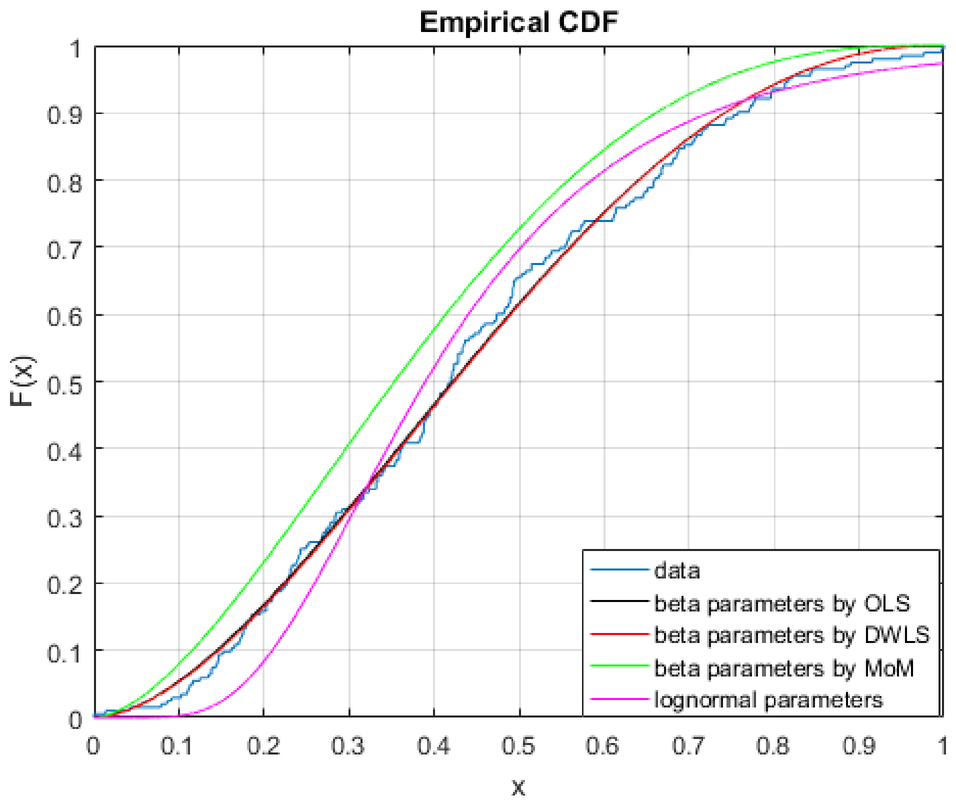

Figure 4.19. Cdf fitting of Hollister (2) data.

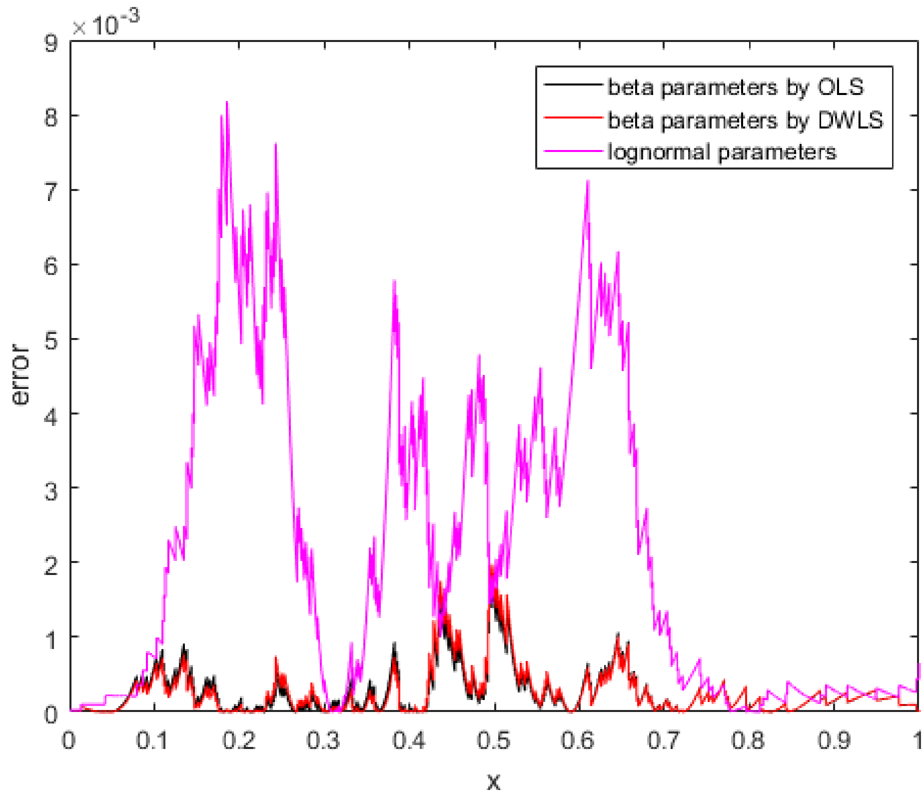

Figure 4.20. Error comparison for Hollister (2) data.

OLS and DWLS pass both Chi-Square and K-S tests while lognormal passes only K-S and MoM fails at both. As they are very close to each other test results of OLS and DWLS are very close too with DWLS is the best result for the Chi-Square test and OLS is for K-S test (Table 4.8).

Table 4.8. Hollister (2) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	0	0	13.1491	**0.0421**
DWLS	0	0	**12.8843**	0.0445
MoM	1	1	33.2673	0.1404
Lognormal	1	0	33.3348	0.0905

4.5 Hollister (3) Data Results

Although it is closer to the boundary K-S value, the data set is inside the region of the bounded distributions (Figure 4.21).

Although its slope varies, the mean excess function shows a downward trend (Figure 4.22). Threshold chosen is 52% with 28 samples left for GPD. Although L2- and L3-

moment k value estimation is close to zero they are all positive for every method applied (Table 4.9). Two samples are very close to the line connecting the threshold to maximum valued sampled in QQ plot in Figure 4.23, but every sample including those two are below the line which means it is convex. The general trend is also convex. Every result points toward boundedness again.

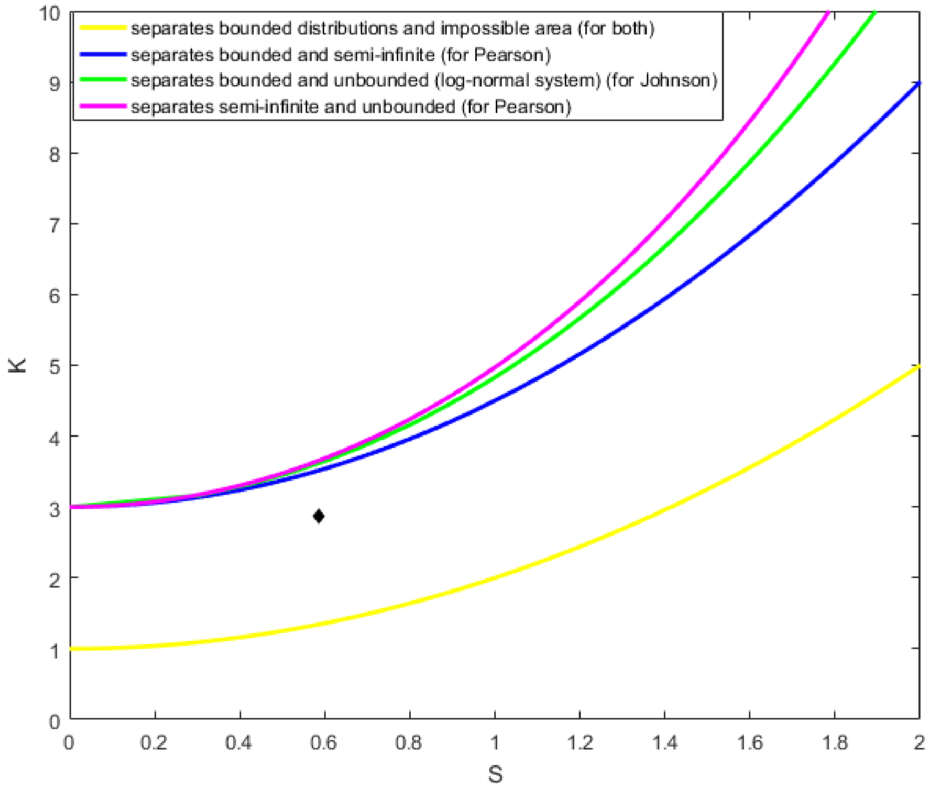

Figure 4.21. Pearson and Johnson distribution systems result of Hollister (3) data.

Table 4.9. k value estimation of Hollister (3) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
24.20308	9.470614	10.61081	0.191113	0.106704	0.076441	0.079364

Cdf graphs in Figure 4.24 show that OLS and DWLS again are very similar and very close fit of the data. Figure 4.25 also shows how well they do against lognormal.

OLS and DWLS pass both goodness of fit tests and MoM and lognormal fail at them both (Table 4.10). Chi-Square test result is the same for OLS and DWLS and for the K-S test, DWLS does just a little bit better.

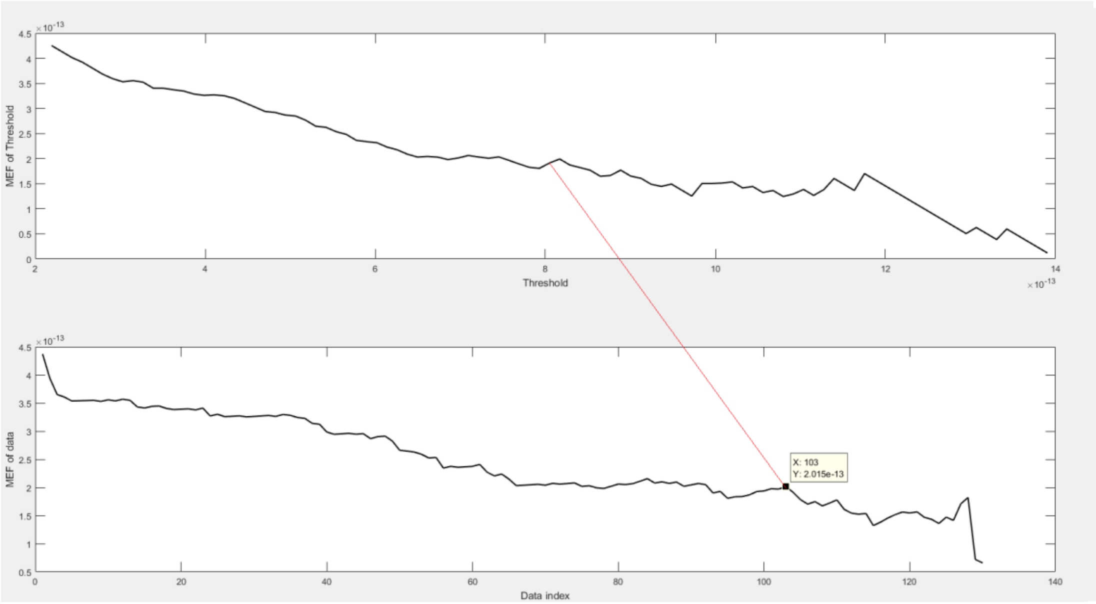

Figure 4.22. Mean excess function of Hollister (3) data.

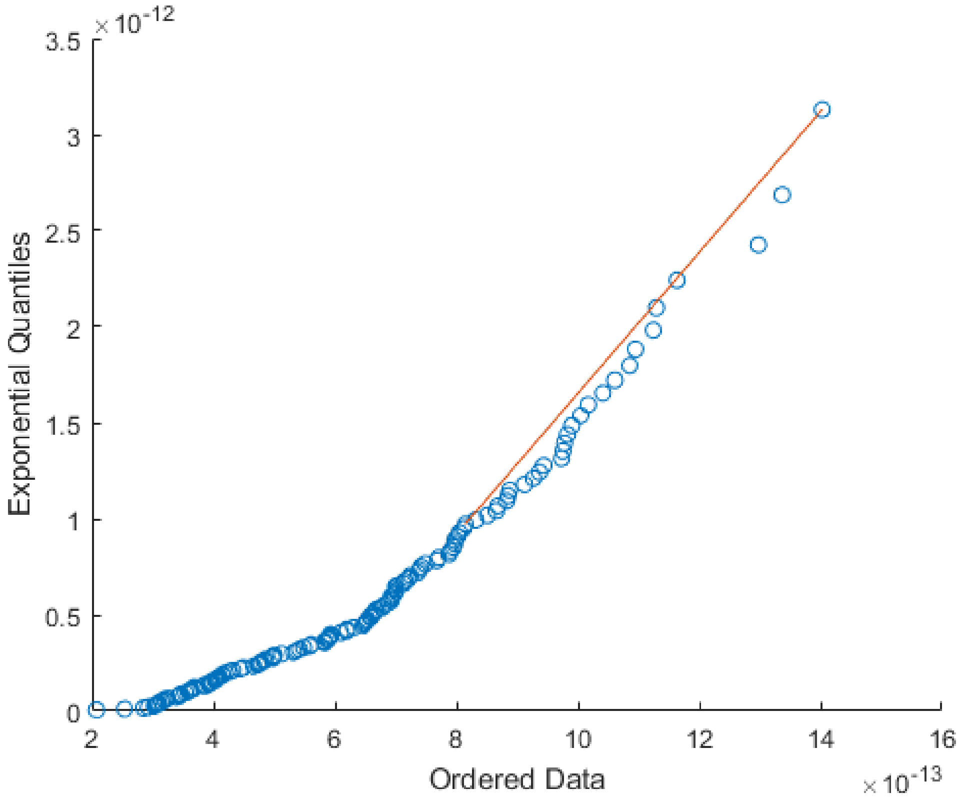

Figure 4.23. QQ plot of Hollister (3) data.

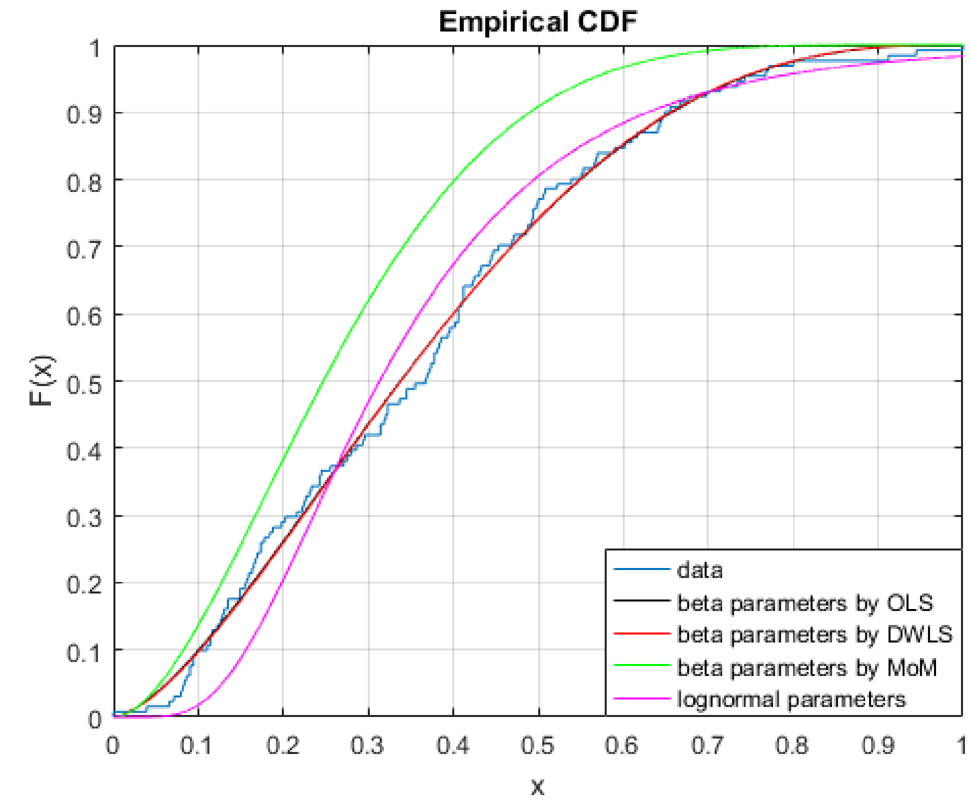

Figure 4.24. Cdf fitting of Hollister (3) data.

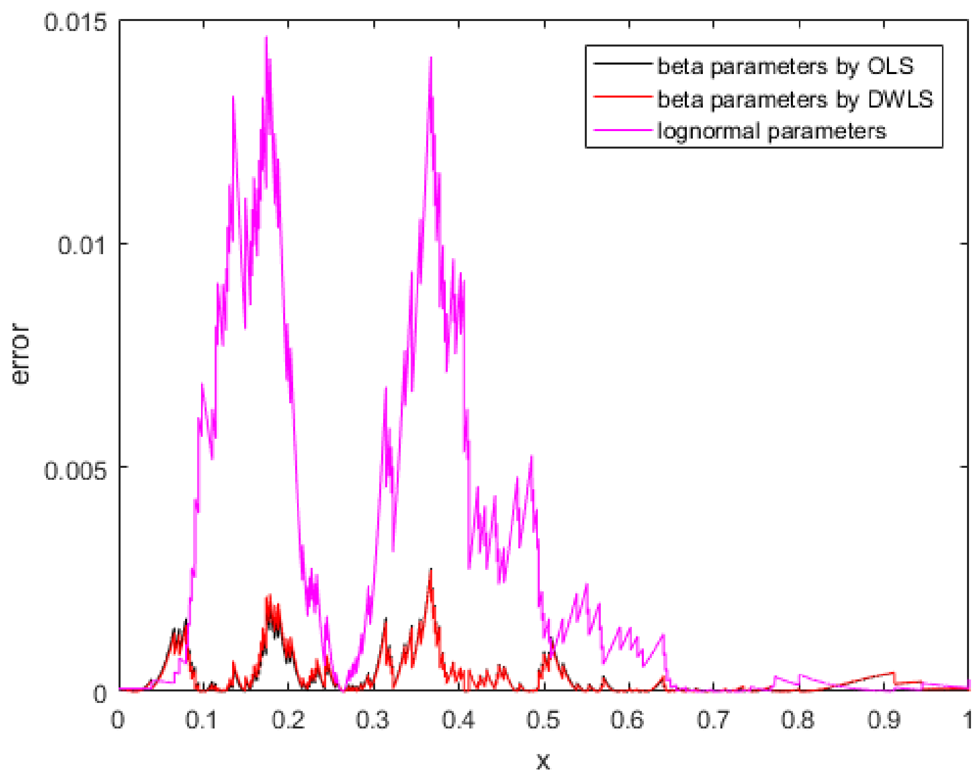

Figure 4.25. Error comparison for Hollister (3) data.

Table 4.10. Hollister (3) goodness of fit test results.

	MATLAB Default Check		Calculation Results	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	0	0	**7.4124**	0.0525
DWLS	0	0	**7.4124**	**0.0520**
MoM	1	1	30.2389	02502
Lognormal	1	1	32.1404	0.1209

4.6 Hollister (4) Data Results

This data set is also inside the region of the bounded distributions of S-K graph (Figure 4.26).

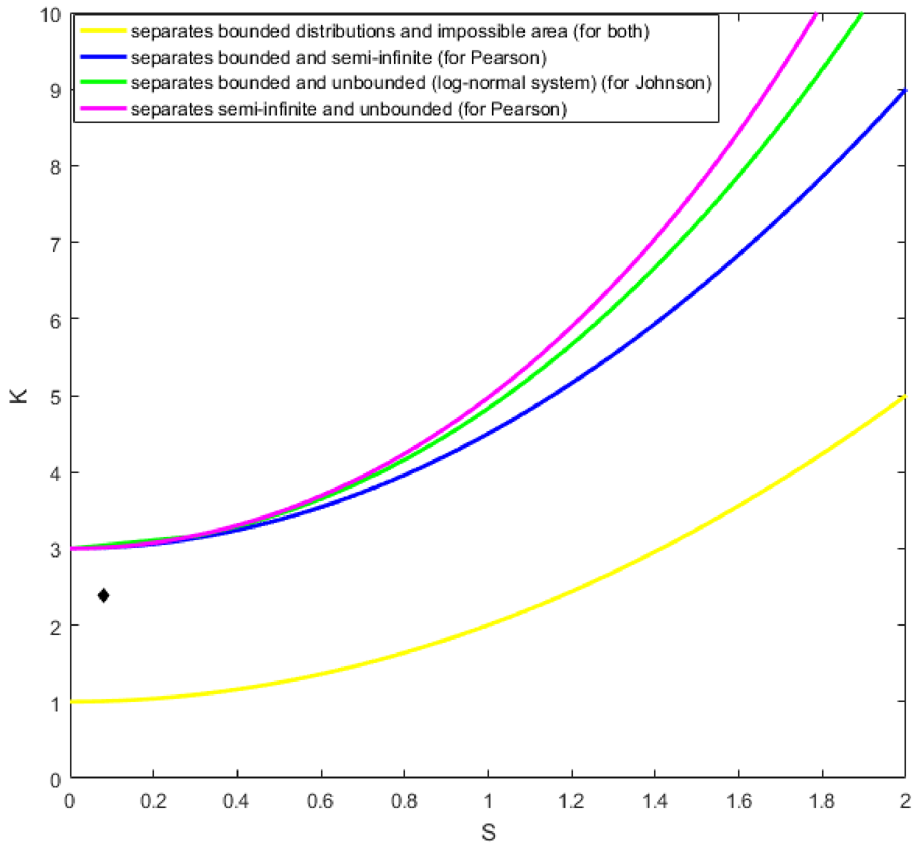

Figure 4.26. Pearson and Johnson distribution systems result of Hollister (4) data.

Mean excess function in Figure 4.27 shows a downward trend with an almost constant slope which makes it hard to select a proper threshold but at 55% there was just a hint of a break in the trend and that point is selected as the threshold. That leaves 60 samples for the GPD and resulting k value estimations are all above zero (Table 4.11). QQ plot in Figure 4.28 shows the clearest convex behaviour of all data sets.

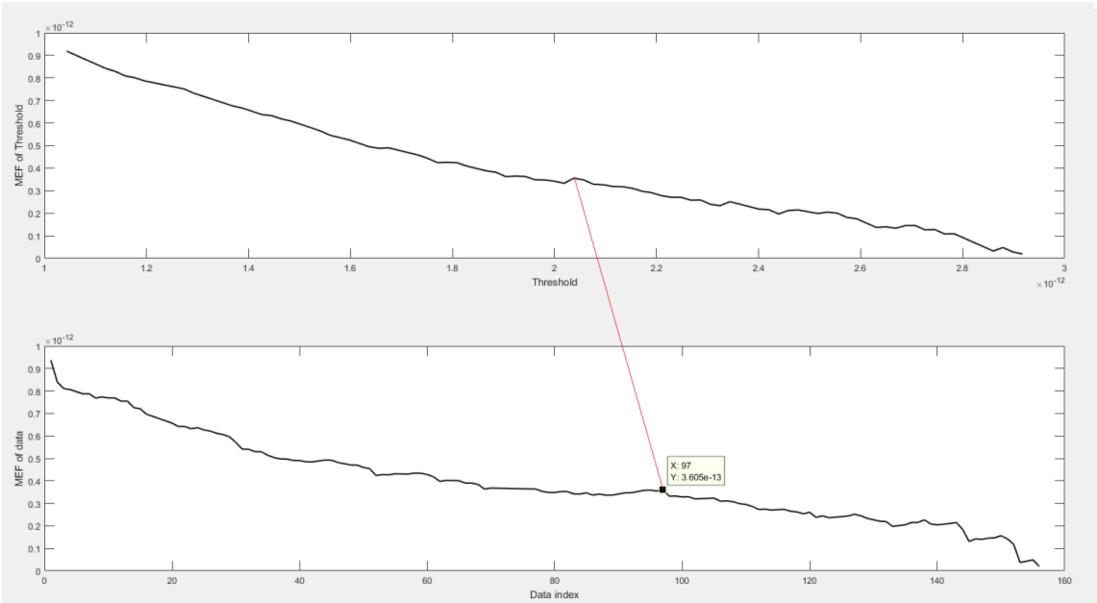

Figure 4.27. Mean excess function of Hollister (4) data.

Table 4.11. k value estimation of Hollister (4) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
52.35026	14.73139	15.95456	0.470229	0.427235	0.417553	0.433759

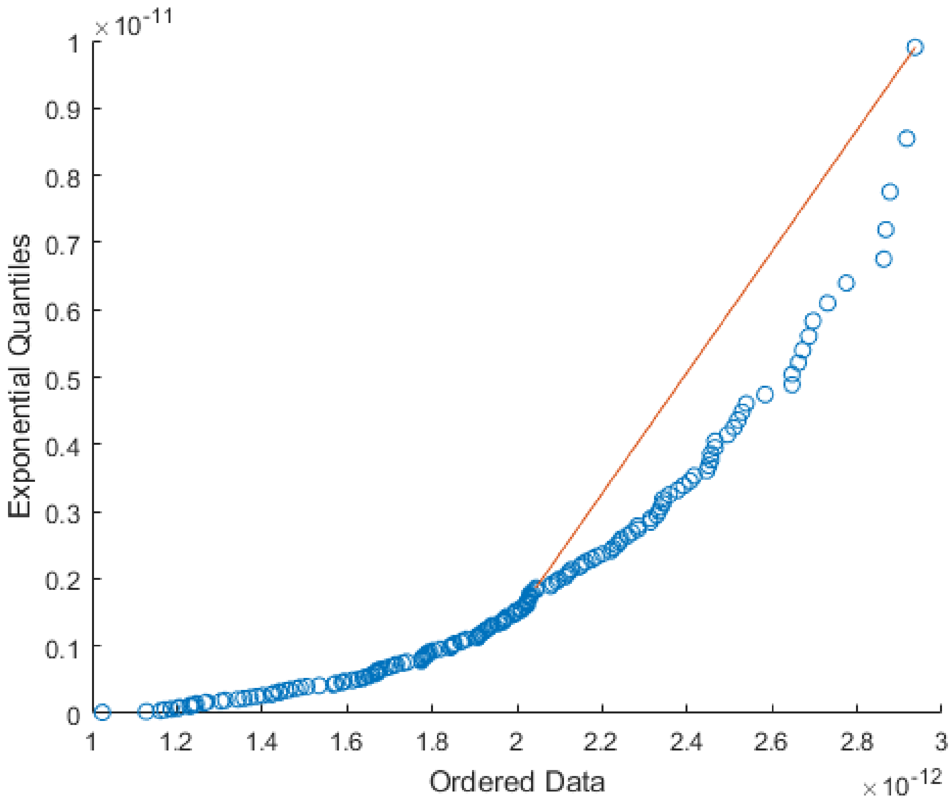

Figure 4.28. QQ plot of Hollister (4) data.

Cdf plot results are very similar to Hollister (3) data as OLS and DWLS are very close to each other and they are better fits than lognormal (Figure 4.29). Error comparison in Figure 4.30 also shows this result.

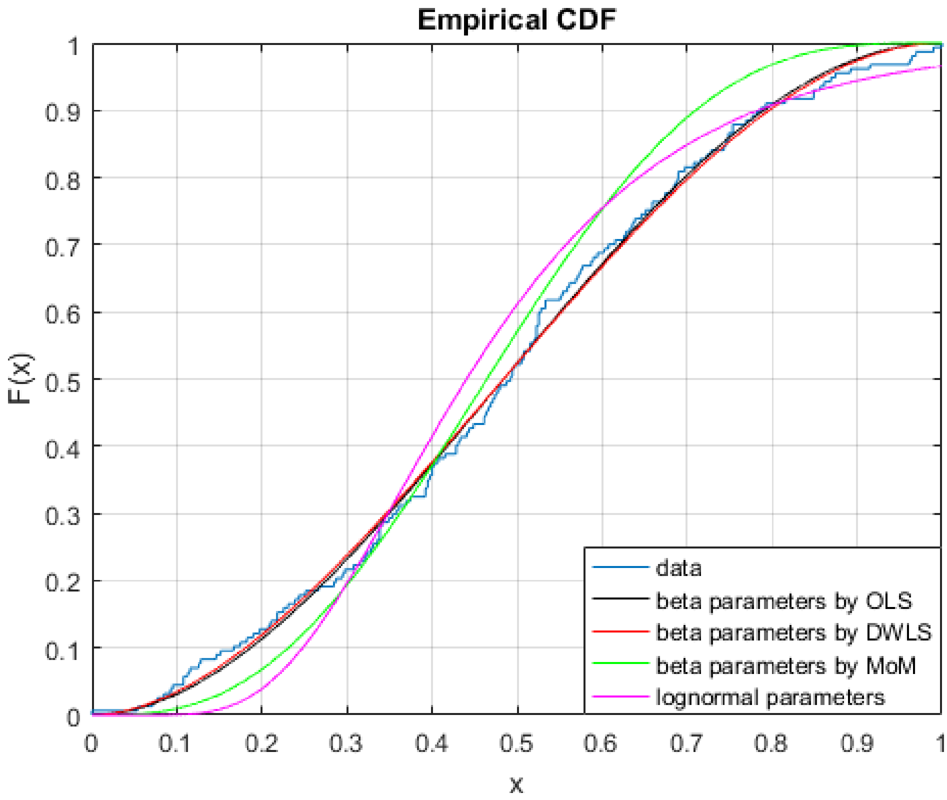

Figure 4.29. Cdf fitting of Hollister (4) data.

OLS and DWLS pass the Chi-Square test while MoM and lognormal fails it. By hand calculation, all of them pass the K-S test but lognormal fails MATLAB's K-S test (Table 4.12). The calculation result is very close to the critical value and there is a possible data loss with the sampling of the theoretical cdf that actually would have resulted in a fail.

Table 4.12. Hollister (4) goodness of fit test results.

	MATLAB Default Check		Calculation Results	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	0	0	2.5765	**0.0437**
DWLS	0	0	**1.9340**	0.0454
MoM	1	0	19.1693	0.0955
Lognormal	1	0	32.1379	0.1079

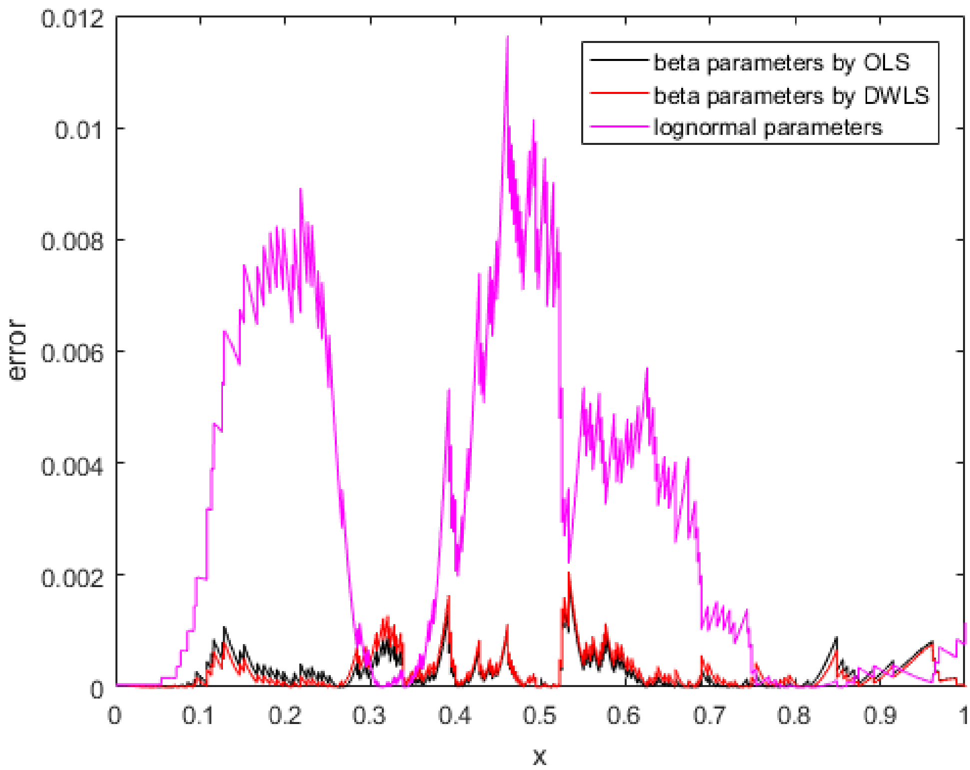

Figure 4.30. Error comparison for Hollister (4) data.

4.7 TUBITAK (1) Data Results

This data set's S-K value is well inside the region of the bounded distributions shown in Figure 4.31.

Mean excess function starts with a very slight upward trend but after the chosen threshold 21% it shows a downward trend (Figure 4.32). There are 1057 samples remaining for GPD shape parameter calculation and all of the results are positive (Table 4.13). QQ plot of the data in Figure 4.33 set shows clear convexity after the chosen threshold.

Table 4.13. k value estimation of TUBITAK (1) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
5.356434	4.069596	4.073985	0.815694	0.498649	0.359043	0.298877

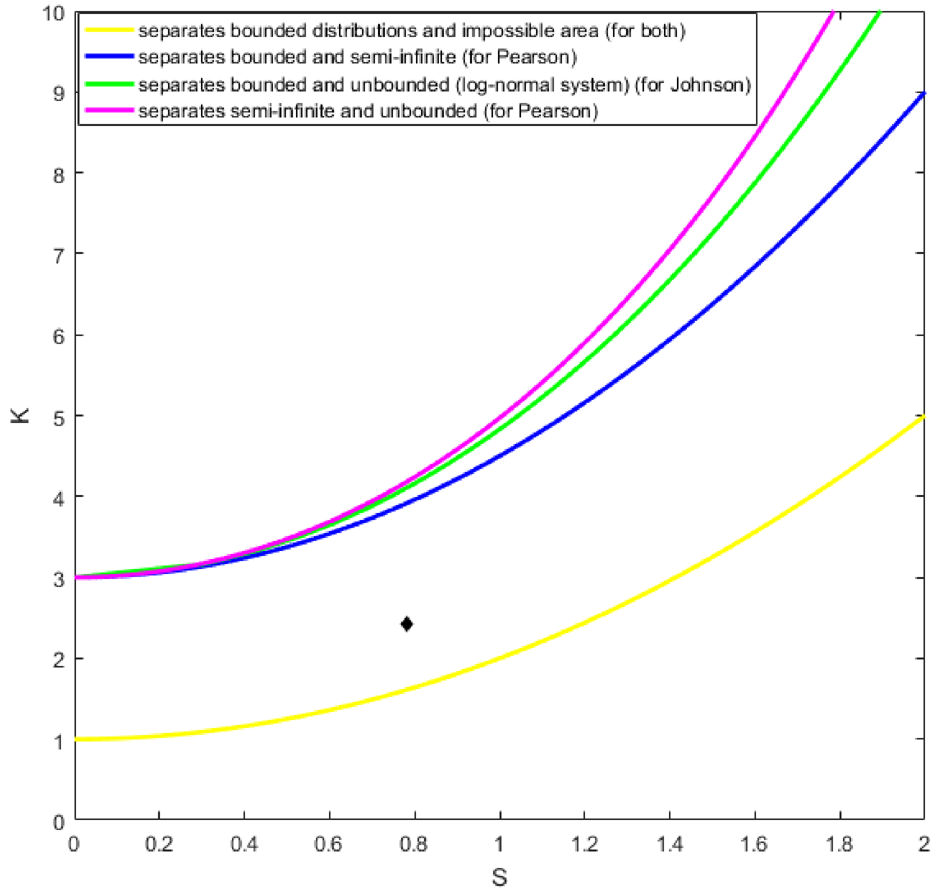

Figure 4.31. Pearson and Johnson distribution systems result of TUBITAK (1) data.

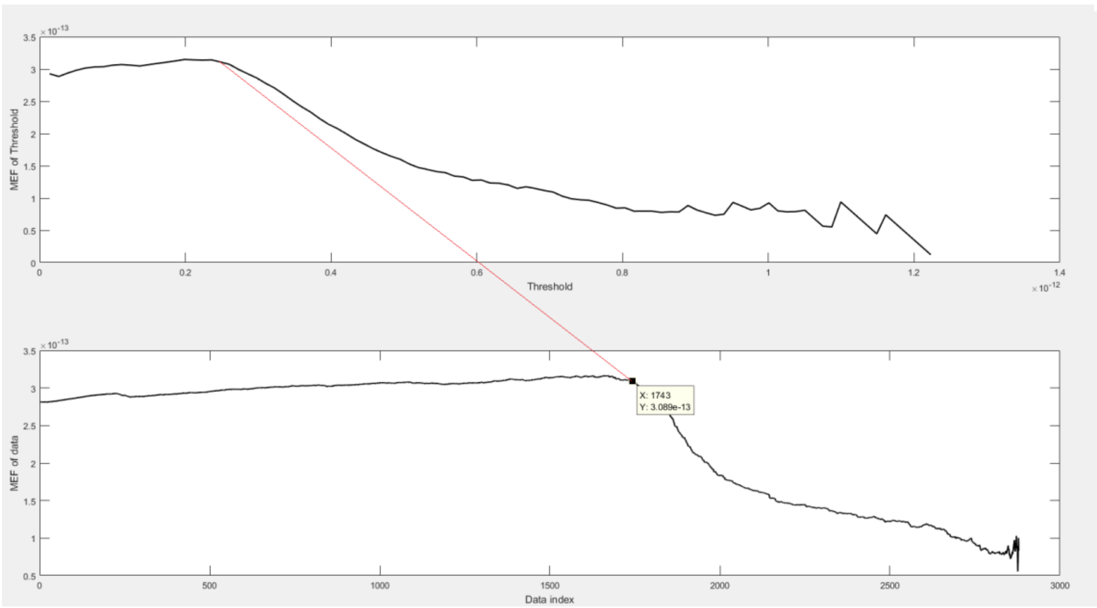

Figure 4.32. Mean excess function of TUBITAK (1) data.

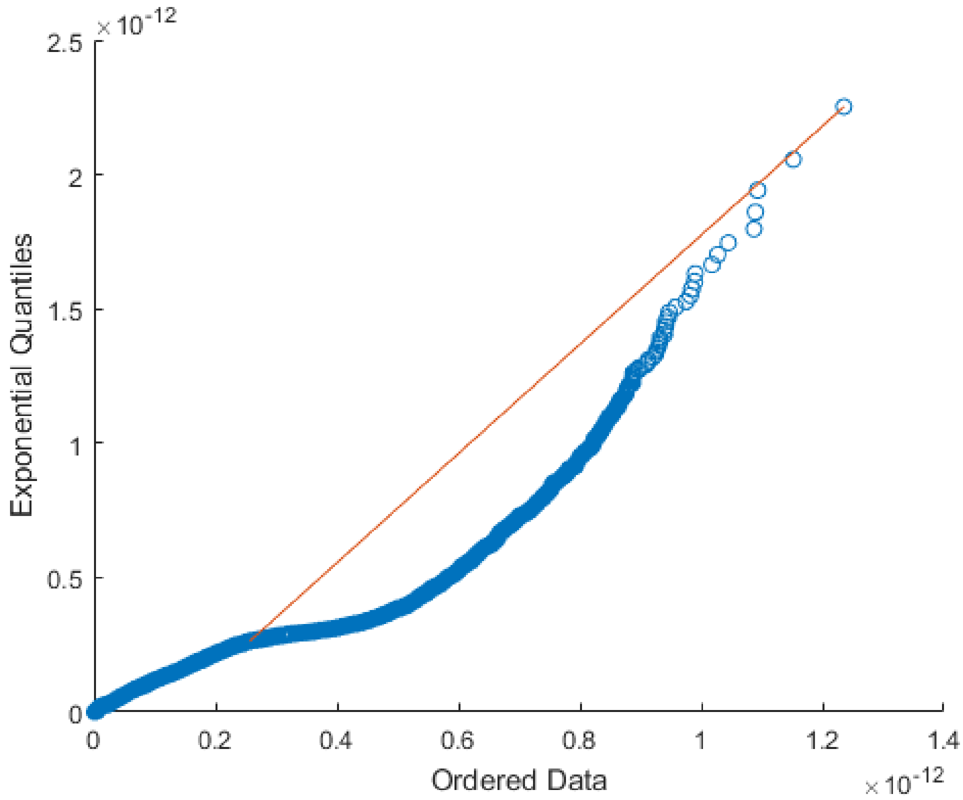

Figure 4.33. QQ plot of TUBITAK (1) data.

Cdf fitting in Figure 4.34 and error comparison in Figure 4.35 show that although they are very close, OLS is better at middle and DWLS is better at the upper tail as excepted and they are better than lognormal.

Although all of the methods fail at both goodness of fit tests, OLS and DWLS results were much better compared to others (Table 4.14).

Table 4.14. TUBITAK (1) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	1	1	681.1634	**0.0561**
DWLS	1	1	**603.7990**	0.0654
MoM	1	1	5810.0076	0.1174
Lognormal	1	1	1767.0188	0.1878

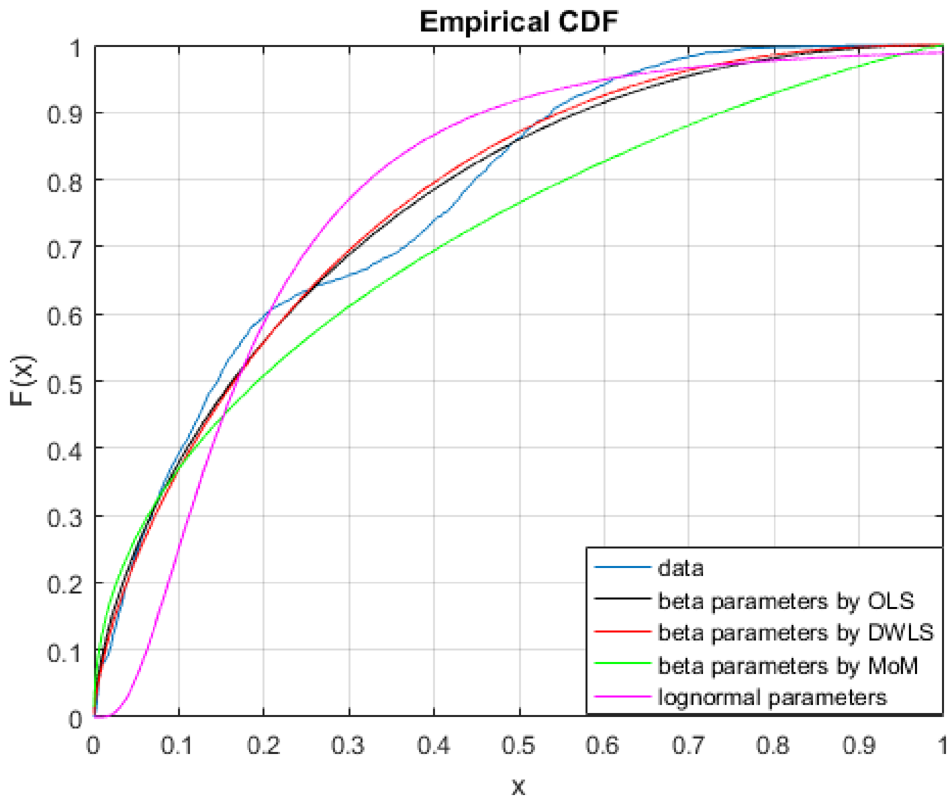

Figure 4.34. Cdf fitting of TUBITAK (1) data.

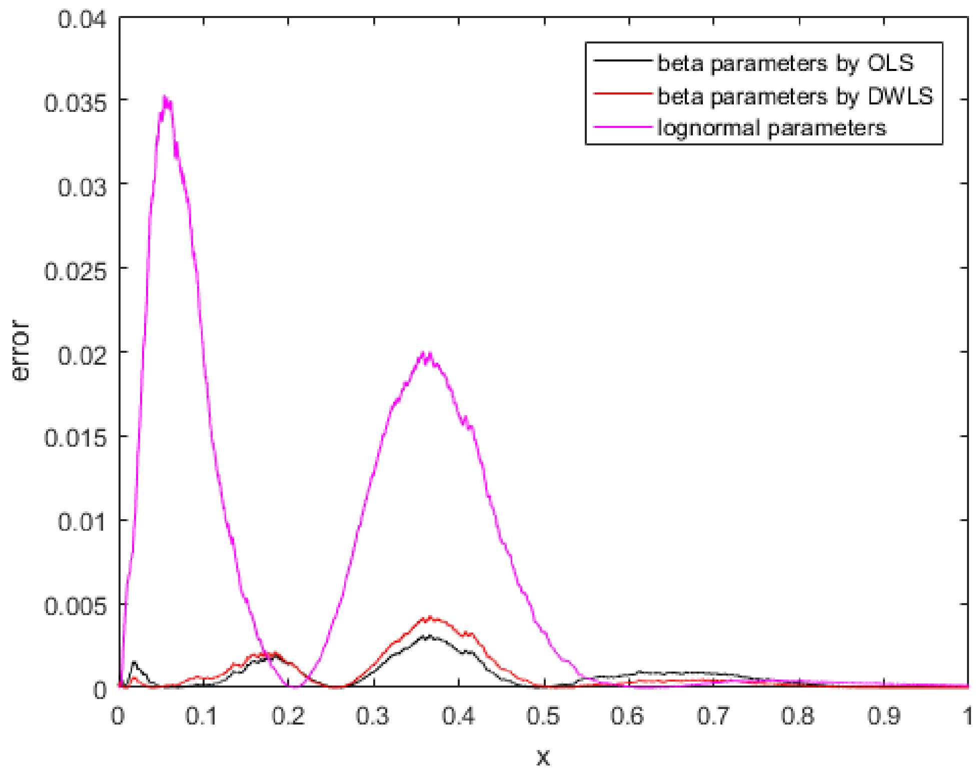

Figure 4.35. Error comparison for TUBITAK (1) data.

4.8 TUBITAK (2) Data Results

This data set is also in the region of the bounded distributions for both Pearson and Johnson distribution systems (Figure 4.36).

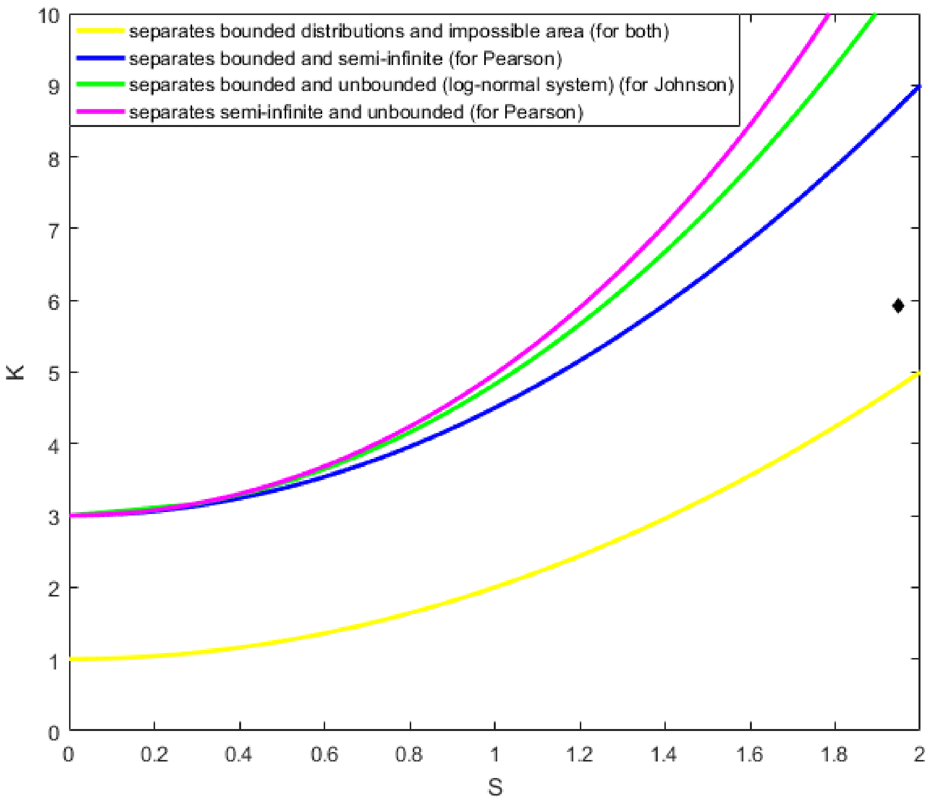

Figure 4.36. Pearson and Johnson distribution systems result of TUBITAK (2) data.

Mean excess function plot in Figure 4.37 starts with a very slanted upward trend but like TUBITAK (1), after a certain threshold, it shows a constant downward trend. This threshold is at 22% and 529 samples remain for GPD. k value estimation results are positive for all methods applied (Table 4.15). QQ plot in Figure 4.38 is clearly convex after the chosen threshold.

Table 4.15. k value estimation of TUBITAK (2) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
4.245607	3.390695	3.397004	0.815708	0.630349	0.509514	0.424344

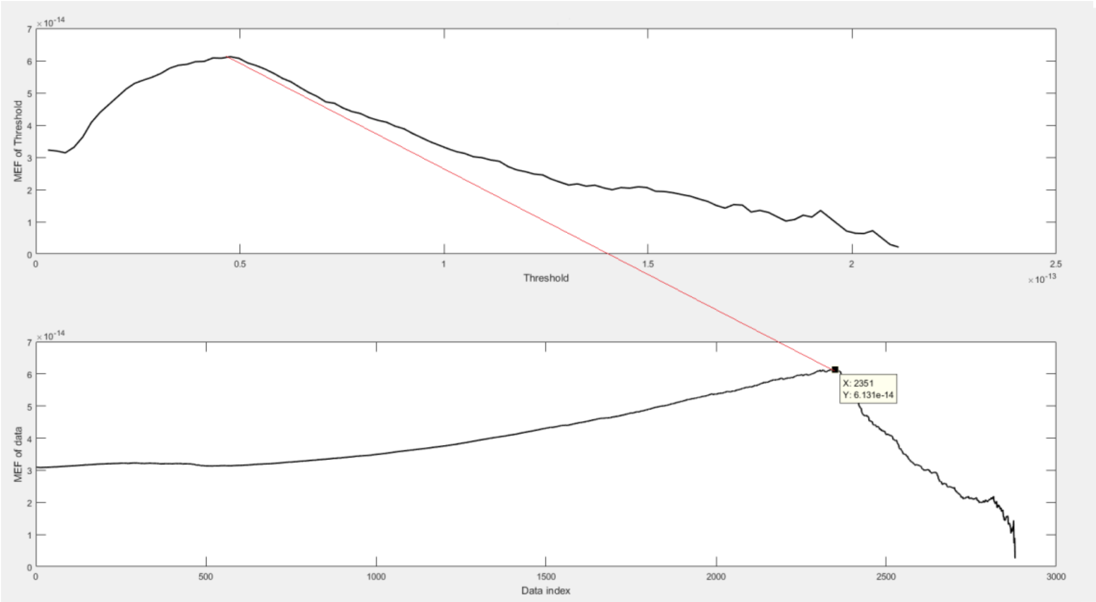

Figure 4.37. Mean excess function of TUBITAK (2) data.

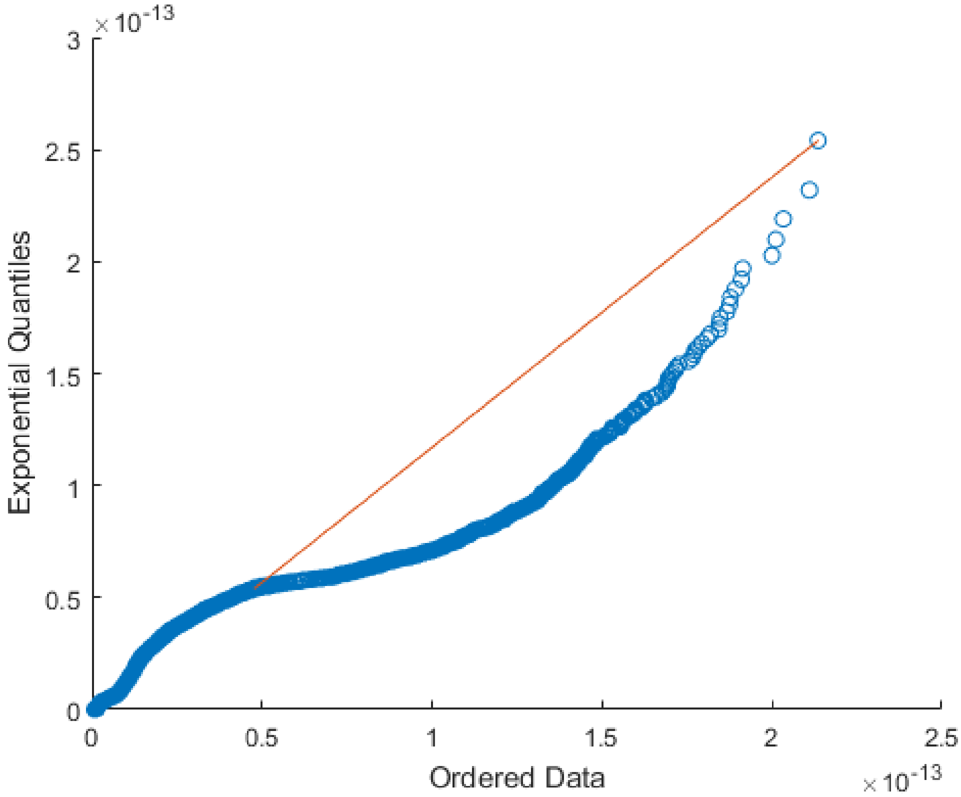

Figure 4.38. QQ plot of TUBITAK (2) data.

This data set is one of the few times OLS and DWLS results are very different (Figure 4.39). MoM again gives very bad results and it can be seen easily that it is a bad fit. OLS error is very high at the start, higher than the maximum of both DWLS and lognormal, but the best fit for the middle part (Figure 4.40).

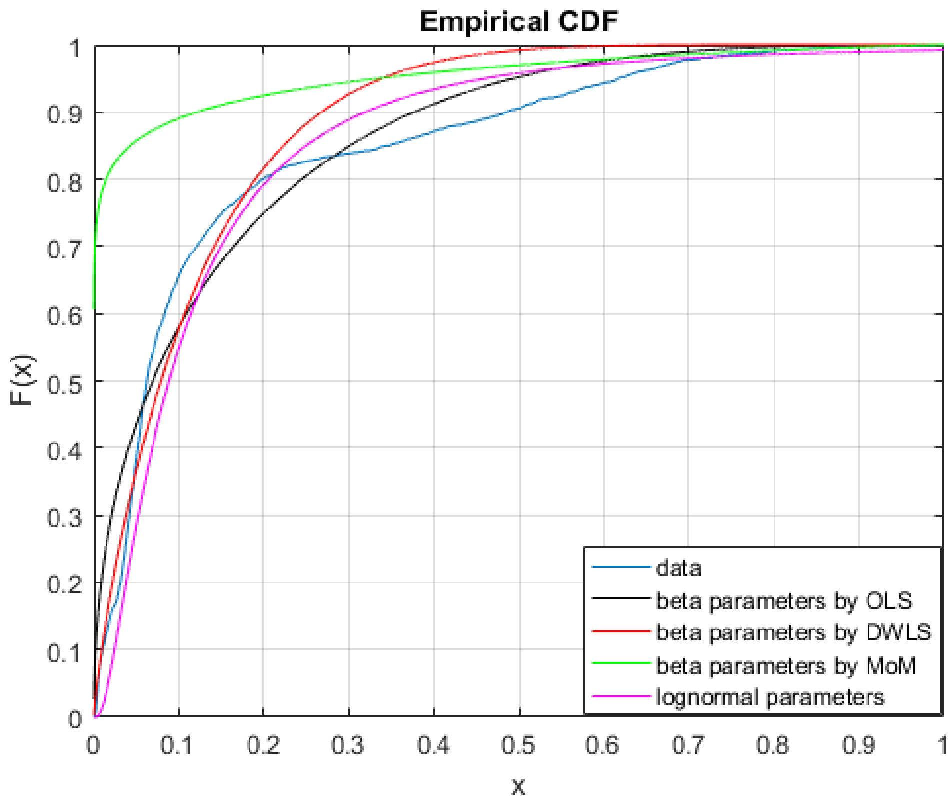

Figure 4.39. Cdf fitting of TUBITAK (2) data.

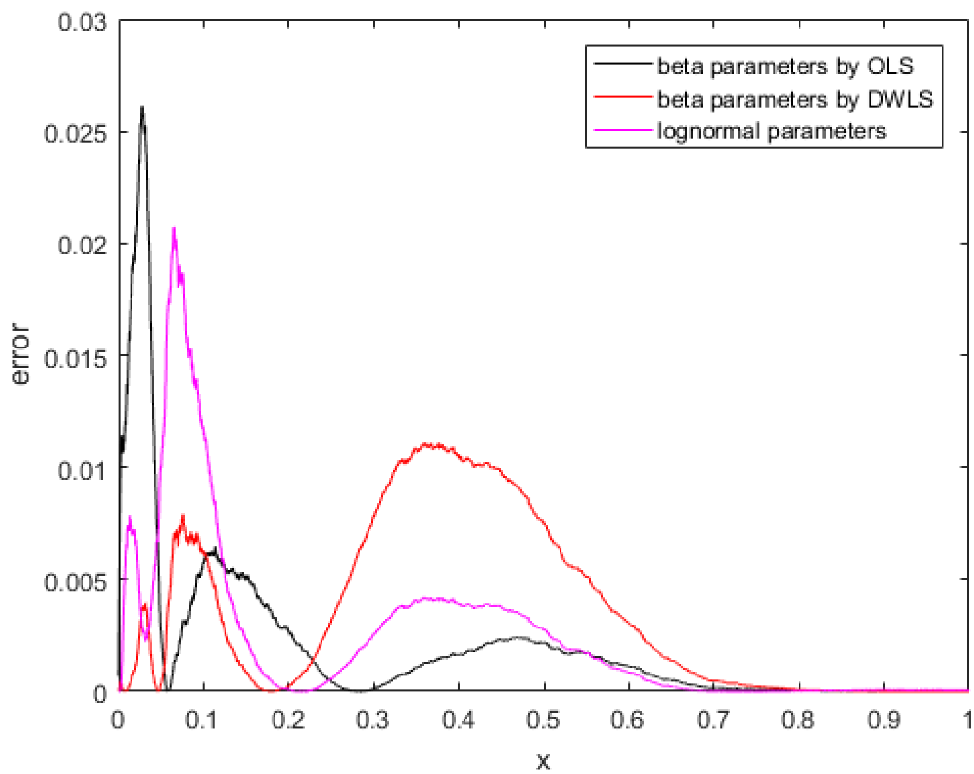

Figure 4.40. Error comparison for TUBITAK (2) data.

While every method fails both tests, OLS gives the best result by far for the Chi-Square test and DWLS gives the best result for the K-S test which (Table 4.16).

Table 4.16. TUBITAK (2) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	1	1	**584.9494**	0.1616
DWLS	1	1	912.1700	**0.1053**
MoM	1	1	1565.6816	0.7143
Lognormal	1	1	753.3747	0.1440

4.9 TUBITAK (3) Data Results

Out of the eleven data sets, this one is the only one that is in the region of the semi-infinite distribution for Pearson distribution system. It is still in the bounded distributions family for Johnson distribution system (Figure 4.41).

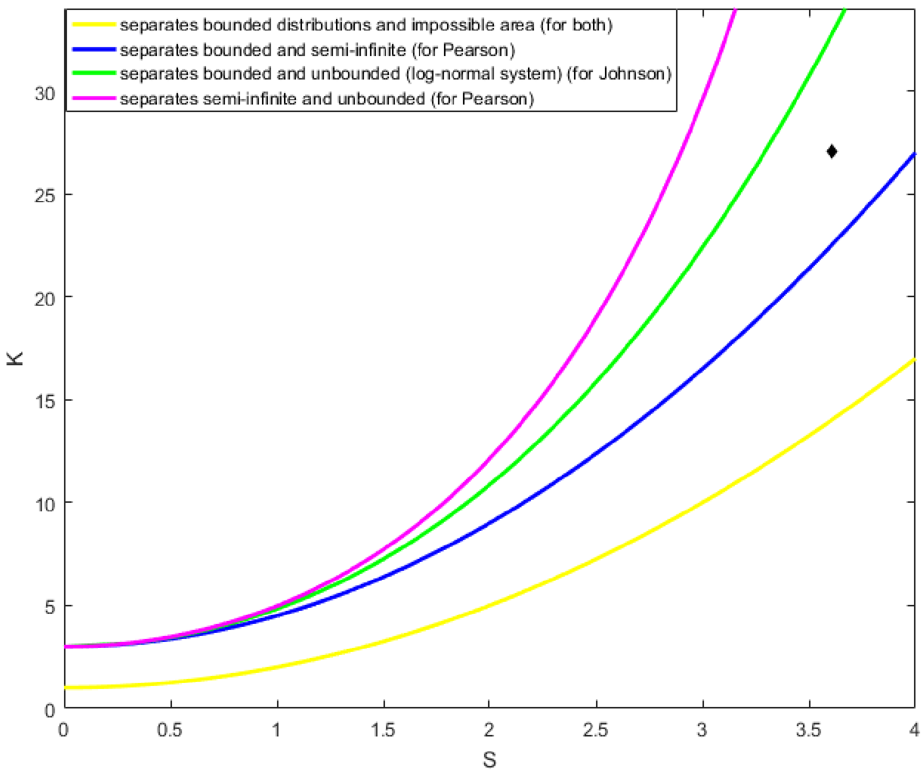

Figure 4.41. Pearson and Johnson distribution systems result of TUBITAK (3) data.

Conventional mean excess function plot starts with an upward trend and around the middle turns to downward trend but when the mean excess function with sample values as thresholds is inspected, it can be seen that most of the data is on the lower end (Figure 4.42). The plot with thresholds as sample values stays at the same level for all the time except for the last part where it increases and decreases sharply. Since it happens only once out of eleven data set it’s not possible to conclude whether a measurement error occurred or not. Examination of more data sets might shed some light on this issue. After the threshold at 63%, only 11 samples are left for GPD which is very low considering the data set consists of 2880 samples. For this threshold, all the estimation methods give positive values for k but results are close to zero for LH-moments method (Table 4.17). QQ plot in Figure 4.43 shows convexity after the threshold but in general, it is concave.

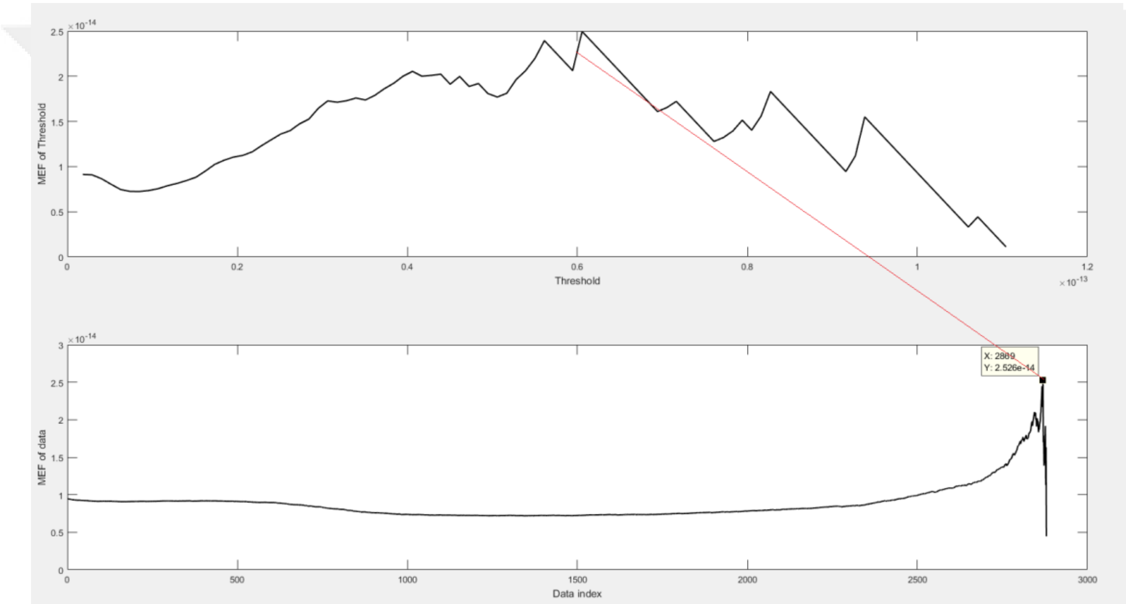

Figure 4.42. Mean excess function of TUBITAK (3) data.

Table 4.17. k value estimation of TUBITAK (3) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
20.4132	6.98914	8.82604	0.157602	0.182819	0.336416	0.562392

As mentioned in the mean excess function plots, cdf graph of the data in Figure 4.44 also shows how stacked this data set is on the lower side as cdf reaches close to 1 even before halfway. Cdf fitting is very hard because of this reason and every method tried is a bad fit for the start (Figure 4.45).

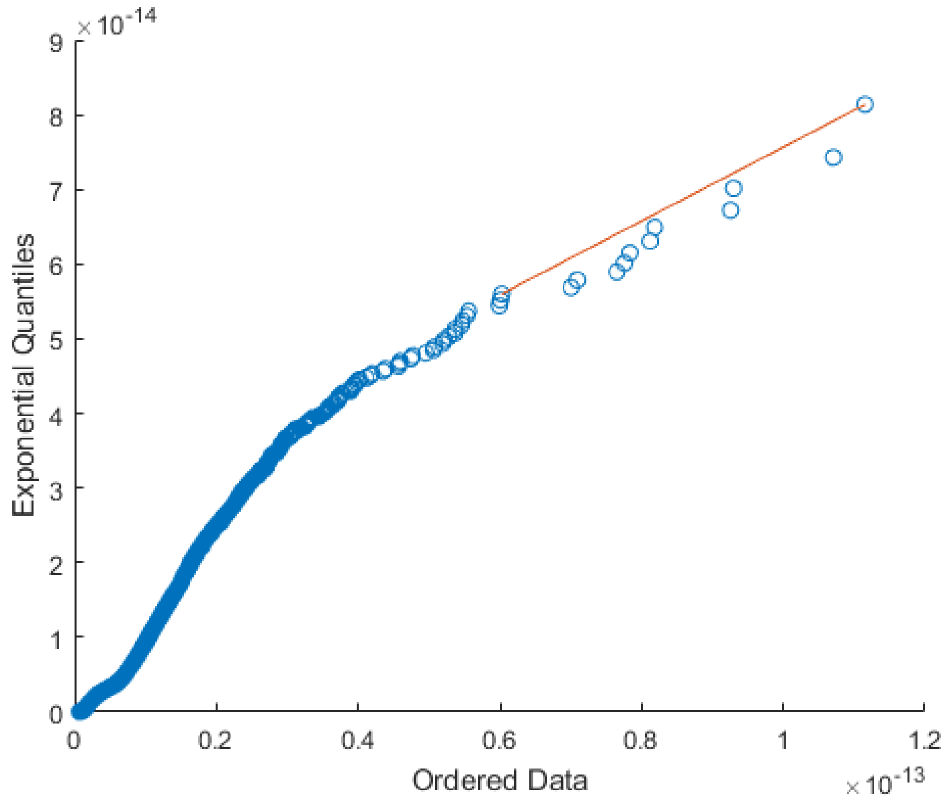

Figure 4.43. QQ plot of TUBITAK (3) data.

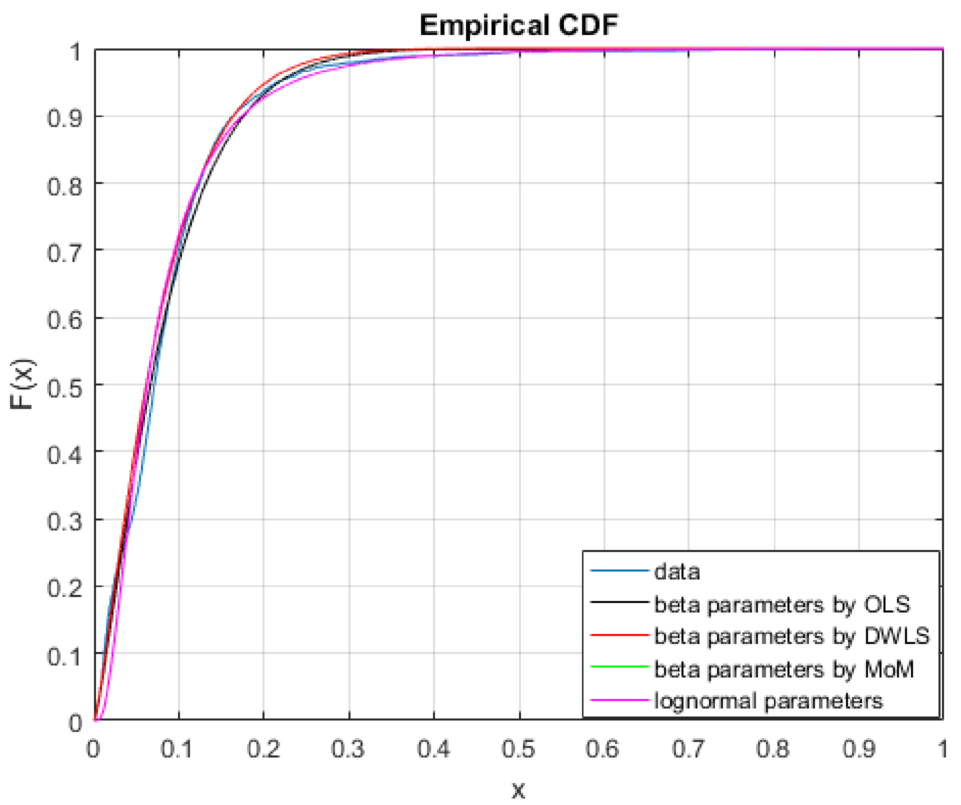

Figure 4.44. Cdf fitting of TUBITAK (3) data.

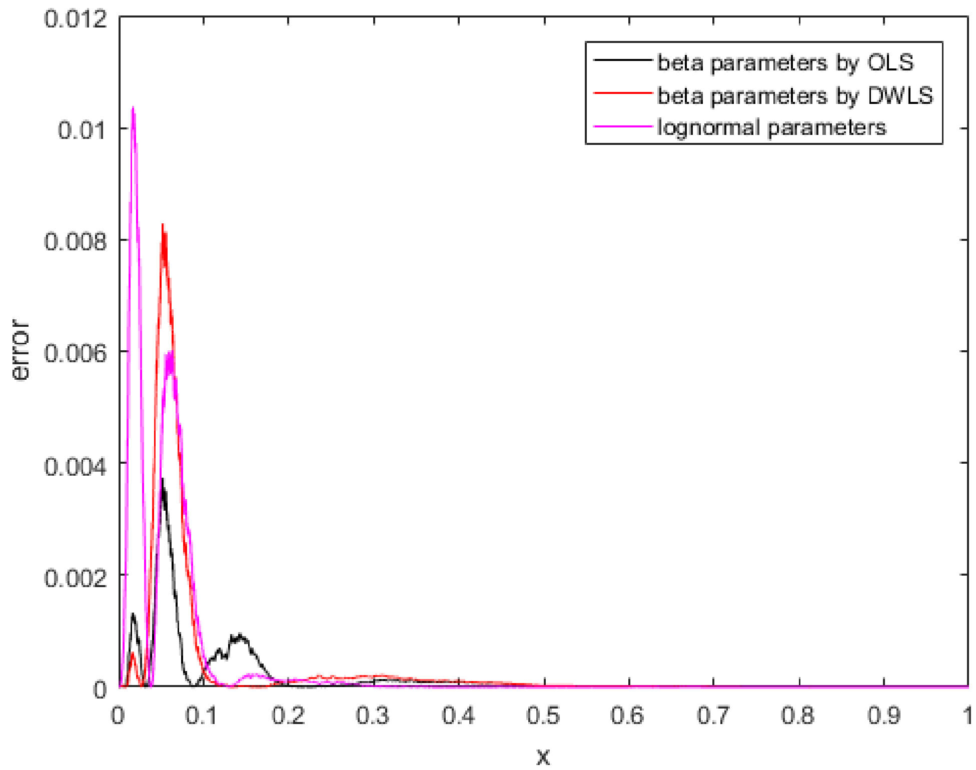

Figure 4.45. Error comparison for TUBITAK (3) data.

MoM could not be tested for goodness of fit because it gives a negative result for one of the parameters. OLS gives the best results for both goodness of fit tests but they are not even close to the critical values (Table 4.18).

Table 4.18. TUBITAK (3) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	1	1	**106.8263**	**0.0611**
DWLS	1	1	122.1348	0.0910
MoM	NaN	NaN	NaN	NaN
Lognormal	1	1	106.8462	0.1019

4.10 TUBITAK (4) Data Results

Figure 4.46 shows that S-K value of the data set is inside the region of the bounded distributions.

Although there are some spikes towards the end, mean excess function shows a downward trend especially after the threshold at 18% (Figure 4.47). 1106 samples are above that threshold and parameter estimation methods are all resulted with positive values for k (Table 4.19). QQ plot in Figure 4.48 shows a clear sign of convexity for both after threshold and in general.

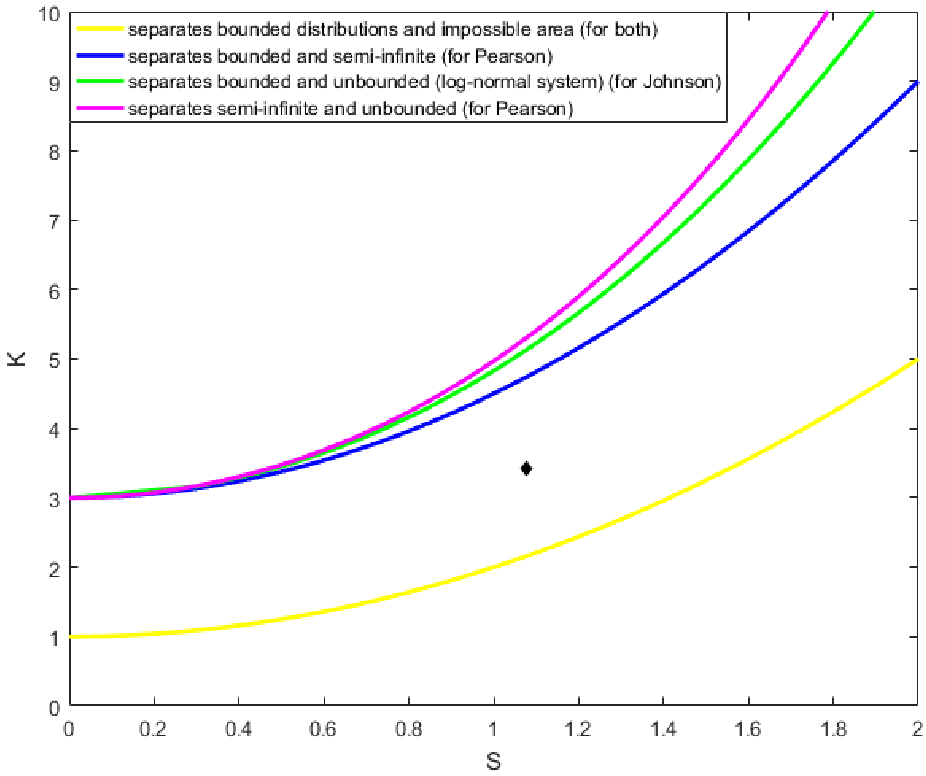

Figure 4.46. Pearson and Johnson distribution systems result of TUBITAK (4) data.

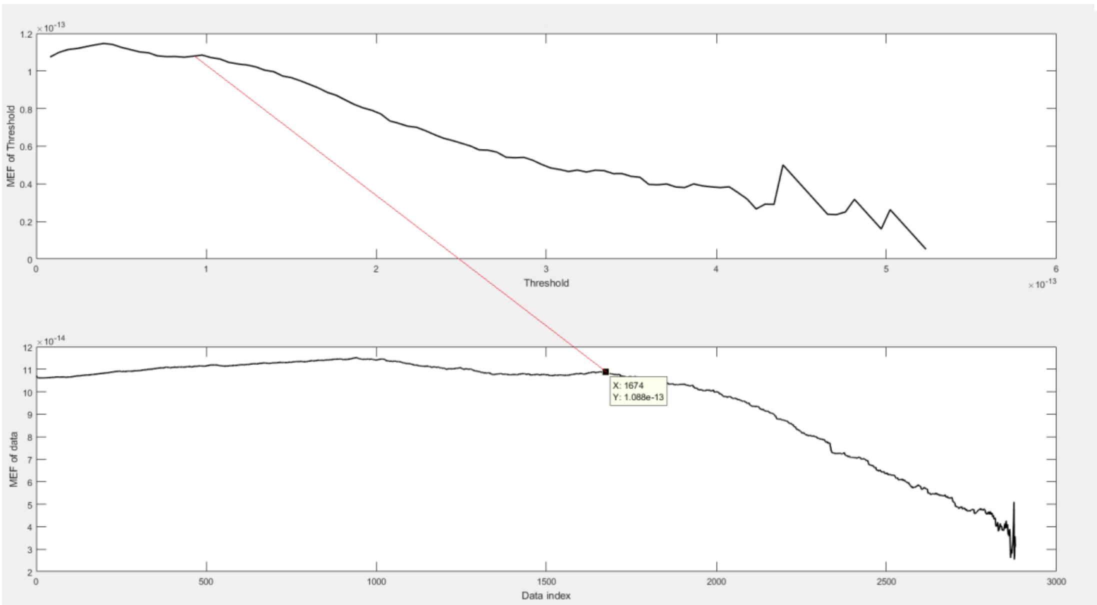

Figure 4.47. Mean excess function of TUBITAK (4) data.

Table 4.19. k value estimation of TUBITAK (4) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
2.697514	2.51722	2.518552	0.43703	0.44433	0.422524	0.390416

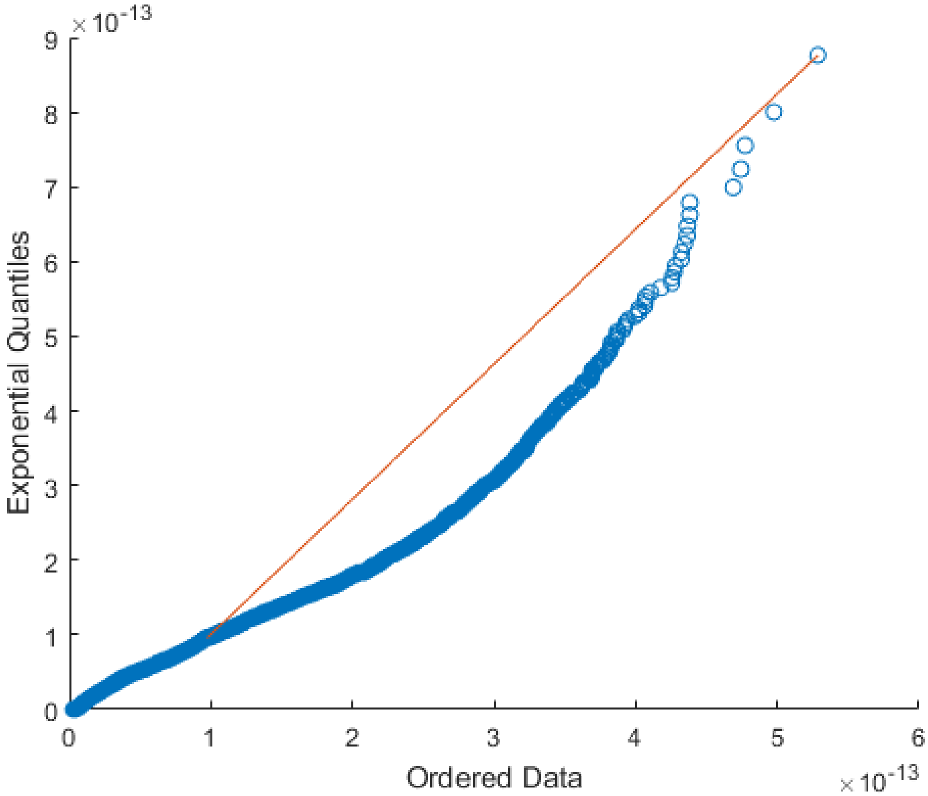

Figure 4.48. QQ plot of TUBITAK (4) data.

Similar to the most data set, OLS and DWLS are very close to each other and better than MoM and lognormal as cdf fit (Figure 4.49). Error comparison in Figure 4.50 also shows how better they are as a fit than lognormal.

OLS gives the best result for the Chi-Square test compared to the others. DWLS comes very close to passing the K-S test with the result just above critical value but every method fails both tests (Table 4.20).

Table 4.20. TUBITAK (4) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	1	1	**42.2114**	0.0400
DWLS	1	1	49.9423	**0.0297**
MoM	1	1	145.4206	0.1050
Lognormal	1	1	866.8770	0.1724

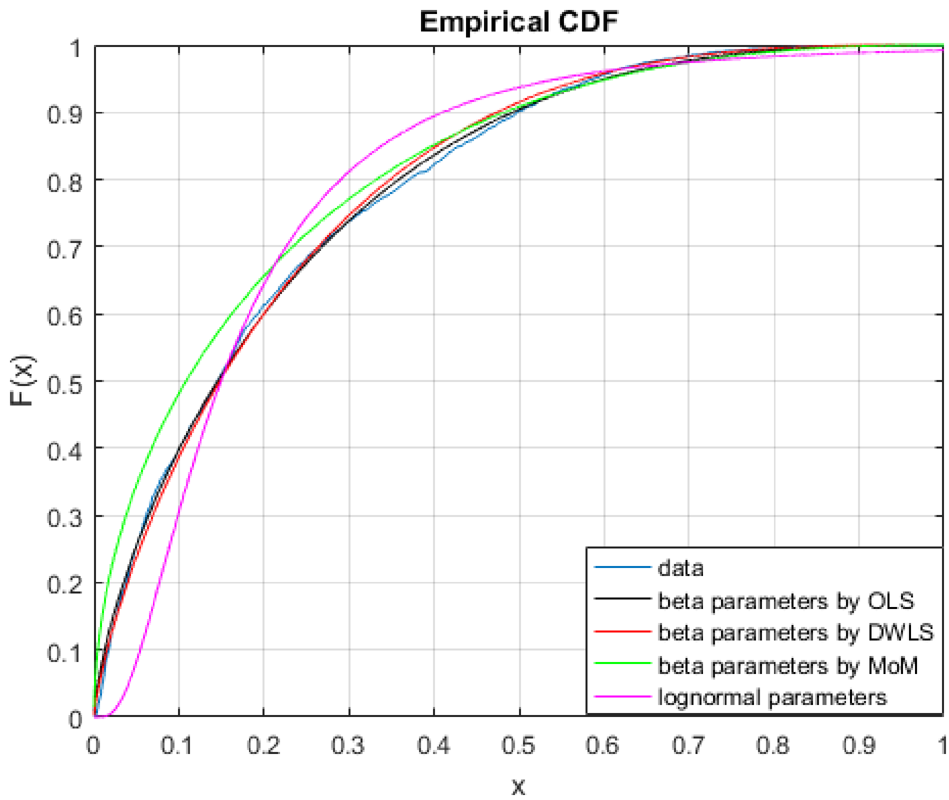

Figure 4.49. Cdf fitting of TUBITAK (4) data.

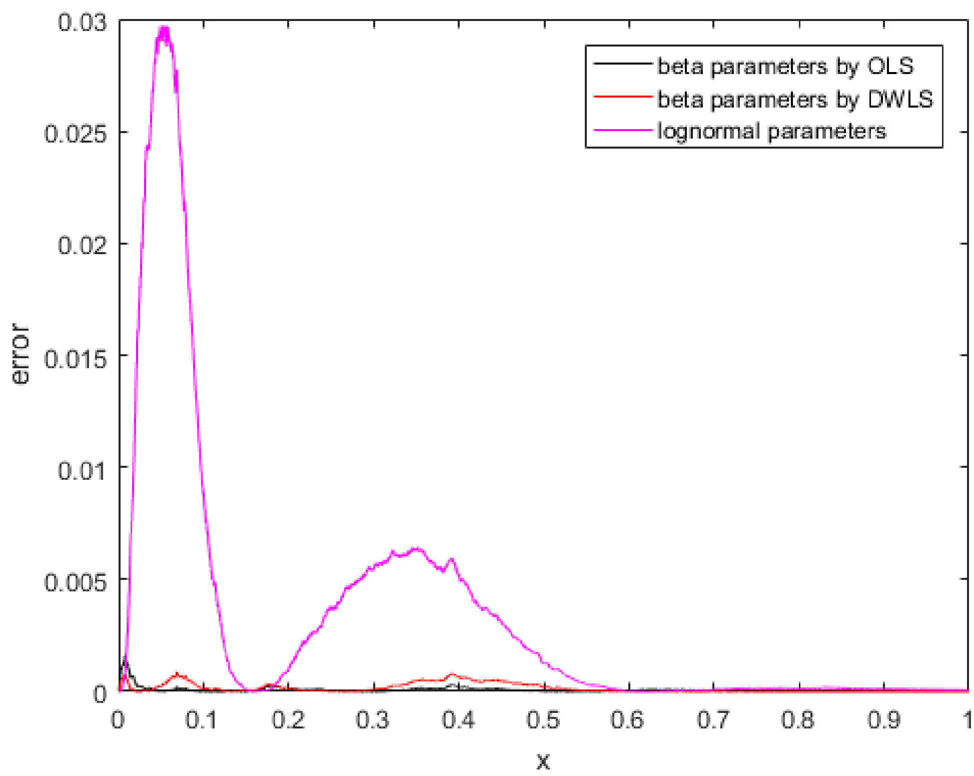

Figure 4.50. Error comparison for TUBITAK (4) data.

4.11 TUBITAK (5) Data Results

S-K value of the data set is well inside the region of the bounded distributions for both Pearson and Johnson distribution systems in Figure 4.51.

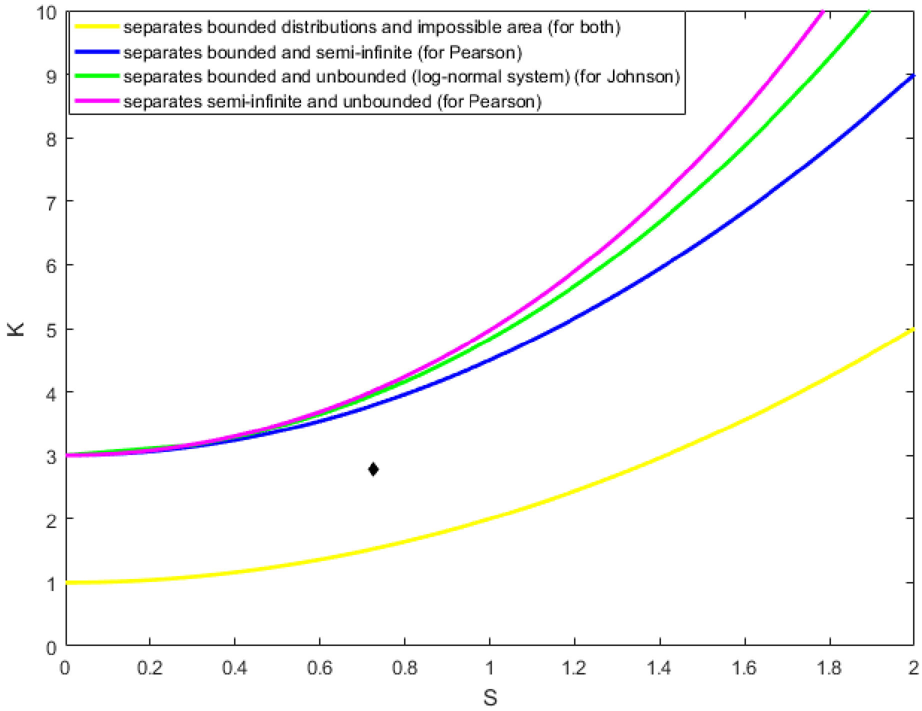

Figure 4.51. Pearson and Johnson distribution systems result of TUBITAK (5) data.

Mean excess function shows a downward trend in general (Figure 4.52). Although both plots, equal-length division of threshold axis and taking sample values as the threshold, shows a breakdown in trend earlier than the chosen threshold, this threshold is chosen to stay in the MATLAB's calculation limits. Threshold chosen is at 25% and the number of samples left for GPD is 1284. Estimations of k yield positive values for every method (Table 4.21). QQ plot in Figure 4.53 of the data set is very similar, almost identical to TUBITAK (4) data's QQ plot and it shows the same convexity for both after threshold and in general.

Table 4.21. k value estimation of TUBITAK (5) data.

MOM	PWM (1)	PWM (2)	L-moment	L1-moment	L2-moment	L3-moment
5.133672	4.023152	4.026942	0.418925	0.398947	0.373105	0.345764

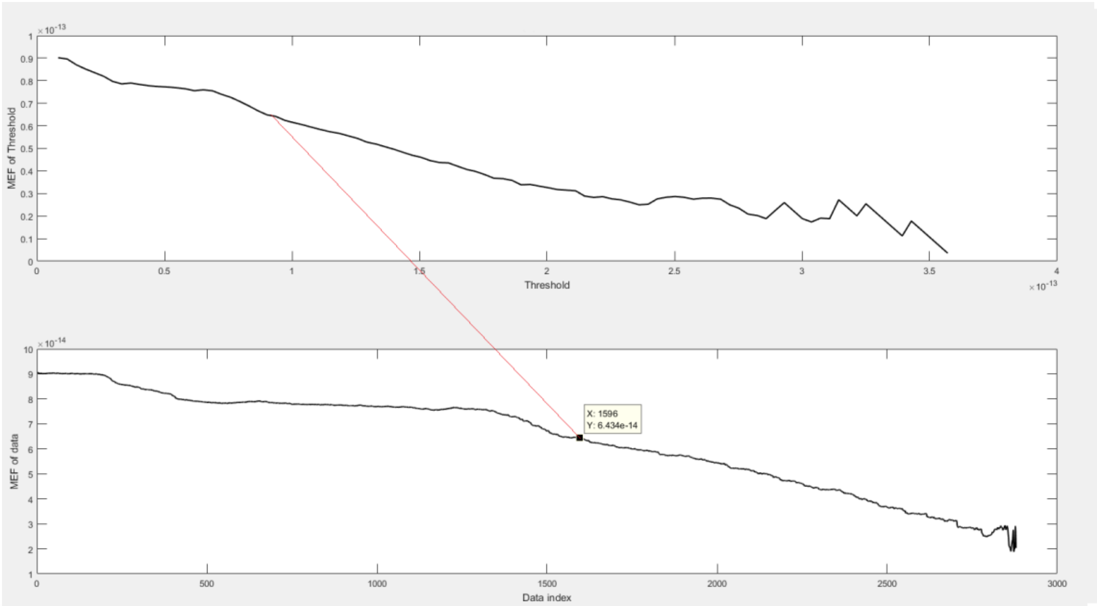

Figure 4.52. Mean excess function of TUBITAK (5) data.

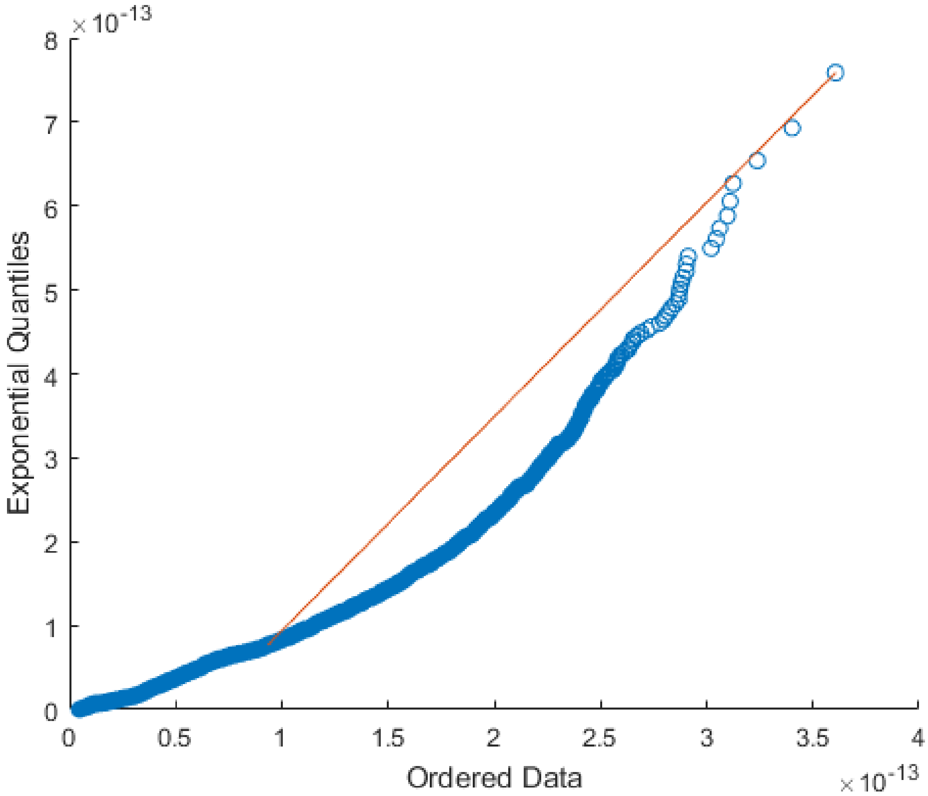

Figure 4.53. QQ plot of TUBITAK (5) data.

Again, cdf fit graph in Figure 4.54 and error comparison graph in Figure 4.55 shows very similar results to TUBITAK (4) data. Lognormal does a little bit better than TUBITAK (4) but not so much.

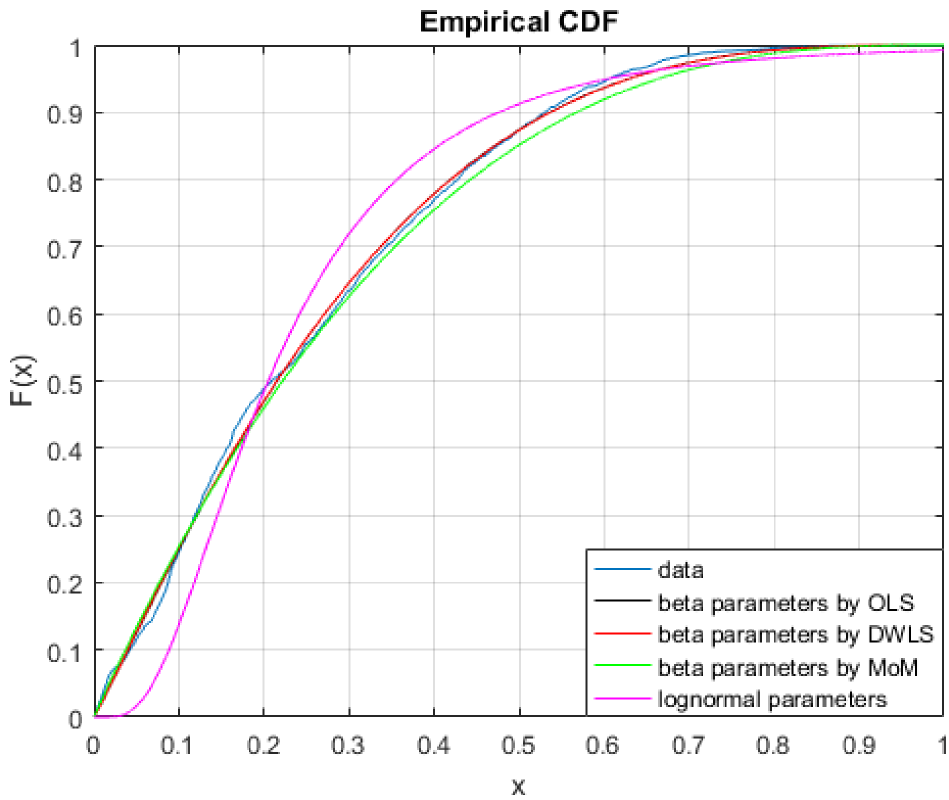

Figure 4.54. Cdf fitting of TUBITAK (5) data.

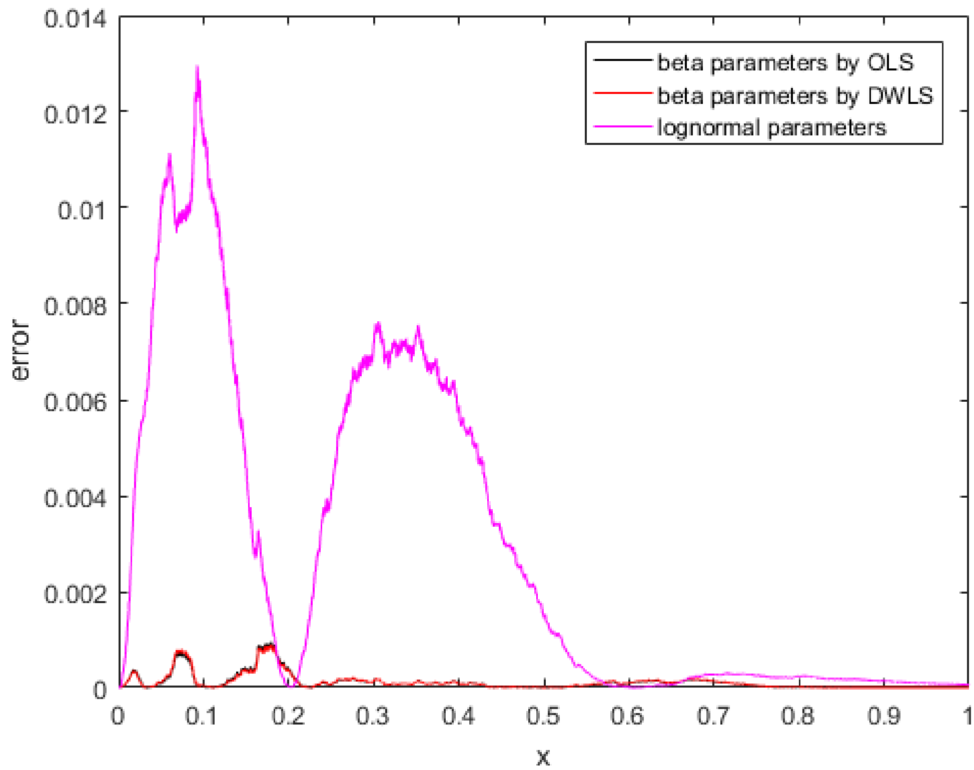

Figure 4.55. Error comparison for TUBITAK (5) data.

OLS and DWLS results for the goodness of fit tests are very close to each other with OLS is just a bit better at Chi-Square test and DWLS is just a bit better at K-S test. Both of them are better than MoM and lognormal (Table 4.22).

Table 4.22. TUBITAK (5) goodness of fit test results.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	1	1	**104.3571**	0.0309
DWLS	1	1	107.4889	**0.0300**
MoM	1	1	215.2604	0.0377
Lognormal	1	1	797.6046	0.1138

4.12 Results of All Data Together

All S-K pairs are placed in the S-K plane and to check whether there is a measurement error in TUBITAK (3) or not, curve fitting algorithm for a second-degree polynomial is applied for twice, with TUBITAK (3) and without TUBITAK (3) data (Figure 4.56 and Figure 4.57 respectively).

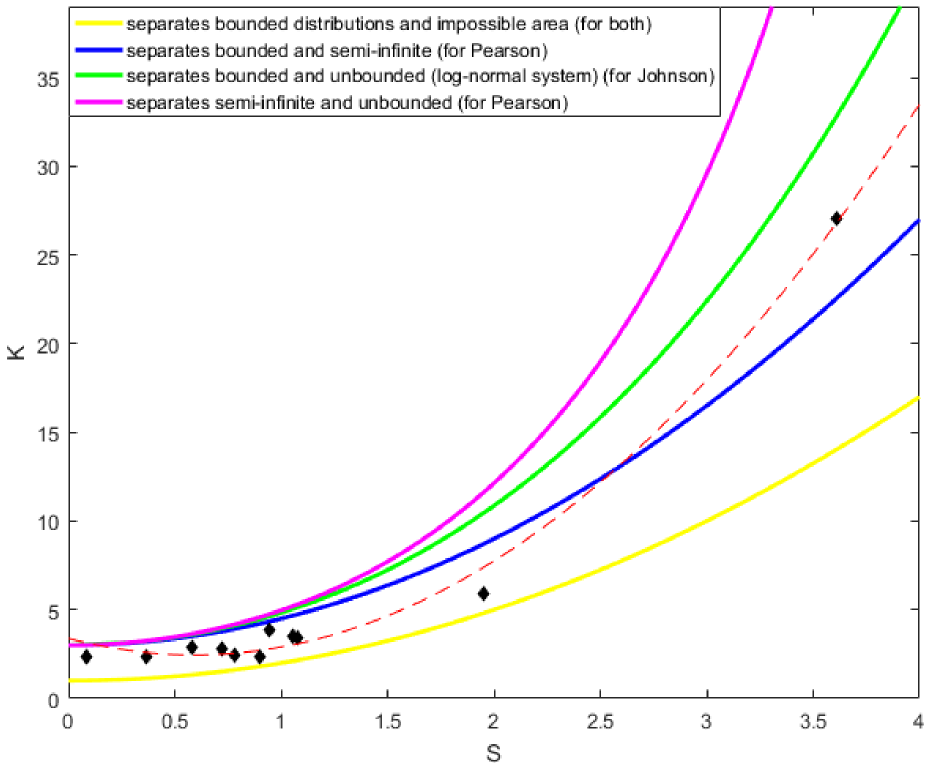

Figure 4.56. Curve fitting with TUBITAK (3) data.

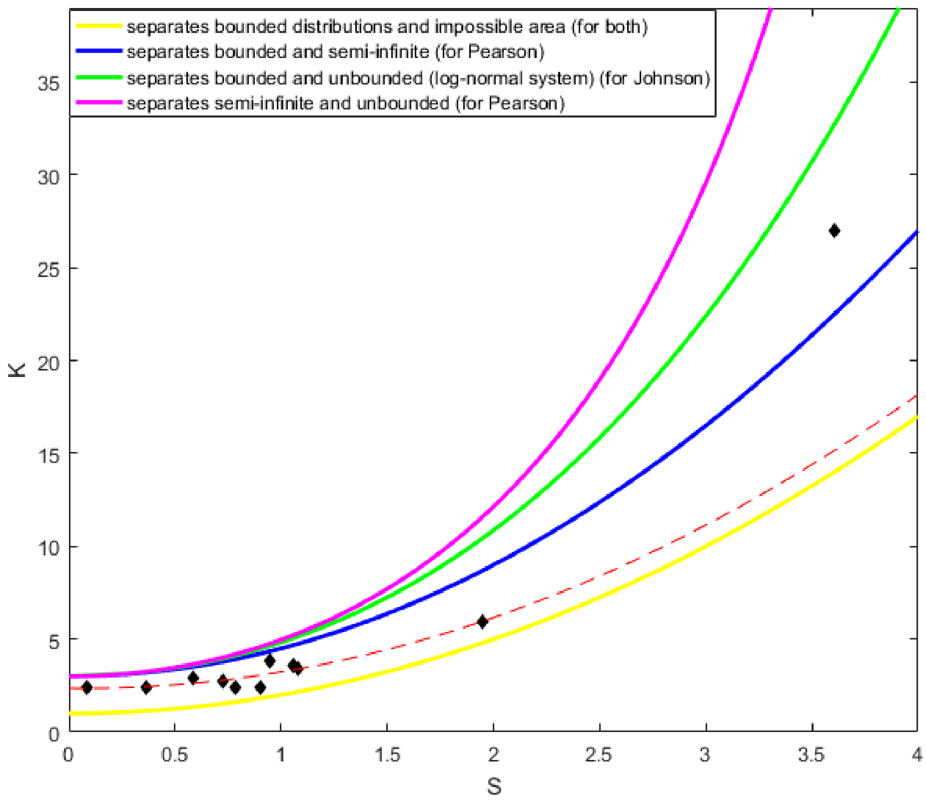

Figure 4.57. Curve fitting without TUBITAK (3) data.

Fitting equation with TUBITAK (3) data is;

$$K = 2.671S^2 - 3.161S + 3.389 \tag{4.1}$$

Fitting equation without TUBITAK (3) data is;

$$K = 1.026S^2 - 0.1479S + 2.357 \tag{4.2}$$

Fitting with TUBITAK (3) data has a high adjusted R-square result of 0.9889 which means very high correlation (with 1 being the highest) but violates the curves that separate regions for both Pearson and Johnson distribution systems. When TUBITAK (3) is left out, the fitted curve is compatible with the boundary lines but has a low adjusted R-square result of 0.8456. Keeping in mind that there are only eleven data sets used in this research, the need for more data set to reach a definite conclusion is apparent.

The goodness of fit test results show that beta distribution is a better fit than lognormal distribution (Table 4.23). Out of three methods to estimate beta distribution parameter, MoM yields the worst results which is expected since the other two are an improvement over MoM. For most cases, OLS and DWLS give similar estimations so their test results are also very similar. OLS does just slightly better since it passes Chi-Square test for all data sets that came from Channel Characterization for Free-Space Optical Communications Project while DWLS fails one of them. It should be noted that DWLS does just a bit better at the fitting of the cdfs at endpoints since it puts more weight on them.

Table 4.23. Performance comparison of estimation methods.

	MATLAB Default Check		**Calculation Results**	
Parameters by	**Chi-Square**	**K-S**	**Chi-Square**	**K-S**
OLS	**6/11**	**5/11**	**7/11**	5/11
DWLS	5/11	**5/11**	4/11	**6/11**
MoM	0/11	1/11	1/11	0/11
Lognormal	1/11	3/11	0/11	0/11

5. CONCLUSION

There is no consensus on which distribution function is the best to model C_n^2, which is the most important parameter for designing outdoor optical wireless communication systems. The tails of the probability distribution function of C_n^2 are especially important because even the errors in the order of one billionth determine the performance of the system. With that in mind, the real focus of this research was determining whether the probability distribution of C_n^2 is bounded as in beta distribution or semi-infinite as in lognormal distribution. These two distributions are chosen because they are the most known and accepted distributions to model C_n^2.

At the start, to get an unbiased result, instead of choosing the distribution beforehand and applying goodness of fit tests, data sets are allowed to determine the best pdf. To do so, Pearson and Johnson distribution systems and Extreme Value Theory are used. It should be noted that Johnson transformation requires the researcher to choose a distribution family beforehand but in this research, Johnson transformation is not applied but only Skewness-Kurtosis plane boundaries of Johnson distribution system are used. Although these methods that help to choose a pdf without prior estimations are used in various scientific fields, as far as known, they are applied to the analysis of C_n^2 for the first time in this research.

With using Skewness-Kurtosis plane to classify data sets, results showed that ten out of eleven data sets are in the bounded distribution family region of the S-K plane for both Pearson and Johnson distribution systems. The remaining one is in the bounded distribution family region for Johnson distribution system but is in the semi-infinite distribution family region for Pearson distribution system. To determine whether this data set is corrupted or not, second order polynomial curve fitting algorithm is applied for two cases. In the first one, a curve is fitted to all eleven data points and although it had a very high correlation, it violated the boundaries of Pearson and Johnson distribution systems. The second case, where the data point in the semi-infinite distribution family region for Pearson distribution system is left out, was compatible with Pearson and Johnson distribution systems but had low correlation. This research is done on limited number of data sets so increasing the number of data sets might give a clearer result.

Extreme Value Theory results were also very similar. Except for the same data set that is in the semi-infinite family region for Pearson distribution system, ten data set showed signs of boundedness. It should be noted that Extreme Value Theory relies on exploratory analysis and some data sets are clearly bounded while some of them depend on the threshold chosen.

During the application of Extreme Value Theory, a new inspection method for the mean excess function is suggested. In traditional mean excess function inspection, threshold axis is divided into equal lengths and because of that, there is a natural downward trend if the data points are separated from each other, which makes interpretation of the plot difficult. Another issue with the traditional method is not knowing how many data points are above the threshold. The proposed method in this thesis is instead of dividing the threshold axis into equal lengths, choosing each data point as a threshold after sorting the data set in ascending order. This way, there're no obsolete thresholds which mean no natural downward trend and also the number of data points that exceed thresholds can be easily seen.

Next step of the research was fitting beta and lognormal distributions to cdf of the data and applying goodness of fit tests, Chi-Square and K-S tests, to see how well they do. With strong hints of boundedness coming from both Pearson and Johnson distribution systems and Extreme Value Theory, beta distribution was the main focus and three methods; MoM, OLS and DWLS are implemented to predict the shape parameters for beta distribution and their results are compared with each other and lognormal distribution parameter estimation result, to see which of them is the best method. Results can be divided into two; small data set results, coming from Channel Characterization for Free-Space Optical Communications Project and big data set results, coming from TUBITAK. For both big data sets and small data sets calculation results of the beta distribution is better than lognormal distribution's. For small data sets, beta distribution passes goodness of fit tests almost every time while lognormal fails most of the time. Comparison between beta distribution parameter estimation methods shows that OLS and DWLS method results are very similar and both viable. They are much better than MoM which is expected since they are an improvement over MoM. For big data sets, both lognormal and beta distribution fail the tests every time. There are two possible ways to overcome this problem. Either another distribution from bounded distributions family might be used as a model or, a method that results in even better improvement for estimating beta distribution parameter can be searched for since improved methods OLS and DWLS did a lot better than MoM.

REFERENCES

AbouRizk, S.M., 1990. Input modelling for construction simulations, Doctorate Thesis, Purdue University, West Lafayette, Indiana, USA.

AbouRizk, S.M., Halpin, D.W. and Wilson, J.R., 1994. Fitting beta distributions based on sample data, Journal of Construction Engineering and Management, 120, 2, 288-305.

Acero, F.J., Gallego, M.C., Garcia, J.A., Usoskin, I.G. and Vaquero, J.M., 2018, Extreme Value Theory applied to the millennial sunspot number series, The Astrophysical Journal, 853, 80, 1-6.

Andreas, E.L., 1989. The refractive-index structure parameter, C_n^2, for a year over the frozen Beaufort Sea, Radio Science, 24, 5, 667-679.

Andreas, E.L., Fairall, C.W., Persson, P.O.G. and Guest, P.S., 2003. Probability distributions for the inner scale and the refractive index structure parameter and their implications for flux averaging, Journal of Applied Meteorology, 42, 9, 1316-1329.

Andrews, L.C. and Phillips, R.L., 2005. Laser beam propagation through random media, Second Edition, SPIE Press, Washington, USA.

Andrews, L.C., Phillips, R.L., Crabbs, R., Leclerc, T. and Sauer, P., 2012. Channel characterization for free-space optical communications, UCF DARPA Project ID: 66016010, Final Report, July 2012.

Baci, N., Şahin, S., Kaya, K. and Yeniay, A., 2018. Yapay sinir ağları ile C_n^2 tahmini, 9. Savunma Teknolojileri Kongresi, Ankara, Turkey.

Balkema, A.A. and De Haan, L., 1974. Residual life time at great age, The Annals of Probability, 2, 5, 792-804.

Bermudez, P. de Z. and Kotz, S., 2010. Parameter estimation of the generalized Pareto distribution – Part I, Journal of Statistical Planning and Inference, 140, 1353-1373.

Chadwick, R.B. and Moran, K.P., 1980. Long-term measurements of C_n^2 in the boundary layer, Radio Science, 15, 2, 355-361.

Chatwin, P.C., Lewis, D.M. and Sullivan, P.J., 1995. Turbulent dispersion and the beta distribution, Environmetrics, 6, 395-402.

Chen, W.Y., 1971. Lognormality of small-scale structure of turbulence, The Physics of Fluids, 14, 8, 1639-1642.

Coles, S., 2001. An introduction to statistical modelling of extreme values, First Edition, Springer-Verlag London Limited, London, Great Britain.

De Haan, L., 1990. Fighting the arch-enemy with mathematics, Statistica Neerlandica, 44, 2, 45-68.

DeBrota J.D., Roberts, D.S., Swain, J.J., Dittus, R.S., Wilson, J.R. and Venkatraman S., 1988. Input modelling with the Johnson system of distributions, Proceedings of the 1988 Winter Simulation Conference, San Diego, California, USA.

Doane, D.P., 1976. Aesthetic frequency classifications, The American Statistician, 30, 4, 181-183.

Embrechts, P., Klüppelberg, C. and Mikosch, T., 1997. Modelling extremal events for insurance and finance, First Edition, Springer-Verlag, Berlin, Germany.

Fisher, R.A. and Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proceedings of the Cambridge Philosophical Society, 24, 180-190.

Freedman, D. and Diaconis, P., 1981. On the histogram as a density estimator: L_2 theory, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 453-476.

George, F., 2007. Johnson's system of distributions and microarray data analysis, Doctorate Thesis, University of South Florida, Tampa, Florida, USA.

Gilli, M. and Kellezi E., 2006. Application of Extreme Value Theory for measuring financial risk, Computational Economics, 27, 1, 1-23.

Gkillas, K. and Katsiampa, P., 2018. An application to Extreme Value Theory to cryptocurrencies, Economics Letters, 164, 109-111.

Gnedenko, B., 1943. Sur la distribution limite du terme maximum d'une série aléatoire, Annals of Mathematics, Second Series, 44, 3, 423-453.

Graf, A., Schüttemeyer, D., Geiß, H., Knaps, A., Möllmann-Coers, M., Schween, J. H., Kollet, S., Neininger, B., Herbst, M. and Vereecken, H., 2010. Boundedness of turbulent temperature probability distributions and their relation to the vertical profile in the convective boundary layer, Boundary Layer Meteorology, 134, 459-486.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form, Water Resources Research, 15, 5, 1049-1054.

Gurvich, A.S. and Yaglom, A.M., 1967. Breakdown of eddies and probability distributions for small scale turbulence, The Physics of Fluid Supplement, 10, 59-65.

Hosking, J.R.M., 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics, Journal of the Royal Statistical Society, 52, 1, 105-124.

Hosking, J.R.M. and Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution, Technometrics, 29, 3, 339-349.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moment, Technometrics, 27, 3, 251-261.

Huyse, L. and Thacker, B.H., 2004. A framework to estimate uncertain random variables, 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, Palm Springs, California, USA

Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation, Biometrika Trust, 36, 1, 149-176.

Johnson, N.L., Kotz, S. and Balakrishnan, N., 1995. Continuous univariate distributions, volume 2, Second edition, John Wiley & Sons, Inc., USA.

Küchenhoff, H. and Thamerus, M., 1995, Extreme value analysis of Munich air pollution data, Sonderforschungsbereich 386, 4, 1-24.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, Water Resources Research, 15, 5, 1055-1064.

McNeill, A.J., 1997. Estimating the tails of loss severity distributions using Extreme Value Theory, ASTIN Bulletin: The Journal of the IAA, 27, 1, 117-137.

Meshgi, A. and Khalili, D., 2009. Comprehensive evaluation of flood frequency analysis by L- and LH-moments. II. Development of LH-moments parameters for the generalized Pareto and generalized logistic distributions, Stochastic Environmental Research and Risk Assessment, 23, 137-152.

Mole, N. and Clarke, E.D., 1995. Relationships between higher moments of concentration and of dose in turbulent dispersion, Boundary-Layer Meteorology, 73, 35-52.

Obukhov, A.M., 1962. Some specific features of atmospheric turbulence, Journal of Geophysical Research, 67, 8, 3011-3014.

Owens, J.C., 1967. Optical refractive index of air: Dependence on pressure, temperature and composition, Applied Optics, 6, 1, 51-59.

Pearson, K., 1895. X. Contributions to the mathematical theory of evolution. II. skew variation in homogeneous material, Philosophical Transactions of the Royal Society of London, 186, 1887-1895.

Pearson, K., 1901. XI. Mathematical contributions to the theory of evolution. – X. Supplement to a memoir on skew variation, Philosophical Transactions of the Royal Society of London, 197, 443-459.

Pickands III, J., 1975, Statistical inference using extreme order statistics, The Annals of Statistics, 3, 1, 119-131.

Podladchikova, O., Lefebvre, B., Krasnoselskikh, V. and Podladchikov, V., 2003. Classification of probability densities on the basis of Pearson's curves with application to coronal heating simulations, Nonlinear Processes in Geophysics, European Geosciences Union (EGU), 10, 4, 323-333.

Smith, R.L., 1989. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone, Statistical Science, 4, 4, 367-393.

Stewart, R.W., Wilson, J.R. and Burling, R.W., 1970. Some statistical properties of small scale turbulence in an atmospheric boundary layer, Journal of Fluid Mechanics, 41, 1, 141-152.

Sturges, H.A., 1926. The choice of a class interval, Journal of the American Statistical Association, 21, 153, 65-66.

Wang, Q.J., 1997. LH moments for statistical analysis of extreme events, Water Resources Research, 33, 12, 2841-2848.

Weng, T.W., Zhang, H., Chen, P.Y., Yi, J., Su, D., Gao, Y., Hsieh, C.J. and Daniel, L., 2018. Evaluating the robustness of neural networks: An Extreme Value Theory approach, 6th International Conference on Learning Representations (ICLR), Vancouver, British Columbia, Canada.

CURRICULUM VITAE

Name and Surname : İbrahim Yazar

Address : Coğlaki Mah. 1215 Sok. No: 21 Merkez/AKSARAY

E-mail address : ibrahimyazar@gmail.com

EDUCATION INFORMATION (Institution and Year)

Bachelor's Degree : İhsan Doğramacı Bilkent University, 2004-2015

Master's Degree : Aksaray University, 2016-Present

JOB EXPERIENCE AND REWARDS

1. Mercedes-Benz Türk A.Ş. (2007)/ Summer Practice
2. Ereğli Demir ve Çelik Fabrikası T.A.Ş. (2013)/ Summer Practice
3. Aksaray University (2016-Present)/ Teaching Assistant

PUBLICATIONS, PRESENTATIONS AND PATENTS FROM THESIS

Conference Paper

1. Yazar, İ., Yenice, Y. E. and Sarı, F., 2017. Probability distribution of refractive index structure parameter, 25th Signal Processing and Communication Applications Conference (SIU), Antalya, Turkey.