ACTUARIAL MODELLING OF EXTREMAL EVENTS USING TRANSFORMED GENERALIZED EXTREME VALUE DISTRIBUTIONS AND GENERALIZED PARETO DISTRIBUTIONS

A Thesis

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of the Ohio State University

By

Zhongxian Han, B.S., M.A.S.

* * * * *

The Ohio State University
2003

Doctoral Examination Committee:

Bostwick Wyman, Advisor

Peter March

Gerald Edgar

Approved by

_____

Advisor
Department of Mathematics

ABSTRACT

In 1928, Extreme Value Theory (EVT) originated in work of Fisher and Tippett describing the behavior of maximum of independent and identically distributed random variables. Various applications have been implemented successfully in many fields such as: actuarial science, hydrology, climatology, engineering, and economics and finance.

This paper begins with introducing examples that extreme value theory comes to encounter. Then classical results from EVT are reviewed and the current research approaches are introduced. In particular, statistical methods are emphasized in detail for the modeling of extremal events. A case study of hurricane damages over the last century is presented using the "excess over threshold" (EOT) method.

In most actual cases, the range of the data collected is finite with an upper bound while the fitted Generalized Extreme Value (GEV) and Generalized Pareto (GPD) distributions have infinite tails. Traditionally this is treated as trivial based on the assumption that the upper bound is so large that no significant result is affected when it is replaced by infinity. However, in certain circumstances, the models can be improved by implementing more specific techniques. Different transforms are introduced to rescale the GEV and GPD distributions so that they have finite supports.

All classical methods can be applied directly to transformed models if the upper bound is known. In case the upper bound is unknown, we set up models with one additional parameter based on transformed distributions. Properties of the transform functions are studied and applied to find the cumulative density functions (cdfs) and probability density functions (pdfs) of the transformed distributions. We characterize the transformed distribution from the plots of their cdfs and mean residual life. Then we apply our findings to determine which transformed distribution should be used in the models. At the end some results of parameter estimation are obtained through the maximum likelihood method.

Dedicated to

My Parents

ACKNOWLEDGMENTS

VITA

December 18, 1975 .... Born - Lanzhou, P.R.China.

1997 .... B.S., Peking University Beijing, China.

1997 - present .... Graduate Teaching and Research Associate, The Ohio State University.

2001 .... Master in Applied Statistics   The Ohio State University.

FIELDS OF STUDY

Major field: Mathematics

Specialization: Extreme Value Theory

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION OF EXTREME VALUE THEORY

1.1 Extremal Events

Extremal events are also called "rare" events. Extremal events share three characteristics: "relatively rareness," "huge impact," and "statistical unexpectness."

From the name of it, extremal events are extreme cases; that is, the chance of occurrence is very small. But over the last decades, we have observed several. We list some recent events with a large impact in time order: Hurricane Andrew (1992), Northridge earthquake (1994), terrorism attack (2001), disintegration of space shuttle Columbia (2003), winter blizzard of the northeast (2003). If we assume those events are in different fields and independent of each other, the probability of the occurrence of each event is small. The terminology **t-year event** has been used by hydrologists to describe how frequent a certain event will occur. In general, suppose for every year, the probability of a certain event is p, where p is relatively small. Given that the events of each year are independent of each other, the number of the year when the event occurs follows a geometric distribution with parameter p, and thus has an expectation of $\dfrac{1}{p}$. On the other words, we say Hurricane Andrew is a 30-year event

means that for every year the probability of having a hurricane which is more severe than Hurricane Andrew is approximately $\frac{1}{30}$.

The enormous impact of catastrophic events on our society is deep and long. Not only we need to investigate the causes of such events and develop plans to protect against them, but also we will have to resolve the resulting huge financial loss. Financial plans have been established for the reduction of the impact of extremal events. For example, the Homeland Security Act of 2002 was passed on Nov 15, 2002 to prevent future terrorism attacks. In the insurance industry of, the magnitude of property-casualty risk has become a major topic of research and discussion. Traditionally, insurers buy reinsurance to hedge against catastrophic risk. For a catastrophic event causing more than $5 billion of insured losses, the overwhelming majority is not covered by reinsurance, according to Froot and Connell [11]. On December 11, 1992, the Chicago Board of Trade launched the first generation of catastrophe insurance futures, also called CAT-Futures. The related future PCS options were launched shortly after. CAT-Futures provides a way of transferring insurance risk to the financial market and CAT-Futures can be traded as standardized securities. Table 1.1 lists 10 catastrophic events that caused the largest losses in history of the United States. Since the population and fortune grow at a relatively constant rate, we expect larger catastrophic events entering the list in the future. The task of understanding and resolving the underlying risk is still a big challenge for researchers.

Year	Catastrophic Event	Estimated Losses (in billions)
2001	Fire and explosion	20.3
1992	Hurricane Andrew	15.5
1994	Northridge earthquake	12.5
1989	Hurricane Hugo	4.2
1998	Hurricane Georges	3.0
2001	Tropical Storm Allison	2.5
1995	Hurricane Opal	2.1
1999	Hurricane Floyd	2.0
2001	St.Louis hailstorm	1.9
1993	Midwest blizzard	1.8

TABLE 1.1 10 Largest Catastrophe Events in US (Source: ISO's PCS).

In the modeling of extremal events, statistical methods are commonly used for inference from historical data. However, classical statistical methods treat those extreme data, exactly the point of our interest, just as any other data. In some cases, these data are even treated as "outliers" and dropped from the analysis. Therefore, any forecasting or prediction could be largely off the target. Table 1.2 shows loss-ratios (yearly data) for earthquake insurance in California from 1991 till 1994. The data are taken from Jaffe and Russell [15].

Note: "Loss-ratio" is the ratio (in percentage) of total claims paid to total premiums collected during a certain period (usually one year) for policies in a portfolio.

Year	Loss Ratio	Year	Loss Ratio	Year	Loss Ratio
1971	17.4	1979	2.2	1987	22.8
1972	0	1980	9.2	1988	11.5
1973	0.6	1981	0.9	1989	129.8
1974	3.4	1982	0	1990	47.0
1975	0	1983	2.9	1991	17.2
1976	0	1984	5.0	1992	12.8
1977	0.7	1985	1.3	1993	3.2
1978	1.5	1986	9.3	1994	?

TABLE 1.2 California Earthquake Data (1971 – 1994).

Suppose we want to make a guess on the 1994 entry by having the data from 1971 to 1993. Any experienced statistician would be impressed by the relatively low values (which means the insurance company makes profits) during all years but 1989 (year of the earthquake in San Francisco). Even though there seems an increasing trend of loss-ratio in time, a reasonable guess would still be between 0 to 50. If we apply a general statistical model (say, a linear regression) for the above data, our prediction to 1994 year data is 31.7, with a 95% confidence interval of $[0.0, 83.0]$. It is time for the true data, which is amazingly 2272.7 ! Indeed, an earthquake recorded at 6.6 Richter scale hit Northridge, California on Jan 17, 1994. The insured damage was about 12.5 billion while the total damage exceeded 30 billion. Obviously, the insurance and reinsurance industry needs to reevaluate the risk in insuring future

4

damages. Extreme Value Theory (EVT) emerges as a basic tool in modeling such risk.

1.2 Modeling Extremal Events

In the modeling of extremal events, different approaches had been proposed for certain circumstances. We would like to introduce some illustrative examples.

On February 1, 1953, a severe storm hit the Netherlands and caused a record storm surge. At various locations the sea dikes collapsed and the flood hit a large part of coastal Holland. The property and casualty loss was so huge that the Delta-committee was established later by the Dutch government. The government required the Delta-committee to estimate the necessary height of the dikes to prevent damages from any future sea surge for a t-year event where t is sufficiently large.

To simplify the problem, suppose $X_1, X_2, ..., X_n$ are hourly or daily observations of sea level in a year. Let

$$M_n = max(X_1, X_2, ..., X_n)$$

be the annual maximum sea level, which we call a **block maximum**. A t-year event corresponds to a value x_t which is defined by:

$$x_t = inf\{x \in R : F(x) \geq (1 - \frac{1}{t})\}$$

where F is the distribution function of M_n. We can interpret x_t as the $(1 - \frac{1}{t})$ upper quantile of the distribution F. As we discussed in section 1.1, assuming that the annual maxima M_n are identically and independently distributed, the expected

time of occurrence of an more extremal event than $\{M_n \geq x_t\}$ is t years. The historical data recorded a previous record surge in year 1570 with a recorded value of $(NAP + 4)meters(m)$, where NAP stands for normal Amsterdam level. The 1953 surge recorded with value $(NAP + 3.85)m$. A quick approximation is that $(NAP + 3.85)m$ corresponds to a 383-year event. The Delta-committee report led to an estimate of $(NAP + 5.14)m$ for 10,000-year event, based on a total of 166 historical observations. For further readings on this topic, the paper by de Haan [13] is recommended.

A second example relates to the stochastic behavior of the stock market. An extremal event occurs when the stock market rallies or crashes. For managers in charge of portfolios worth millions, lowering the risk during market crashes is as important as maximizing the profit. Analysis can be applied to a single stock, stocks in a particular sector, or major indices that represent the overall market movement.

Using stock market crashes as examples, we can trace the overall stock market performance for the last century by the Dow Jones Index. To identify stock market crashes, as pointed out in Mishkin and White [16], October 1929 and October 1987 can be used as benchmarks. On October 28 and 29, 1929, the Dow Jones declined 12.8 and 11.7 percent; and on October 19, 1987, the Dow Jones fell 22.6 percent. To capture the speed and duration characteristics of a market crash, we can use a moving window with a fixed length. The length could be chosen as one day, two days, one week, one month and one year. The identification of a market crash is equivalent to detect windows that have an overall decline of more than 20%.

In modeling market crashes as an extremal events, we let S_n be the close market index on the nth window. Define $C_n = \dfrac{S_n - S_{n-1}}{S_{n-1}}$ be the percentage change in the nth window. We can split C_n into a positive part P_n and a negative part N_n. To study the market crashes, we define

$$M_n = \min \{N_n : \text{for n such that } C_n < 0\}$$

The objective is to fit historical data into a model to get a better understanding of the conditional distribution of $(M_n - u \,|\, M_n < u)$ where M_n is defined as above, and u is a "high threshold" chosen so that the Extreme Value Theory conditions are satisfied.

On the other hand, we can rewrite $-M_n$ as a maximum:

$$M_n = -\max \{-N_n : \text{for n such that } C_n < 0\}$$

In the similar way, we can see that any minimization problem can be treated as a maximization problem. Our concern is to fit historical data into a model to get an approximation of its distribution.

Finally, modeling extremal events in climatology is a very interesting task. In general, the iid assumptions are not valid and thus more sophiscated techniques need to be employed. Two examples will be presented as follows.

Weather conditions are an important factor in crop growth of and affect the whole agricultural economy. The study of extreme weather conditions at a certain location attracts more and more scientists for the purpose of accurate prediction and prevention procedures. With available historical data, analysis can be done on a vast of aspects, including long period of drought, heavy precipitation, extreme low temperatures during winter, etc.

7

One important feather of extreme weather data is the attached spatial character. Different locations may focus on different categories of weather conditions. For example, snow storms and extreme low temperature during winter are not concerns at south Florida, but they would affect most areas in the northeast. Also, if we compare data collected from New York City and Boston, it is not a surprise to find some positive correlation. How to include spatial factors into the mathematical model is still a task which needs to be explored.

Seasonality is another feather of most weather data. Using extreme low temperature as example, the subject of the study could be simply the absolute low temperature ($< 0F^\circ$ any time) or including relative low temperature ($36F^\circ$ below average). In the later case, we would have to model the non-stationary process of annual temperatures. In [5], Coles and Tawn fit the expected annual temperature with a sinusoidal curve and then use the difference of expected and actual in the analysis of extreme cases. In the similar way, long term's trend in the data can be introduced into the model easily.

A second example is the hazard of hurricanes for the southeast coastal area. Causing the costliest losses among natural disasters, the annual variation of hurricane hurricanes frequency and intensity needs to be studied. There are many ways to evaluate the intensity of a hurricane, including Saffir-Simpson Hurricane Scale which uses the wind speed at the center of the storm when it is land-falling. For example, hurricane Andrew of 1992 was firstly assessed to be a Category 4 and later upgraded to be a Category 5 - the highest intensity scale possible - at its land fall in southeastern Florida with peak sustained winds of 165 mph.

However, the total damages that are caused by a hurricane is a more important concern to the insurance industry. In order to make use of historical data, we need to estimate the total loss if a certain hurricane would happen today. There are several factors we need to count: intensity of the hurricane, population affected by hurricane, average fortune index, inflation factors, and the insured property ratio. This issue will be revisited in Chapter 3 in detail, where we will do analysis on the adjusted data. From next section, we will introduce the main extreme value theory and necessary mathematical and statistical tools for the analysis.

1.3   Extreme Value Theory

Theorem 1.1. *Extreme Value Theorem (Fisher and Tippett, 1928) Suppose $X_1, X_2, ..., X_n$ are iid with distribution (df) F. If there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$\frac{M_n - d_n}{c_n} \to Y, \quad n \to \infty \tag{1.1}$$

where $M_n = X_{1,n} = max(X_1, ..., X_n)$, Y is non-degenerate with distribution function G. Then G is of one the following types:

1. (Gumbel)

$$\Lambda(x) = exp\{-e^{-x}\}, \quad x \in \mathbb{R}$$

2. (Frechet)

$$\Phi_\alpha(x) = \begin{cases} 0, & if\ x \leq 0 \\ exp\{-x^{-\alpha}\}, & if\ x > 0 \end{cases}$$

3. (Weibull)

$$\Psi_\alpha(x) = \begin{cases} exp\{-(-x)^\alpha\}, & \text{if } x < 0 \\ 0, & \text{if } x \geq 0 \end{cases}$$

Sketch of the Proof   The following theorem is important in the proof.

Theorem 1.2 (Convergence to types theorem). *Let $A, B, A_1, A_2, \cdots$ be random variables and $b_n > 0$, $\beta_n > 0$ and $a_n, \alpha_n \in \mathbb{R}$ be constants. Suppose that*

$$\frac{A_n - a_n}{b_n} \xrightarrow{d} A$$

Then the relation

$$\frac{A_n - \alpha_n}{\beta_n} \xrightarrow{d} B \tag{1.2}$$

holds iff

$$\lim_{n\to\infty} \frac{b_n}{\beta_n} = b \geq 0, \quad \lim_{n\to\infty} \frac{a_n - \alpha_n}{\beta_n} = a \in \mathbb{R} \tag{1.3}$$

If (1.2) holds then $B \xrightarrow{d} bA + a$ and a, b are the unique constants for which this holds. And A is non-degenerate iff $b > 0$, and then A and B belong to the same type.

Proofs of theorem (1.2) could be found at Gnedenko [12] or Resnick [18]. Equation (1.1) implies that for any $t > 0$,

$$F^{\lfloor nt \rfloor}(c_{\lfloor nt \rfloor}x + d_{\lfloor nt \rfloor}) \to G(x), \quad x \in \mathbb{R}$$

where $\lfloor \cdot \rfloor$ denotes the function taking integer part. On the other hand,

$$F^{\lfloor nt \rfloor}(c_{\lfloor nt \rfloor}x + d_{\lfloor nt \rfloor}) = \left[F^{\lfloor n \rfloor}(c_{\lfloor nt \rfloor}x + d_{\lfloor nt \rfloor})\right]^{\lfloor nt \rfloor/n} \to G^t(x), \quad x \in \mathbb{R}$$

So that by theorem (1.2), there exists functions $\gamma(t) > 0$, $\delta(t) \in \mathbb{R}$ satisfying

$$\lim_{n\to\infty} \frac{c_n}{c_{\lfloor nt \rfloor}} = \gamma(t), \quad \lim_{n\to\infty} \frac{d_n - d_{\lfloor nt \rfloor}}{c_{\lfloor nt \rfloor}} = \delta(t), \quad t > 0$$

10

and

$$H^t(x) = H[\gamma(t)x + \delta(t)] \tag{1.4}$$

then

$$H[\gamma(st)x + \delta(st)] = H^{st}(x) = H^s[\gamma(t)x + \delta(t)] = H[\gamma(s)\gamma(t)x + \gamma(s)\delta(t) + \delta(s)]$$

thus, we obtain

$$\gamma(st) = \gamma(s)\gamma(t), \quad \delta(st) = \gamma(s)\delta(t) + \delta(s) \tag{1.5}$$

The solutions of the functional equations (1.5) are (up to some constants):

$$\begin{cases} \gamma(t) &= 1 \\ \delta(t) &= -\ln t \end{cases} \qquad \begin{cases} \gamma(t) &= t^\xi \\ \delta(t) &= 0 \end{cases} \qquad \begin{cases} \gamma(t) &= a^t \\ \delta(t) &= 0 \end{cases} \tag{1.6}$$

Finally, plugging the above results into (1.4) leads to the three types Λ, Φ_α and Ψ_α.
Details of the proof are to be found in Resnick [18], Proposition 0.3. $\qquad\square$

The Central Limit Theorem (CLT) is one of the most important tools in probability theory and statistics, stating that the normal distribution is the only distribution of iid sums under certain conditions. As an analog, EVT states the three types (standard GEV) are the only possibilities for the limiting distribution of the maxima. We have a brief comparison of these two theorem in the following table:

Analog	CLT	EVT
Conditions on (X_n)	iid with finite second moment	iid and df is regularly varying
Study Object	S_n (Sum)	M_n (Maxima)
Limiting Distribution	Normal	GEV

TABLE 1.3 Analog between CLT and EVT.

11

For the next three examples, we simply provide simple distributions that lead to three types of limiting distribution in Theorem(1.1).

Example 1 Maxima for exponential distribution

Let (X_i) be a sequence of iid standard exponential random variables. So the cdf is $F(x) = 1 - e^{-x}$ and then

$$
\begin{aligned}
P(M_n - \ln n \le x) &= [F(x + \ln n)]^n \\
&= \left(1 - e^{-x - \ln n}\right)^n \\
&= \left(1 - \frac{e^{-x}}{n}\right)^n \\
&\to \exp\left\{-e^{-x}\right\} \\
&= \Lambda(x)
\end{aligned}
$$

For $x \in \mathbb{R}$. Here the normalizing constants $c_n = 1$, $d_n = \ln n$. □

Example 2 Maxima for Cauchy distribution

Let (X_i) be a sequence of iid standard Cauchy random variables. So the pdf is $f(x) = \dfrac{1}{\pi(1 + x^2)}$ and then

$$
F(x) = \int_{-\infty}^{x} \frac{1}{\pi(1 + t^2)} dt = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x)
$$

Apply l'Hopital's rule we obtain

$$
\bar{F}(nx/\pi) = 1 - F(nx/\pi) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1}(nx/\pi)
$$

$$
\lim_{x \to \infty} \frac{\bar{F}(nx/\pi)}{(nx)^{-1}} = \lim_{x \to \infty} \frac{-\dfrac{1}{\pi}\dfrac{n}{\pi}\dfrac{1}{(nx/\pi)^2 + 1}}{-\dfrac{1}{n}\dfrac{1}{x^2}} = \lim_{x \to \infty} \frac{n^2 x^2}{n^2 x^2 + \pi^2} = 1
$$

12

Hence

$$
\begin{aligned}
P(M_n \leq \frac{nx}{\pi}) &= \left[1 - \bar{F}(\frac{nx}{\pi})\right]^n \\
&= \left[1 - \frac{1}{n}\left(\frac{1}{x} + o(1)\right)\right]^n \\
&\rightarrow \exp\left\{-x^{-1}\right\} \\
&= \Phi_1(x)
\end{aligned}
$$

For $x > 0$. Here the normalizing constants are $c_n = \dfrac{n}{\pi}$, $d_n = 0$. $\qquad\square$

Example 3 Maxima for uniform distribution

Let (X_i) be a sequence of iid uniform random variables on $[0, 1]$. So the cdf is $F(x) = x$ for $0 \leq x \leq 1$ and then

$$
\begin{aligned}
P(n(M_n - 1) \leq x) &= \left(1 + \frac{x}{n}\right)^n \\
&\rightarrow \exp\left\{x\right\} \\
&= \Psi_1(x)
\end{aligned}
$$

For $x < 0$. Here the normalizing constants $c_n = \dfrac{1}{n}$, $d_n = 1$. $\qquad\square$

1.4   Generalized Extreme Value Distributions (GEV)

The three types of limiting distribution in section 1.3 are in standard form. We can parameterize them within the location and scale families.

1. (Gumbel)
$$
\Lambda(x) = \exp\left\{-\exp\left[-\left(\frac{x-d}{c}\right)\right]\right\}, x \in \mathbb{R}
$$

2. (Frechet)
$$
\Phi_\alpha(x) = \begin{cases} 0, & \text{if } x \leq d \\ \exp\{-(\frac{x-d}{c})^{-\alpha}\}, & \text{if } x > d \end{cases}
$$

13

3. (Weibull)

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-\frac{x-d}{c})^\alpha\}, & \text{if } x < d \\ 0, & \text{if } x \geq d \end{cases}$$
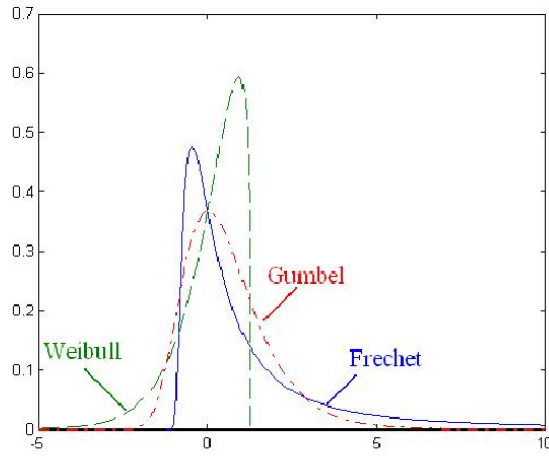
Some examples of extreme value distributions for
alpha = 1.25 (Frechet), 0 (Gumbel), -1.25 (Weibull)

FIGURE 1.1: Density Plots of Generalized Extreme Distributions

It is clear from Figure(1.1) that the three types of distributions have different support and tail behavior. The density of the Gumbel distribution decays exponentially as the decaying in algebraic order for the left tail in Weibull distribution and the right tail in Frechet distribution. It is advantageous to reformulate these three types into one family of distributions.

Lemma The generalized Gumbel, Frechet and Weibull families can be combined into a single family of distributions in the form

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad \text{where } 1 + \xi\left(\frac{x - \mu}{\sigma}\right) > 0 \qquad (1.7)$$

It is straightforward to check the result by letting

$$\alpha = \frac{1}{\xi}, \qquad d = \mu - \frac{\sigma}{\xi}, \qquad c = \begin{cases} \dfrac{\sigma}{\xi} & \text{when} \quad \xi > 0 \\ -\dfrac{\sigma}{\xi} & \text{when} \quad \xi < 0 \end{cases} \qquad (1.8)$$

$\square$

The model has three parameters: a location parameter $\mu \in \mathbb{R}$, a scale parameter $\sigma > 0$ and a shape parameter $\xi \in \mathbb{R}$. The generalized Frechet and Weibull distributions correspond respectively to the cases $\xi > 0$ and $\xi < 0$. When $\xi = 0$, we can take the limit of (1.7) as $\xi \to 0$ to obtain the following Gumbel family with distribution:

$$G(x) = \exp\left\{-\exp\left[-\left(\frac{x - \mu}{\sigma}\right)\right]\right\} \quad \text{where } x \in \mathbb{R} \qquad (1.9)$$

In general, the normalizing constants in Theorem(1.1) are difficult to find. This problem can be solved by the introduced GEV family of distributions. It is easy to see that if

$$Pr\{\frac{M_n - d_n}{c_n} \leq x\} \sim G(x)$$

for large enough n, then

$$Pr\{M_n \leq x\} \sim G(\frac{x - d_n}{c_n})$$

Since G is a member of GEV family, $G(\frac{x - d_n}{c_n})$ is also a member of GEV family with different location and scale parameters. In practice, the distribution of M_n itself

15

can also be approximated by a member of GEV family. Therefore the location and scale parameters μ, σ in the estimation depend on the data size n but not the shape parameter ξ. To emphasize it for the later statistical analysis, we state the above result as a theorem:

Theorem 1.3. *If there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$Pr\{\frac{M_n - d_n}{c_n} \leq x\} \to G(x) \quad as \ n \to \infty \tag{1.10}$$

for a non-degenerate distribution G, then G is a member of the GEV family as in equation(1.7) where $\mu, \xi \in \mathbb{R}$ and $\sigma > 0$. Moreover, the approximate distribution of M_n itself is a different member of the same family with the same shape parameter ξ.

$\square$

In practice, data are blocked into sequences of independent observations $X_1, X_2, ..., X_n$ for some large integer n, to generate a series of block maxima $M_{n,1}, M_{n,2}, ..., M_{n,m}$. The block maxima will be fitted to a GEV distribution and corresponding parameters will be estimated. A natural way of choosing the blocks is to use data from a certain time period so that seasonality effects can be neglected. For example, hourly observed sea level data can be blocked into monthly maxima and daily observed rainfall data can be blocked into annual maxima.

The method above has one disadvantage that it may throw away lots of useful information since only the block maxima are recorded. An alternative method is to use upper order statistics, where the EVT can be used as the fundamental tool to approximate the joint distributions of the k largest observations in a block. For reference books, David([8]) and Arnold([1]) are recommended.

16

1.5 Generalized Pareto Distributions (GPD)

As the GEV in the previous section describes the limit distribution of normalized maxima, the Generalized Pareto Distribution (GPD) is the limit distribution of scaled excess of high thresholds. The main connection is in the following theorem.

Theorem 1.4. *the Generalized Pareto Distribution (GPD). Suppose $X_1, X_2, ..., X_n$ are iid with distribution F. As in Theorem(1.3),*

$$G(x) = exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad where \; 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0$$

is the limit distribution of the maxima $M_n = X_{1,n} = max(X_1, ..., X_n)$. Then for a large enough threshold u, the conditional distribution function of $Y = (X - u | X > u)$, is approximately

$$Pr\{X - u < x | X > u\} \sim H(x) = 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi} \tag{1.11}$$

defined on $\{x : x > 0 \; and \; (1 + \xi x/\tilde{\sigma}) > 0\}$, where

$$\tilde{\sigma} = \sigma + \xi(u - \mu)$$

In case $\xi = 0$, the limit distribution should be interpreted as

$$H(x) = 1 - exp\left(1 - x/\tilde{\sigma}\right), \quad for \; x > 0$$

Sketch of the proof   First, notice that

$$Pr\{X - u > x | X > u\} = \frac{Pr\{X > u + x\}}{Pr\{X > u\}} = \frac{1 - F(u + x)}{1 - F(u)} \tag{1.12}$$

17

For large u, $F(u) \to 1$ and by Taylor expansion

$$\ln[F(u)] = \ln\{1 + [F(u) - 1]\} \approx -[1 - F(u)]$$

and since $Pr\{M_n < u\} = F^n(u) \approx G(u)$, we have

$$F^n(u) \approx \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad \text{where } 1 + \xi\left(\frac{u-\mu}{\sigma}\right) > 0$$

Taking $\ln(\cdot)$ on both sides gives the approximation of $\ln[F(u)]$

$$\ln[F(u)] \approx -\frac{1}{n}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}$$

Put it back into equation(1.12) so we have

$$Pr\{X - u > x | X > u\} \approx \frac{-\dfrac{1}{n}\left[1 + \xi\left(\dfrac{u+x-\mu}{\sigma}\right)\right]^{-1/\xi}}{-\dfrac{1}{n}\left[1 + \xi\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\xi}}$$

The result of the theorem comes from the simplification of the right hand side of the above equation. □

The family of distributions defined by equation(1.11) is called the **General Pareto family** (GPD). For a fixed high threshold u, the two parameters are the shape parameter ξ and the scale parameter $\tilde{\sigma}$. For simpler notation, we may just use σ for the scale parameter if there is no confusion. The GPD distribution has many good properties, see for example Embrechs [9], page 165. For the purpose of this paper, we particularly compute the expectation of the GPD.

Definition 1.1. *Let X be a random variable with distribution function $F(x)$.*

$$F_u(x) = Pr\{X - u \le x | X > u\}, \quad \text{for } x \ge 0$$

is the excess distribution of X over the threshold u. Define

$$e(u) = E(X - u | X > u) \tag{1.13}$$

and e(u) is called the **mean excess function** *of X.*

Suppose a random variable X follows GPD with parameters $\xi > 0$ and σ. Then

$$
\begin{aligned}
e(u) &= E(X - u | X > u) \\
&= \int_u^\infty \frac{x - u}{1 - H(u)} dH(x) \\
&= \frac{1}{1 - H(u)} \int_u^\infty -(x - u) \, d[1 - H(x)] \\
&= \frac{1}{1 - H(u)} \left\{ \lim_{x \to \infty} -x[1 - H(x)] + \int_u^\infty [1 - H(x)]dx \right\}
\end{aligned}
$$

where $H(x)$ is defined in equation(1.11), so $1 - H(x) = \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}$. The integration $\int_u^\infty -(x-u) \, d[1-H(x)]$ converges only when $\xi < 1$, where $\lim_{x \to \infty} -x[1-H(x)] = 0$. And when $0 < \xi < 1$,

$$\int_u^\infty [1 - H(x)]dx = \int_u^\infty \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi} dx = \frac{\sigma}{1 - \xi} \left(1 + \frac{\xi u}{\sigma}\right)^{-1/\xi + 1}$$

thus, the mean excess function of GPD is linear in u when $0 < \xi < 1$.

$$
e(u) = \begin{cases} \dfrac{\sigma + \xi u}{1 - \xi} & \text{for } 0 < \xi < 1 \\ \infty & \text{for } \xi \geq 1 \end{cases}
$$

At the end of this section, we present the three examples in Section 1.4 in terms of the excess of threshold model with GPD. In other words, we will compute the limit distribution function $Pr\{X - u > x | X > u\} = \dfrac{1 - F(u + x)}{1 - F(u)}$ for each example.

Example 1 For the standard exponential model, $F(x) = 1 - e^{-x}$ for $x > 0$. Then

$$\frac{1 - F(u + x)}{1 - F(u)} = \frac{e^{-(u+x)}}{e^{-x}} = e^{-u}$$

19

for all $x > 0$. The distribution of exceedances over excess itself is exponential distribution, corresponding to $\xi = 0$ and $\sigma = 1$ in the General Pareto family. $\square$

Example 2 For the standard Frechet model, $F(x) = e^{-1/x}$ for $x > 0$. Then

$$\frac{1 - F(u + x)}{1 - F(u)} = \frac{1 - e^{-1/(u+x)}}{1 - e^{-1/u}} \approx \frac{1/(u + x)}{1/u} = \left(1 + \frac{x}{u}\right)^{-1} \quad \text{as } u \to \infty$$

This corresponds to the case of $\xi = 1$ and $\sigma = u$ in the General Pareto family. $\square$

Example 3 For the uniform distribution model, $F(x) = x$ for $0 \le x \le 0$. Then

$$\frac{1 - F(u + x)}{1 - F(u)} = \frac{1 - (u + x)}{1 - u} = 1 + \frac{(-1)\, x}{1 - u}$$

This corresponds to $\xi = -1$ and $\sigma = 1 - u$ in the General Pareto family. $\square$

CHAPTER 2

STATISTICAL METHODS IN APPLICATION OF EVT

2.1 Exploratory Data Analysis

In the real world, the extreme value theory as we described in chapter 1 needs to be applied through statistical data: the observed sea level, major insurance claims, large variation of security market values over a certain time period, daily records of temperature and precipitation at a certain location, etc. We hope the modeling of the empirical data through extremes would manifest most of these variations. Appropriate models can be used to manage financial risks, to set up prevention procedure or to obtain estimations and predictions.

Before getting into any statistical analysis, we want to learn our best knowledge of the data set: How is the data collected? Is there any missing or unreported data? Since it is particularly important that the data is error-free for extreme observations, we should give special attentions to data that are outliers.

The first step in exploratory analysis is to have a scatter plot of all observations. It is important just by looking at data to obtain the following information: range of data, several extremes, trends and seasonalities, any violation of independence and stationarity conditions. If trends and seasonalities are suspected, we recommend to

fit the location, scale and shape parameters with a time variable t. See for example, chapter 6 in Coles [6] and references therein.

Embrechts, Kluppelberg and Mikosch's book [9] and its references give a thorough list of both graphical and analytical methods, and in Bassi [3], a survival kit of some terminologies on quantile estimation is provided. We have introduced the mean excess function in section 1.5, which is crucial in the later analysis on the determination of choice of transforms. We will emphasize two other methods used in this paper: QQ-plots, the return period and return levels.

Probability and Quantile Plots (QQ-plot)

Suppose $X_1, X_2, ..., X_n$ are continuous random variables and $x_1, x_2, ..., x_n$ are independent observations from a common population with unknown distribution F. An estimate of F, say $\hat{F}$, has been obtained. The probability and quantile plots provide a graphical accessment to the fitted distribution $\hat{F}$. It follows from the Quantile Transformation Lemma ([9] Lemma 4.1.9) that $F(X_i)$ has a uniform distribution on (0,1) for $i = 1, ..., n$. Furthermore, if $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ are the ordered sample, then the expectation of these quantiles can be computed as:

$$E[F(X_{(i)})] = \frac{i}{n+1}, \quad \text{for } i = 1, ..., n$$

and this leads to the following definition.

Definition 2.1. *Given an ordered sample of independent observations*

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

22

from a population with estimated distribution function $\hat{F}$, a graph

$$\left\{ \left(\hat{F}(x_{(i)}), \frac{i}{n+1} \right) : \ i = 1, ..., n \right\}$$

is called a **probability plot (PP-plot)**. □

If $\hat{F}$ is a reasonable estimation of the population distribution, the points on the probability plots should lie close to the unit diagonal. Substantial deviation from linearity provides evidence of the lack of fit of $\hat{F}$ to the model. This plot can be modified by using the data values as the x-values of the points. The following is more commonly used:

Definition 2.2. *Given an ordered sample of independent observations*

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

from a population with estimated distribution function $\hat{F}$, a graph

$$\left\{ \left(\hat{F}^{-1}(\frac{i}{n+1}), x_{(i)} \right) : \ i = 1, ..., n \right\}$$

is called a **quantile plot (QQ-plot)**. □

The quantities $x_{(i)}$ is the empirical $\dfrac{i}{n+1}$ quantile of the population distribution F while $\hat{F}^{-1}(\dfrac{i}{n+1})$ is the estimation. If $\hat{F}$ is a reasonable estimation of F, the quantile plot should look roughly linear. This remains true if the data come from a linear transformation of the distribution. And since that, the change of location and scale parameters only change the plot through y-intercept and slope.

Outliers can be easily identified on the QQ-plot in a general statistical analysis. While the subject of the extreme value study concerns the upper tail, we should be

particularly cautious about any point that substantially deviates from the model on the large observation end. Since the shape parameter ξ determines how heavy the tail distribution is, some difference in distributional shape may be deduced from the plot. In general, an overestimation of ξ (heavy tail) will result in a concave down curve in the QQ-plot, and an underestimation of ξ (light tail) will result in a concave up curve in the QQ-plot.

The Return Period and The Return Level

The return period and the return level are very silimar to the t-year event we mentioned in section 1.1. We make the definition precise as below:

Definition 2.3. *Let* (X_i) *be a sequence of iid random variables with continuous distribution* F *and* u *a given threshold. Consider the sequence* $(I_{(X_i>u)})$ *of iid Bernoulli random variables with success probability* $p = 1 - F(u)$*. Then*

$$E[L(u)] = \frac{1}{p}$$

is called the **return period** *of the event* $(X_i > u)$*, where* $L(u) = min\,\{i \geq 1 : X_i > u\}$ *is the time of first exceedance of the threshold* u*.* $L(u)$ *follows a geometric distribution with parameter* p*.* $\square$

From the definition, return period is the expected revisiting period corresponding to the threshold u. This leads to the other definition:

Definition 2.4. *Let* (X_i) *be a sequence of iid random variables with continuous distribution* F*. if*

$$F(z_p) = 1 - p \tag{2.1}$$

24

then z_p is called the **return level** *associated with the return period* $\dfrac{1}{p}$. $\qquad\square$

For the sea dikes example in chapter 1, in the above language, one would say that the 1953 sea level corresponds to a return period of 382 years. The expectation from the government is a dike level that can sustain a 1,000 or even 10,000 year event. And it is clear from this example that the solution of extremal events needs to be able to make inferences about very high quantiles.

Combining equation 1.7 and 2.1, we can solve the return level z_q in terms of parameters in the model and p.

$$z_p = \mu - \frac{\sigma}{\xi}\left\{1 - [-\ln(1-p)]^{-\xi}\right\} \quad \text{for } \xi \neq 0$$

In the similar way, combining 1.9 and 2.1 in case $\xi = 0$, we obtain:

$$z_p = \mu - \sigma \ln[-\ln(1-p)] \quad \text{for } \xi = 0$$

Define $y_p = -\ln(1-p)$, we have

$$z_p = \begin{cases} \mu - \dfrac{\sigma}{\xi}[1 - y_p^{-\xi}] & \text{for} \quad \xi \neq 0 \\[2mm] \mu - \sigma \ln y_p & \text{for} \quad \xi = 0 \end{cases} \tag{2.2}$$

The plot of z_p against $-\ln y_p$ is a referred as **return level plot**. When $\xi = 0$, the plot is linear with μ and σ as y-intercept and slope. If $\xi > 0$, the plot is convex with asymptotic limit $\mu - \dfrac{\sigma}{\xi}$ as $p \to 0$; if $\xi > 0$, the plot is concave and unbounded as $p \to 0$. The return level plots are convenient for both model presentation and validation, for its simple interpretation and highlighted extrapolation of the tail distribution under the compressed scale.

Often the number of blocks n is chosen to correspond to a time period of length one year, then the block maxima is annual maxima and the return period is in years.

2.2 Parameter Estimation for the GEV

Once the model has been set up, parameters in the model need to be estimated using appropriate procedure(s). Suppose $X_1, X_2, ..., X_n$ is a sequence of random variables of iid GEV distribution with parameters ξ, μ and $\sigma > 0$, and $x_1, x_2, ..., x_n$ are the recorded observations. As in section 1.4, the distribution function $G(x)$ is

$$G(x) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{when } \xi \neq 0, \ 1+\xi\left(\dfrac{x-\mu}{\sigma}\right) > 0 \\ \exp\left\{-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right\} & \text{when } \xi = 0, \ x \in \mathbb{R} \end{cases}$$

In this section, we introduce two methods: maximum likelihood method and method of probability-weighted moments. For Bayesian estimations, Coles and Powell [7] give a general review.

Maximum Likelihood Method The method of maximum likelihood is, by far, the most popular technique for deriving estimators.

Definition 2.5. *Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X} = (X_1, X_2, ..., X_n)$. Given that $\mathbf{X} = \mathbf{x} = (x_1, x_2, ..., x_n)$ is observed, the function of θ defined by*

$$L(\theta|\mathbf{X}) = f(\mathbf{X}|\theta)$$

is called the **likelihood function***, and*

$$l(\theta|\mathbf{X}) = \ln L(\theta|\mathbf{X})$$

defines the **log-likelihood function***. The* **maximum likelihood estimator (MLE)** *$\hat{\theta}(\mathbf{X})$ is the function of $\mathbf{X}$ that maximizes $L(\theta|\mathbf{X})$ or $l(\theta|\mathbf{X})$.* $\qquad\square$

As an example, we consider the Gumbel family where $\xi = 0$.

$$L((0; \mu, \sigma)|\mathbf{x}) = \exp\left[-\sum_{i=1}^{n} \exp\left(-\frac{x_i - \mu}{\sigma}\right)\right] \cdot \exp\left[-\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)\right] \cdot \frac{1}{\sigma^n}$$

Thus, the log-likelihood is

$$l((0; \mu, \sigma)|\mathbf{x}) = -\sum_{i=1}^{n}\left[\exp\left(-\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right) - n\ln\sigma$$

One approach to locate the maximum is to differentiate the above equation with respect to μ and σ and set them to zero. This leads to the following likelihood equations:

$$\begin{cases} -\dfrac{1}{\sigma}\sum_{i=1}^{n}\left[\exp\left(-\dfrac{x_i - \mu}{\sigma}\right)\right] + \dfrac{n}{\sigma} & = 0 \\[4mm] \sum_{i=1}^{n}\left\{\dfrac{x_i - \mu}{\sigma^2}\left[1 - \exp\left(-\dfrac{x_i - \mu}{\sigma}\right)\right]\right\} - \dfrac{n}{\sigma} & = 0 \end{cases} \tag{2.3}$$

There exists no explicit solution to equations(2.3). Numerical procedures to find maximums are called for the approximation. The likelihood equations for the case $\xi \neq 0$ is more complicated. For the stability and asymptotic properties of the MLE, it is shown in Smith [19] that the classical properties of the MLE hold whenever $\xi > -\dfrac{1}{2}$; but not for $\xi \leq -\dfrac{1}{2}$ in general.

$\square$

Method of Probability-Weighted Moments The technique of moments method is to equate the model-moments with empirical moments. Even though the general properties of obtained estimators can be unreliable, the method of moments can be very useful in obtaining an initial approximation of the intended statistics. The class of probability-weighted moments stands out to be more promising in obtaining a

good first estimate. The results are used as the initial values for other methods when numerical techniques are applied. Let $\theta = (\xi; \mu, \sigma)$ and define

$$w_r(\theta) = E\left[XG_\theta^r(X)\right] \quad \text{for } r = 0, 1, 2, ...$$

In case $\xi < 1$, calculation yields

$$w_r(\theta) = \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\xi}[1 - \Gamma(1-\xi)(1+r)^\xi]\right\} \tag{2.4}$$

where Γ denotes the Gamma function $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ for $t > 0$. Choosing $r = 0, 1, 2$, we can then immediately obtain

$$\begin{cases} w_0(\theta) &= \mu - \dfrac{\sigma}{\xi}[1 - \Gamma(1-\xi)] \\[2mm] 2w_1(\theta) - w_0(\theta) &= \dfrac{\sigma}{\xi}\Gamma(1-\xi)(2^\xi - 1) \\[2mm] 3w_2(\theta) - w_0(\theta) &= \dfrac{\sigma}{\xi}\Gamma(1-\xi)(3^\xi - 1) \end{cases} \tag{2.5}$$

$(\xi; \mu, \sigma)$ can be explicitly solved from the above system (2.5). For example

$$\frac{2^\xi - 1}{3^\xi - 1} = \frac{2w_1(\theta) - w_0(\theta)}{3w_2(\theta) - w_0(\theta)} \tag{2.6}$$

Parameter estimation is obtained by replacing the model-moments $w_r(\theta)$ in (2.6) by empirical moments $\hat{w}_r(\theta)$. To obtain empirical moments, notice

$$\hat{w}_r(\theta) = \frac{1}{n}\sum_{j=1}^n X_{(j)}G_\theta^r(X_{(j)}), \quad \text{for } r = 0, 1, 2 \tag{2.7}$$

where $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are order sample. Again it follows from the Quantile Transformation Lemma ([9] Lemma 4.1.9) that

$$(G_\theta(X_{(1)}), ..., G_\theta(X_{(n)})) \overset{d}{=} (U_{(1)}, ..., U_{(n)})$$

where $U_{(1)}, U_{(2)}, ..., U_{(n)}$ are the order statistics of an iid sequence $U_1, U_2, ..., U_n$ uniformly distributed on $(0, 1)$. Thus, 2.7 can be rewritten as:

$$\hat{w}_r(\theta) = \frac{1}{n} \sum_{j=1}^{n} X_{(j)} U_{(j)}^r, \quad \text{for } r = 0, 1, 2 \tag{2.8}$$

where $U_{(j)}^r$ are often approximated by their expectations. $\qquad\qquad\square$

2.3 The Excess Over Threshold Method

Model framework and parameter estimation

The modeling using the excess over threshold follows the assumptions and conclusions in Theorem(1.4). Suppose $x_1, x_2, ..., x_n$ are raw observations independently from a common distribution F. Given a high threshold u, assume $x_{(1)}, x_{(2)}, ..., x_{(k)}$ are observations that exceeds u. Here we abuse our notation and define the exceedances as $x_i = x_{(i)} - u$ for $i = 1, 2, ..., k$. By Theorem(1.4), x_i may be regarded as realization of independently random variable which follows a General Pareto family with unknown parameters ξ and σ. General parameter estimation procedure can be applied to obtain references, followed by goodness of fit procedure and further extrapolation.

As we applied two parameter estimation methods to GEV in Section 2.2, we use both two methods in the parameter estimation of GPD model. In case $\xi \neq 0$, the likelihood function can be obtained directly from (1.11):

$$L((\xi, \sigma)|\mathbf{x}) = \prod_{i=1}^{k} \left[\frac{1}{\sigma} \left(1 + \frac{\xi x_i}{\sigma} \right)^{-1/\xi - 1} \right]$$

The log-likelihood function is

$$l((\xi, \sigma)|\mathbf{x}) = -k \ln \sigma - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^{k} \ln \left(1 + \frac{\xi x_i}{\sigma} \right)$$

29

Taking partial derivatives of $l((\xi, \sigma)|\mathbf{x})$ with respect to ξ and σ, we obtain

$$
\begin{cases}
\dfrac{1}{\xi^2} \sum_{i=1}^{k} \ln\left(1 + \dfrac{\xi x_i}{\sigma}\right) - \sum_{i=1}^{k} \dfrac{x_i}{\sigma + \xi x_i} & = \ 0 \\[3mm]
-k + (1+\xi) \sum_{i=1}^{n} \dfrac{x_i}{\sigma + \xi x_i} & = \ 0
\end{cases}
\tag{2.9}
$$

We can solve (2.9) through a reparametrization $(\xi, \sigma) \rightarrow (\xi, \lambda)$ where $\lambda = \dfrac{\xi}{\sigma}$. The solution is determined through

$$
\xi = \frac{\displaystyle\sum_{i=1}^{k} \ln(1 + \lambda x_i)}{\displaystyle\sum_{i=1}^{k} \frac{\lambda x_i}{1 + \lambda x_i}} \quad \text{and} \quad \xi = \frac{\displaystyle\sum_{i=1}^{k} \frac{\lambda x_i}{1 + \lambda x_i}}{\displaystyle\sum_{i=1}^{k} \frac{1}{1 + \lambda x_i}}
$$

Again the MLE properties like consistency and asymptotic efficiency hold if $\xi > -\dfrac{1}{2}$.

Assume $H_\theta(x)$ is the GPD distribution where $\theta = (\xi, \sigma)$. Let $\hat{H}_\theta(x) = 1 - H_\theta(x)$ and define the probability-weighted moment

$$
w_r(\theta) = E\left[X\hat{H}_\theta^r(X)\right] \quad \text{for } r = 0, 1, 2, \ldots
$$

Direct calculation leads to

$$
w_0 = \frac{\sigma}{1 - \xi} \quad \text{and} \quad w_1 = \frac{\sigma}{4 - 2\xi}
$$

and we obtain the solution

$$
\xi = \frac{w_0 - 4w_1}{w_0 - 2w_1} \quad \text{and} \quad \sigma = \frac{2w_0 w_1}{w_0 - 2w_1}
$$

And finally replace w_0 and w_1 by empirical moments to obtain estimation. $\qquad \square$

Threshold selection As we vary threshold u from low to high, there is an important trade-off between bias and variance. If u is chosen small, the asymptotic basis of Theorem(1.4) is violated and there may be systematic bias. A very large choice of u would directly decrease the number of exceedance so the variance of estimation is high. The standard practice is to adopt as low a threshold as possible such that the limiting model still provides a reasonable estimation. Two methods are available for this purpose: one is through an explanatory plot (mean residual life plot) prior to model estimation; the other is through the analysis of the stability of model estimation, based on the fitting of models across a range of different thresholds.

As we discussed in Section 1.5, the mean excess function of GPD

$$e(u) = \frac{\sigma + \xi u}{1 - \xi} \quad \text{for } 0 < \xi < 1$$

is linear in u. Therefore we can check the linearity in the plot described blow:

Definition 2.6. *Given a sample of iid observations $x_1, x_2, ..., x_n$ and a threshold u. Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n_u)}$ be the observations that exceeds u, a graph*

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

is called a **mean residual life plot (mrl)**. $\qquad\qquad$ □

The mean residual life plot provides an accessible approximation to the mean excess function. As we will discuss in the next section, confidence intervals can be added to the plot based on the approximate normality of sample means.

In practice, the interpretation of a mean residual life plot may not be simple. Often the linearity is vague for small choice of u and for large u, the sparseness of

the data available for calculation causes the large variation of the plot toward the right end. For our purposes, the mlr plot is used as a graphical tool in distinguishing between light- and heavy-tail models.

A complementary technique is to fit the generalized Pareto distribution at a range of thresholds, and to look for stability of parameter estimates. Suppose the conditions in Theorem(1.4) are satisfied, then for threshold u which is large enough, the shape parameter is nearly constant. And in the estimation

$$\tilde{\sigma} = \sigma + \xi(u - \mu)$$

if we reparametrize $\sigma^* = \tilde{\sigma} - \xi u$, then the estimation of σ^* should be also nearly constant if the threshold u is large enough that the Theorem is approximately valid.

This argument suggests that by plotting estimations $\hat{\xi}$ and $\hat{\sigma}^*$ against a range of u, we can select the threshold as the lowest value of u such that these estimates remain near-constant. Again, confidence intervals for each estimate can be combined into the plot after finding approximate normality means. $\qquad\square$

The return level for a N-year period

In the approach of the excess over threshold method, the number of exceedances over a time period follows a binomial distribution for a chosen threshold u. Suppose that a generalized Pareto distribution with parameters ξ and σ is a suitable model, then for $x > 0$,

$$Pr\{X > x | X > u\} = \left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}$$

Let $p_u = Pr\{X > u\}$, then the **m-observation return level** x_m is the solution of

$$Pr\{X > x_m\} = \frac{1}{m}$$

use the fact that $Pr\{X > x_m\} = Pr\{X > x_m | X > u\} \cdot Pr\{X > u\}$, we obtain

$$\left[1 + \xi \left(\frac{x_m - \mu}{\sigma}\right)\right]^{-1/\xi} \cdot p_u = \frac{1}{m}$$

or explicitly

$$x_m = \mu + \frac{\sigma}{\xi}\left[(mp_u)^\xi - 1\right] \tag{2.10}$$

provided $\xi \neq 0$ and m is large enough to ensure $x_m > u$. For the case $\xi = 0$, we can obtain an explicit return level in the same way:

$$x_m = \mu + \sigma \ln(mp_u)$$

Suppose we have approximately n_y observations annually, then for a N year period, the number of observations that exceeds the threshold u follow a binomial distribution with parameter $N \cdot n_y$ and p_u. Thus the corresponding return level

$$x_N = \begin{cases} \mu + \dfrac{\sigma}{\xi}\left[(N \cdot n_y \cdot p_u)^\xi - 1\right] & \text{for} \quad \xi \neq 0 \\ \mu + \sigma \ln(N \cdot n_y \cdot p_u) & \text{for} \quad \xi = 0 \end{cases} \tag{2.11}$$

$\square$

2.4 Approximate Normality of the MLE and Profile Likelihood

Various standard and widely applicable approximations are available for many useful sampling distributions if the maximum likelihood method is used for parameter estimations. For references, Azzalini [2] and Casella & Berger [4] provide the basics of statistics inference and modeling in these aspects. Here, we briefly list the major

results without introducing any detailed proofs. Each of the results is an asymptotic law which becomes more accurate as the sample size n increases to infinity. And they are valid under certain regularity conditions. We will assume these conditions to be satisfied (or nearly satisfied).

Theorem 2.1. *Let $x_1, x_2, ..., x_n$ be independent observations from a common distribution within a parametric family $\mathcal{F}$. Let $l(\theta|\mathbf{x})$ and θ_0 denote the log-likelihood function and the MLE of the d-dimensional model parameter θ_0. Then, under suitable regularity conditions, for large n*

$$\hat{\theta}_0 \overset{d}{\sim} MVN_d(\theta_0, I_E(\theta_0)^{-1})$$

where MVN_d denotes the d-dimensional multivariate normal distribution and

$$I_E(\theta) = \left\{ E\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\theta) \right] \right\}_{1 \leq i,j \leq d}$$

is a $d \times d$ matrix which measures the expected curvature of the log-likelihood surface.

□

Theorem(2.1) can be used to obtain approximate confidence intervals for individual components of $\theta_0 = (\theta_1, ..., \theta_d)$. Denoting $I_E(\theta_0) = \{\psi_{i,j}\}$, it follows from the properties of the multivariate normal distribution that for large n,

$$\hat{\theta}_i \overset{d}{\sim} N(\theta_i, \psi_{i,i})$$

Since the true value of θ_0 is generally unknown, it is usual to approximate the terms of $I_E(\theta)$ with those

$$I_O(\theta) = \left\{ -\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\hat{\theta}) \right\}_{1 \leq i,j \leq d}$$

34

where $\hat{\theta}$ is the MLE estimate of θ. Again, denote $I_O(\theta_0) = \left\{ \hat{\psi}_{i,j} \right\}$, then an approximate $(1 - \alpha)$ confidence interval for θ_i is

$$\hat{\theta}_i \pm z_{\frac{\alpha}{2}} \sqrt{\hat{\psi}_{i,i}}$$

$I_E(\theta)$ and $I_O(\theta)$ are usually referred to as the **expected information matrix** and **observed information matrix** respectively.

Perhaps one of the most useful properties of maximum likelihood estimators is what has come to be known as the invariance property. Suppose that a distribution is parameterized by θ, but our interest is in finding inferences for some function of θ, say $\tau(\theta)$, where $\tau(\theta)$ may have a different dimension to θ. We state the invariance property as a theorem:

Theorem 2.2. Invariance Property of MLEs *If $\hat{\theta}$ is the MLE of θ, then for any function $\tau(\theta)$, $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.* □

It is easy to verify the result in case $\tau(\theta)$ is a one-to-one function. For more general function of $\tau(\theta)$, yet the above theorem holds only for the *induced likelihood function L^*, given by*

$$L^*(\eta|\mathbf{x}) = \sup_{\{\theta : \tau(\theta) = \eta\}} L(\theta|\mathbf{x})$$

The value $\hat{\eta}$ which maximizes $L^*(\eta|\mathbf{x})$ will be called the MLE of $\eta = \tau(\theta)$.

Theorem 2.3. *Let $\hat{\theta}_0$ be the large-sample MLE of the d-dimensional parameter θ_0 with approximate variance-covariance matrix V_θ. Then if $\phi = g(\theta)$ is a scalar function, the maximum likelihood estimator of $\phi_0 = g(\theta_0)$ satisfies*

$$\hat{\phi}_0 \overset{d}{\sim} N(\phi_0, V_\phi)$$

35

where $V_\phi = \nabla \phi^T V_\theta \nabla \phi$ and $\nabla \phi = \left[\dfrac{\partial \phi}{\partial \theta_1}, ..., \dfrac{\partial \phi}{\partial \theta_d} \right]^T$ evaluated at $\hat{\theta}_0$. $\qquad\qquad\square$

Theorem (2.3) is usually referred to as the **delta method**, which enables the approximate normality of ϕ_0 to be used in obtaining confidence intervals for ϕ. As an alternative way of making inferences on a parameter component θ_i of a parameter vector θ, a profile log-likelihood method often provide unsymmetric, yet more accurate results. The log-likelihood for θ can be written as $l(\theta_i, \theta_i^c)$, where θ_i^c denotes all components of θ excluding θ_i. The **profile log-likelihood** for θ_i is defined as

$$l_p(\theta_i) = \max_{\theta_i^c} \ l(\theta_i, \theta_i^c)$$

From the definition, given a fixed θ_i, the profile log-likelihood is the maximized log-likelihood with respect to all other components of θ. In other words, $l_p(\theta_i)$ is the profile of the log-likelihood surface view from the θ_i axis. A profile log-likelihood plot consists points of $(l_p(\theta_i), \theta_i)$ for a range of θ_i and a confidence region described below.

A generalized situation to the profile likelihood is to partition θ into two parts, $(\theta^{(1)}, \theta^{(2)})$, of which $\theta^{(1)}$ is the k-dimensional vector of interest and $\theta^{(2)}$ corresponds to the remaining $(d-k)$ components. the profile log-likelihood for $\theta^{(1)}$ is defined similarly,

$$l_p(\theta^{(1)}) = \max_{\theta^{(1)}} \ l(\theta^{(1)}, \theta^{(2)})$$

The previous definion is the case when $k = 1$.

Theorem 2.4. *Let $x_1, x_2, ..., x_n$ be independent observations from a common distribution within a parametric family $\mathcal{F}$. Let $\hat{\theta}_0$ denote the MLE of the d-dimensional*

model parameter $\theta_0 = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(1)}$ is a k-dimensional subset of θ_0. Then, under suitable regularity conditions, for large n

$$D_p(\theta^{(1)}) \triangleq 2\{l(\hat{\theta}_0) - l_p(\theta^{(1)})\} \overset{d}{\sim} \chi_k^2$$

□

$D_p(\theta^{(1)})$ is usually referred to as the **deviance statistic** and χ_k^2 is the known chi-square distribution with index k. Applying Theorem 2.4 directly to the profile likelihood, we can obtain a confidence interval for θ_i.

Corollary 2.1. *Suppose conditions in Theorem 2.4 are all satisfied and $k = 1$, $\theta^{(1)} = \theta_i$, $\theta^{(2)} = \theta_i^c$ then,*

$$C_\alpha = \{\theta_i : D_p(\theta_i) \le c_\alpha\}$$

is a $(1-\alpha)$ confidence interval, where c_α is the $(1-\alpha)$ quantile of the χ_1^2 distribution.

□

Another application of Theorem 2.4 is to model selection or model reduction. Suppose our interest is whether some factors should be introduced in explaining the variations of the data. The procedure is to obtain the maximized likelihood for both models with and without the intended factors. The magnitude of the improvement in the maximized log-likelihood when the new factors are added determines the acceptance or rejection of our hypothesis.

Corollary 2.2. *Suppose $\mathcal{M}_0$ with parameter $\theta^{(2)}$ is the sub-model of model $\mathcal{M}_1$ with parameter $\theta_0 = (\theta^{(1)}, \theta^{(2)})$ under the constraint that the k-dimensional sub-vector $\theta^{(1)} = \mathbf{c}$, where $\mathbf{c}$ is known. Then*

$$D \triangleq 2\{l_1(\mathcal{M}_1) - l_0(\mathcal{M}_0)\} \stackrel{d}{\sim} \chi_k^2$$

where $l_0(\mathcal{M}_0)$ and $l_1(\mathcal{M}_1)$ are the maximized log-likelihood values for model $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively. A hypothesis test of the validity of model $\mathcal{M}_0$ relative to $\mathcal{M}_1$ at a level of significance is to reject $\mathcal{M}_0$ in favor of $\mathcal{M}_1$ if $D > c_\alpha$, where c_α is the $(1 - \alpha)$ quantile of the χ_k^2 distribution. □

CHAPTER 3

CASE STUDY: HURRICANE DAMAGES OVER THE
LAST CENTURY

3.1 Data Resource and Information

As discussed in chapter 1, we need to work on the raw data so that the total damage
(in millions) reflects the impact of the corresponding hurricane if it happened today,
and thus make the estimation and prediction meaningful. As in Pielke [17], it is as-
sumed that losses are proportional to three factors: inflation, wealth and population.
The result of normalizing the data is to produce the estimated impact of any storm
as if it had made land-falling in 1995. Before we get into detailed analysis of the data,
let's take a look at some descriptive statistics of the data and the scatter plot:

Statistics	Observed Data
Maximum	$72.3 Billion
Mean	$6.7 Billion
the Ratio of Maximum and Sum	0.04
$10B Return Period	9 years (approx)

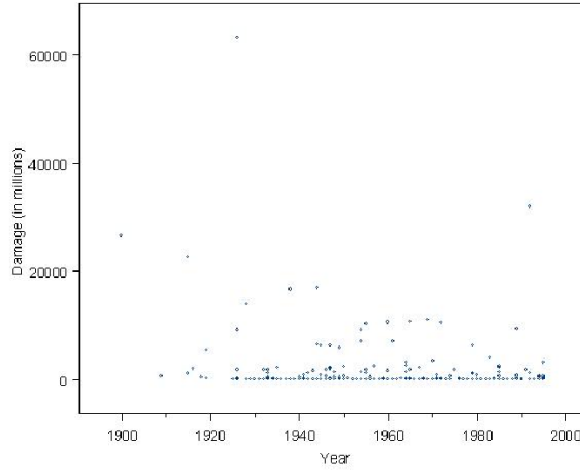TABLE 3.1 Some descriptive statistics of hurricane data.

FIGURE 3.1: Scatter plot of the hurricane data

There are 253 normalized data in total, range from year 1900 to 1995. In the data from year 1900 to year 1925, all hurricanes that have small damages amount were omitted from the records. Since the omission of small amount of data does not affect the result when excess over threshold method is used, we will keep data from this period in our study.

The plot of the data does not indicate any serious violation of the independence and stationarity assumptions. However, there is seemingly an over normalization for data before year 1940. While the costliest hurricane (normalized to year 1995) happened in year 1926 with an actual damage of \$105M and a normalized damage of \$72.3B, hurricane Andrew (1992) ranks second with a normalized loss of \$33.1B. It takes on average 9 years to observe a hurricane with a normalized damage over \$10B.

40

3.2 Fitting Data by the EOT Method: Threshold Selection

Since the number of observations is very limited for each year, we use the EOT method for the analysis of the data. It follows from section 2.4 that the excesses follows approximately a generalized Pareto distribution, thus the mean excess function is linear in threshold u. The mean residual life plot in figure(3.2) looks like a straight line except at the very high end. We can also try to estimate ξ from the slope of the line and set it equal to $\dfrac{\xi}{1-\xi}$. For example, if slope 1 is observed from the plot, we can have a very rough estimation for ξ as 0.5.
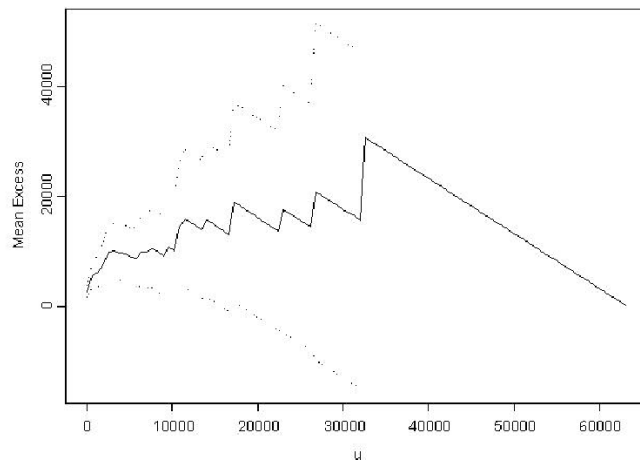

FIGURE 3.2: Mean residual plot of hurricane data

As an alternative method mentioned in section 2.4, the selection of a threshold can be carried out more carefully through a stability check of MLEs of parameters. Suppose the CAT future defines any total damage loss over $500 million as a catastrophe

41

event, we are particularly interested in finding out if \$500 million is an appropriate threshold. Figure(3.3) are the MLEs of parameters $\tilde{\sigma}$ and ξ together with a 95% confidence interval of the estimates. We can directly tell from these plots that when $u = 500$, both parameter estimations are nearly constant. So the rest of our analysis uses the threshold $u = 500M$.
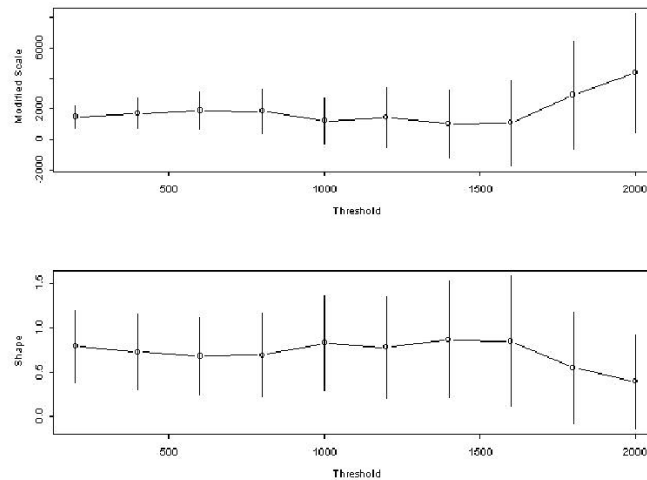

FIGURE 3.3: Parameter estimations for different thresholds

Usually after the threshold is taken, we generate another scatter plot consisting only the excesses of the data and recheck the assumptions in Theorem(1.4). In Coles [6] (page 90), the excesses of transformed Dow Jones Index series demonstrate the tendency that extreme events are clustered. In this example, the plot does not indicate any serious violation of those assumptions.

42

3.3 Statistical Analysis of S-Plus Plots

Figure(3.4) gives the diagnostic plots for the fitted GPD model of excesses data over threshold $u = 500$. The probability plot and the quantile plot show a good fit of the model, especially the higher end fit in the quantile plot. The return level plot and the density plot can be hardly distinguished since the large scale in the data. This can be improved by a log-log plot of the results.
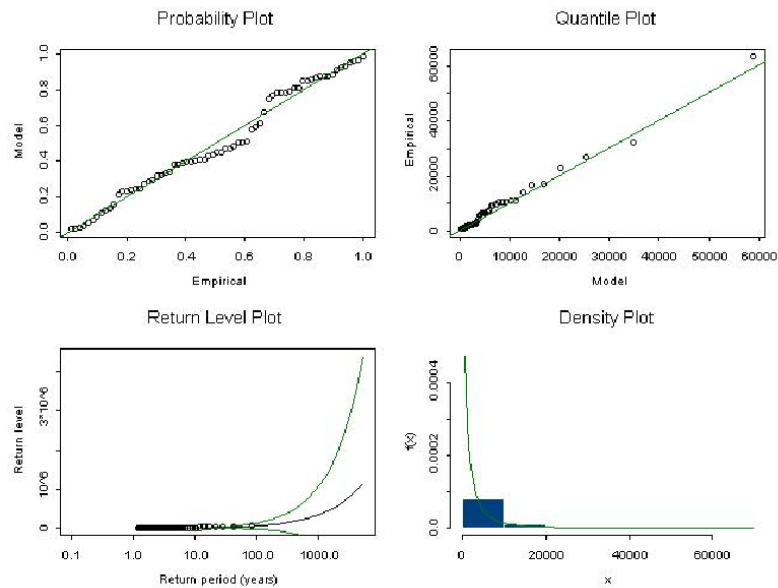

FIGURE 3.4: Goodness-of-fit Plots for hurricane data, u = 500

As a comparison, we present both the estimates by the MLE method and by the profile log-likelihood method. Maximum likelihood estimation for parameters are:

$$(\hat{\sigma}, \hat{\xi}) = (2108.4, 0.71479)$$

43

with a corresponding maximized log-likelihood of -614.82. The variance-covariance
matrix is calculated as

$$\begin{bmatrix} 247167.6 & -64.83216 \\ -64.83216 & 0.0485135 \end{bmatrix}$$

leading to standard error of 497.16 and 0.22026 for $\hat{\sigma}$ and $\hat{\xi}$ respectively. In particular,
it follows that a 95% confidence interval for ξ is obtained as $0.71479 \pm 1.96 \times 0.22026 = $
$[0.2831, 1.1465]$. Since the 95% interval for ξ does not include 0, there is strong
evidence that $\xi > 0$ with the exact p-value less than 0.001.

As a comparison, the profile log-likelihood plot, figure(3.5), suggests non-symmetric
confidence intervals for the shape parameter ξ. A 95% confidence interval for ξ can
be obtained from the plot as $[0.35, 1.25]$. This is not so different from the previous
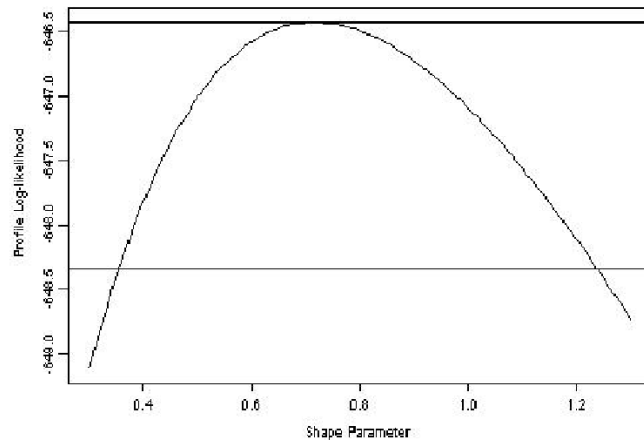$[0.2831, 1.1465]$ estimation, but strengthens the conclusion that $\xi > 0$.

FIGURE 3.5: Profile log-likelihood estimate of shape parameter ξ

There are 69 exceedances of the threshold $u = 500$ in the complete set of 253 observations, the maximum likelihood estimate of the exceedance probability is $\hat{p_u} = \frac{69}{253} = 0.27273$. Therefore the approximate variance is about $\frac{\hat{p_u}(1 - \hat{p_u})}{253} = 0.000784$. The complete variance-covariance matrix for $(\hat{p_u}, \hat{\sigma}, \hat{\xi})$ is

$$V = \begin{bmatrix} 0.000784 & 0 & 0 \\ 0 & 247167.6 & -64.83216 \\ 0 & -64.83216 & 0.0485135 \end{bmatrix}$$

To obtain the 10-year return period, we need make assumption on the frequencies of the hurricane occur each year. Since we have about 253 observations during the 90 year span and there are probably some reports missing between 1905-1920, we assume the frequency to be 3, even though it seems in recent years, this number should be slightly higher. Then the 10-year return period corresponds to the 30-observation period (m=30) in (2.10). Substitution with parameter estimates, we obtain $\hat{x}_m = 10,802$. And apply delta method to obtain an estimate of the variance

$$\text{Var}(\hat{x}_m) \approx \nabla x_m^T V \nabla x_m = 7,403,393$$

The standard deviation is 2,721. Thus a 95% confidence interval for 10-year return period x_m is $10,802 \pm 1.96 * 2,721 = [5469, 16135]$. Figure(3.6) gives the profile log-likelihood plot for the 10-year return period. From the plot we can obtain a non-symmetric confidence intervals for the shape 10-year return period $[7300, 19000]$. For a quick comparison, the confidence interval shifts to the higher end as the result of the curvature in the parameter surface. A nice interpretation of this result is, we

45

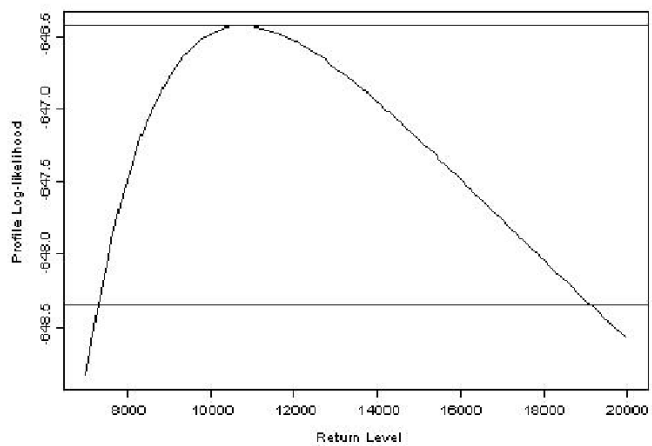are 95% confidently to say, that the costliest hurricane in the next 10 year period is between \$7.3B and \$19B.

FIGURE 3.6: Profile log-likelihood estimate of 10 year return level

CHAPTER 4

TRANSFORMS OF GEV AND GPD

4.1 Class of Transforms

The origin of transforms

In the modeling of extremal events, the asymptotic distributions of block maxima and excess over threshold were stated as GEV and GPD. In case the shape parameter ξ is positive, the support of the distribution function is in the form $[a, \infty]$ where

$$
a = \begin{cases} \mu - \dfrac{\sigma}{\xi}, & X \sim \mathrm{GEV}(\mu, \sigma, \xi) \\[2mm] 0, & X \sim \mathrm{GPD}(\tilde{\sigma}, \xi) \end{cases}
$$

In many examples we mentioned earlier, the support of the collected data has a compact support, i.e., the range is bounded. In the example of stock market returns, the daily percentage change has a range of $[-100\%, \infty)$. Particularly, in the analysis of market crashes, the data we use would be all negative returns which have range of $[0, 100\%]$ after negation. In the sea level example, we expect that there is an upper bound for the surge in the model. In modeling the total loss caused by a catastrophe event, a natural upper bound would be the total wealth of the covered region. In the low temperature example, the natural limit of any low temperature would be the absolute zero degree in physics.

Different examples require different strategies. If we are tracing a major stock market index such as Dow Jones or S&P 500, the probability of a near -100% drop would be the same as the default rate of treasury bond, which is negligible. If our concern is on the return of a single stock or a small portfolio, based on the recent bankcruptcy of corporations including Enron and Worldcom, the default risk may be still small, yet should be counted.

Suppose we denote the percentage change of a stock price during a trading day by r, where $-1 \leq r < \infty$. Positive r and negative r corresponds to the up and down of the stock price. Since the stock price can never go negative, the distribution of r is not symmetric. More importantly, the percentage change r does not satisfy the **addition property**: the total percentage change over a year is not the sum of all daily percentage changes. To overcome these difficulties, we have two solutions:

Solution 1: Use log-returns on both positive r and negative r. Let $x = \ln(1 + r)$ and the distribution of x will be symmetric and addition property is satisfied.

Solution 2: We will fix positive r and try to find an 'equivalent' negative r that will 'negate' the effect of positive r. Let

$$
r = \begin{cases} u, & \text{if } r > 0 \\ -d, & \text{if } r < 0 \end{cases}
$$

Suppose we observe the behavior of a stock price in two consecutive days. The stock price goes up by u on the first day and then goes down by d on the second day in percentage. If on the close of the second day, the stock price gets back to the original price, we have $(1 + u)(1 - d) = 1$, or $u = \dfrac{d}{1 - d}$, which means a drop of d is

'equivalent' to a rise of $\dfrac{d}{1-d}$ in value. This suggests a transform of $x = \dfrac{d}{1-d}$ on the negative return data will make the comparison to positive return data meaningful.

Transforms from $[0,1)$ to $[0,\infty)$

From the discussion above, it is necessary to transform data with bounded support to an unbounded domain to make the application of the extreme value theory more suitable. We first consider transforms that map $[0,1)$ to $[0,\infty)$, and that are smooth, simple and meaningful for plausible applications. Based on the discussion of the stock return example, three types of transforms can be found from the elementary functions:

Type I (standard hyperbolic transform with index p):

$$T(x) = \frac{x}{(1-x)^p}, \quad \text{where } p > 0 \tag{4.1}$$

Type II (standard logarithmic transform):

$$T(x) = -\ln(1-x) \tag{4.2}$$

Type III (standard trigonometric transform)

$$T(x) = \frac{2}{\pi}\tan(\frac{\pi}{2}x) \tag{4.3}$$

Note: These transforms are selected from elementary functions that have vertical asymptotes. In (4.1), p is typically chosen as 1 if there is no other information that favors other choices. It would be in general similar if we incorporate the parameter p in our model. To keep the calculation simple, we will use this assumption in later

analysis. For subject matter, a hyperbolic transform should be suitable for most subjects, while a logarithmic transform can apply to financial data and a trigonometric transform to engineering data.

Transformed GEV and GPD families

Using transforms similar to those listed in (4.1)–(4.3), we can map any bounded interval $[a, b]$ to $[a, \infty)$. Suppose T is the transform, we assume that T will fix a and map b to ∞, i.e., $T(a) = a$ and $\lim_{x \to b} T(x) = \infty$. We first use linear transform $\dfrac{x - a}{b - a}$ to map $[a, b)$ onto $[0, 1)$. Then apply the standard transforms to obtain the following transforms:

Type I (general hyperbolic transform with index p):

$$\frac{T_1(x) - a}{b - a} = \frac{\dfrac{x - a}{b - a}}{\left(1 - \dfrac{x - a}{b - a}\right)^p}, \quad \text{where } p > 0 \tag{4.4}$$

Type II (general logarithmic transform):

$$\frac{T_2(x) - a}{b - a} = -\ln\left(1 - \frac{x - a}{b - a}\right) \tag{4.5}$$

Type III (general trigonometric transform)

$$\frac{T_3(x) - a}{b - a} = \frac{2}{\pi} \tan\left(\frac{\pi}{2} \frac{x - a}{b - a}\right) \tag{4.6}$$

Apply these transforms to the GEV and GPD distribution, where $a = \mu - \dfrac{\sigma}{\xi}$ and $a = 0$ respectively. We introduce a new parameter $\tau = b - a$ to be the spread of the domain. τ will be referred as the range parameter in later use.

Definition 4.1. *We say X is from a transformed GEV or GPD distribution if Z follows GEV or GPD distribution, where Z is defined by the following:*

Type I (general hyperbolic transform with index p of GEV or GPD):

$$Z = \frac{X - a}{\left(1 - \dfrac{X - a}{\tau}\right)^p} + a \tag{4.7}$$

Type II (general logarithmic transform of GEV or GPD):

$$Z = -\tau \ln\left(1 - \frac{X - a}{\tau}\right) + a \tag{4.8}$$

Type III (general trigonometric transform of GEV or GPD)

$$Z = \frac{2\tau}{\pi} \tan\left[\frac{\pi(X - a)}{2\tau}\right] + a \tag{4.9}$$

where $a = \mu - \dfrac{\sigma}{\xi}$ in GEV and $a = 0$ in GPD. And $p > 0$, $\sigma > 0$, $\tau > 0$. $\qquad\square$

From the definition, the supports of the transformed GEV and GPD distributions are bounded, i.e. $[\mu - \dfrac{\sigma}{\xi}, \mu - \dfrac{\sigma}{\xi} + \tau)$ and $[0, \tau - \tilde{u})$ respectively.

Example 4.1. *CDF of hyperbolic transform $(p = 1)$ of GEV distribution $G_{T_1}(x)$*

The cumulative distribution function for X can be easily derived. Suppose Z follows a GEV distribution and the transform is

$$Z = \frac{X - a}{1 - \dfrac{X - a}{\tau}} + a = \frac{\tau(X - a)}{\tau - (X - a)} + a \qquad where\ a = \mu - \frac{\sigma}{\xi}$$

Since the transform function is strictly increasing (properties of the transforms will be proved in next section), we obtain:

$$G_{T_1}(x) \;=\; Pr(X \leq x)$$

51

$$= Pr\left(Z \le \frac{x-a}{1-\dfrac{x-a}{\tau}}+a\right)$$

$$= exp\left\{-\left[1+\xi\left(\frac{\dfrac{\tau(x-a)}{\tau-(x-a)}+a-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

$$= exp\left\{-\left[1+\xi\left(\frac{\dfrac{\tau(x-a)}{\tau-(x-a)}-\dfrac{\sigma}{\xi}}{\sigma}\right)\right]^{-1/\xi}\right\}$$

$$= exp\left\{-\left[\frac{\xi\tau}{\sigma}\frac{x-a}{\tau-(x-a)}\right]^{-1/\xi}\right\} \tag{4.10}$$

$\square$

Example 4.2. *CDF of logarithmic transform of GPD distribution $H_{T_2}(x)$*

Suppose $(Y-u|Y>u)$ follows a GPD distribution for large enough threshold u (say, when $u \ge u_0$) and the transform is

$$Y = -\tau \ln\left(1 - \frac{X-a}{\tau}\right) + a$$

Using the monotonicity of the transform, we obtain:

$$
\begin{aligned}
H_{T_2}(x) &= Pr\left(X-u \le x \,|\, X>u\right)\\
&= Pr\left[X \le x+u \,|\, Y > -\tau\ln(1-\frac{u-a}{\tau})+a\right]\\
&= Pr\left[Y \le -\tau\ln(1-\frac{x+u-a}{\tau})+a \,|\, Y > -\tau\ln(1-\frac{u-a}{\tau})+a\right]
\end{aligned}
$$

Denote $u_\tau = -\tau\ln(1-\dfrac{u-a}{\tau})+a$, then

$$H_{T_2}(x) = Pr\left[Y - u_\tau \le -\tau\ln(1-\frac{x+u-a}{\tau})+a-u_\tau \,|\, Y > u_\tau\right]$$

$$= Pr\left[Y - u_\tau \leq \tau \ln\left[\frac{\tau - (u - a)}{\tau - (x + u - a)}\right] \mid Y > u_\tau\right]$$

$$= 1 - \left\{1 + \frac{\xi}{\tilde{\sigma}}\tau \ln\left[\frac{\tau - (u - a)}{\tau - (x + u - a)}\right]\right\}^{-1/\xi} \tag{4.11}$$

where

$$\tilde{\sigma} = \sigma + \xi(u_\tau - \mu)$$

$$= \sigma + \xi\left[-\tau \ln\left(1 - \frac{u - a}{\tau}\right) + a - \mu\right]$$

$$= -\xi\tau \ln\left(1 - \frac{u - a}{\tau}\right) \tag{4.12}$$

Replace $\tilde{\sigma}$ in (4.11) by (4.12):

$$H_{T_2}(x) = 1 - \left[1 - \frac{\ln(\tau + a - u) - \ln(\tau + a - x - u)}{\ln(\tau + a - u) - \ln\tau}\right]^{-1/\xi}$$

$$= 1 - \left[\frac{\ln\tau - \ln(\tau + a - x - u)}{\ln\tau - \ln(\tau + a - u)}\right]^{-1/\xi}$$

The CDF function solely depends on ξ and $u - a$. By denoting $\tilde{u} = u - a = u - \mu + \frac{\sigma}{\xi}$ and $\tilde{u}_0 = u_0 - a = u_0 - \mu + \frac{\sigma}{\xi}$, we get

$$H_{T_2}(x) = 1 - \left[\frac{\ln\tau - \ln(\tau - \tilde{u} - x)}{\ln\tau - \ln(\tau - \tilde{u})}\right]^{-1/\xi}$$

where

$$0 < \tilde{u}_0 \leq \tilde{u} < \tau, \qquad 0 \leq x < \tau - \tilde{u}$$

53

4.2 Properties of Transformed GEV and GPD Distributions

Transforms $T_1(x), T_2(x), T_3(x)$ in (4.4) – (4.6) are elementary functions in simple form. At the same time, these functions have nice properties so that they can be easily used in the calculation of probabilities. Theroem (4.1) lists the main results.

Theorem 4.1. *The transform functions $T_1(x), T_2(x), T_3(x)$ in (4.4) – (4.6) satisfy the following properties:*

(a) All functions are strictly increasing and concave up.

(b) All functions are smooth, and the their inverse functions exist and are smooth.

(c) $T_i(x) \geq (x - a)$ for all $x \in [a, b)$, $i = 1, 2, 3$.

(d) Fix a and $x \in [a, b)$, $|T_i(x) - x| \to 0$ as $\tau = (b - a) \to \infty$, $i = 1,2,3$.

Proof. (a) We need to show the first and second derivatives of $T_i(x)$ are positive.

$$T_1'(x) = \left(\frac{b - a}{b - x} \right)^p \left(1 + p \frac{x - a}{b - x} \right) > 1$$

$$T_1''(x) = \left(\frac{b - a}{b - x} \right)^p \frac{2p(b - x) + p(p + 1)(x - a)}{(b - x)^2} > 0$$

$$T_2'(x) = \frac{b - a}{b - x} > 1$$

$$T_2''(x) = \frac{b - a}{(b - x)^2} > 0$$

$$T_3'(x) = \sec^2 \left(\frac{\pi}{2} \frac{x - a}{b - a} \right) > 1$$

$$T_3''(x) = \frac{\pi}{2(b - a)} \sec^2 \left(\frac{\pi}{2} \frac{x - a}{b - a} \right) \tan(\frac{\pi}{2} \frac{x - a}{b - a}) > 0$$

Note: In all above equations, $a \leq x < b$.

(b) We can find all inverse functions except $p \neq 1$.

$$T_1^{-1}(x) = \frac{bx - a^2}{x + b - 2a}, \qquad x \geq a$$

$$T_2^{-1}(x) = b - (b - a) \exp\left(-\frac{x - a}{b - a}\right), \qquad x \geq a$$

$$T_3^{-1}(x) = \frac{2}{\pi}(b - a) \tan^{-1}\left(\frac{\pi}{2} \frac{x - a}{b - a}\right) + a, \qquad x \geq a$$

In case $p \neq 1$, the existence of an inverse function is obvious from (a) that all functions are monotone. The smoothness follows from the Implicit Function Theorem.

(c) Let $G_i(x) = T_i(x) - x$. $G_i(a) = 0$ and $G_i'(x) > 0$ for $x \geq a$ from (a). Thus we conclude $G_i(x) \geq 0$, i.e., $T_i(x) \geq x$ for all $x \in [a, b)$, $i = 1, 2, 3$.

(d) For $T_1(x)$, we have

$$T_1(x) - x = \frac{x - a}{\left(1 - \dfrac{x - a}{b - a}\right)^p} + a - x = (x - a)\left[\left(1 - \frac{x - a}{b - a}\right)^{-p} - 1\right]$$

where $\left(1 - \dfrac{x - a}{b - a}\right)^{-p} \downarrow 1$ as $(b - a) \to \infty$, thus $(T_1(x) - x) \to 0$.

For $T_2(x)$, the Taylor expansion of the function $-\ln(1 - z)$ is:

$$-\ln(1 - z) = z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots \qquad \text{for } 0 \leq z < 1 \qquad (4.13)$$

Let $z = \dfrac{x - a}{b - a}$ and apply (4.11) to $T_2(x)$, we obtain

$$
\begin{aligned}
T_2(x) &= -(b - a)\ln\left(1 - \frac{x - a}{b - a}\right) + a \\
&= (b - a)\left[\frac{x - a}{b - a} + O\left(\left(\frac{x - a}{b - a}\right)^2\right)\right] + a \\
&= x + (x - a)^2 O\left(\frac{1}{b - a}\right)
\end{aligned}
$$

55

Similarly, from the Taylor expansion of the function $\tan z$:

$$\tan z = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \cdots \qquad \text{for } 0 \le z < \frac{\pi}{2} \qquad (4.14)$$

Let $z = \frac{\pi}{2} \frac{x-a}{b-a}$ and apply (4.12) to $T_3(x)$, we obtain

$$
\begin{aligned}
T_3(x) &= \frac{2}{\pi}(b-a)\tan\left(\frac{\pi}{2}\frac{x-a}{b-a}\right) + a \\
&= \frac{2}{\pi}(b-a)\left[\frac{\pi}{2}\frac{x-a}{b-a} + O\left(\left(\frac{\pi}{2}\frac{x-a}{b-a}\right)^2\right)\right] + a \\
&= x + \frac{\pi}{2}(x-a)^2 O\left(\frac{1}{b-a}\right)
\end{aligned}
$$

And we conclude $(T_2(x) - x) \to 0$ and $(T_3(x) - x) \to 0$ as $(b-a) \to 0$. $\qquad\square$

Comments 4.1. *(1) It can be seen that from the construction of the transform, the transform depends only on the relative position of x to the end points a and b. To be exact, suppose $\frac{x-a}{b-a} = \lambda$ where $0 \le \lambda < 1$. Then $\frac{T_i(x) - a}{b-a} = f_i(\lambda)$, $i = 1, 2, 3$.*

(2) Property (b) indicates that all transforms are isomorphisms from $[a, b)$ to $[a, \infty)$. From (a) and (c), the transform performs an expansion on $[a, b)$. Point a is fixed, and point b has been pulled to infinity as the scale of the expansion is higher if the point is closer to point b.

(3) Property (d) indicates the consistence of the transformed distribution with the original GEV or GPD in modeling. The new model converges asymptotically to the original GEV or GPD model as the range parameter $\tau = b - a$ goes to infinity.

Next we study the behavior of the pdfs of transformed distributions.

56

Example 4.3. *pdf of hyperbolic transform of GEV distribution* $g_{T_1}(x)$

Suppose Z follows a GEV distribution and the transform is

$$Z = \frac{X - a}{\left(1 - \frac{X-a}{\tau}\right)^p} + a \qquad where \ a = \mu - \frac{\sigma}{\xi}$$

We can obtain cdf of X similar as in Example 4.1,

$$G_{T_1}(x) = exp\left\{ -\left[\frac{\xi}{\sigma} \frac{x - a}{\left(1 - \frac{x-a}{\tau}\right)^p}\right]^{-1/\xi} \right\}$$

Taking derivative, the pdf is:

$$
\begin{aligned}
g_{T_1}(x) &= G_{T_1}(x) \cdot \frac{1}{\xi} \left[\frac{\xi}{\sigma} \frac{x - a}{\left(1 - \frac{x-a}{\tau}\right)^p}\right]^{-1/\xi - 1} \cdot \\
&\quad \cdot \frac{\xi}{\sigma} \left[(x - a)(-p)\left(1 - \frac{x - a}{\tau}\right)^{-p-1} \cdot \frac{-1}{\tau} + \left(1 - \frac{x - a}{\tau}\right)^{-p}\right] \\
&= G_{T_1}(x) \cdot \frac{1}{\sigma} \left[\frac{\xi}{\sigma}(x - a)\right]^{-1/\xi - 1} \cdot \left(1 - \frac{x - a}{\tau}\right)^{p/\xi - 1} \cdot \left[1 + (p - 1)\frac{x - a}{\tau}\right]
\end{aligned}
$$

Since $\left(1 - \dfrac{x - a}{\tau}\right)^{p/\xi - 1} \cdot \left[1 + (p - 1)\dfrac{x - a}{\tau}\right] \to 1$ *as* $x \downarrow a$, *the pdf of X is close to the pdf of Z at the left end point* $x = a$. *In other words,*

$$\lim_{x \downarrow a} \frac{g_{T_1}(x)}{g(x)} = 1$$

When x approaches the right end point $a + \tau$, *we obtain:*

$$\lim_{x \uparrow (a+\tau)} g_{T_1}(x) = \begin{cases} 0, & \text{if } \xi < p \\[2mm] \dfrac{p}{\sigma}\left(\dfrac{\xi\tau}{\sigma}\right)^{-1/\xi - 1}, & \text{if } \xi = p \\[2mm] \infty, & \text{if } \xi > p \end{cases}$$
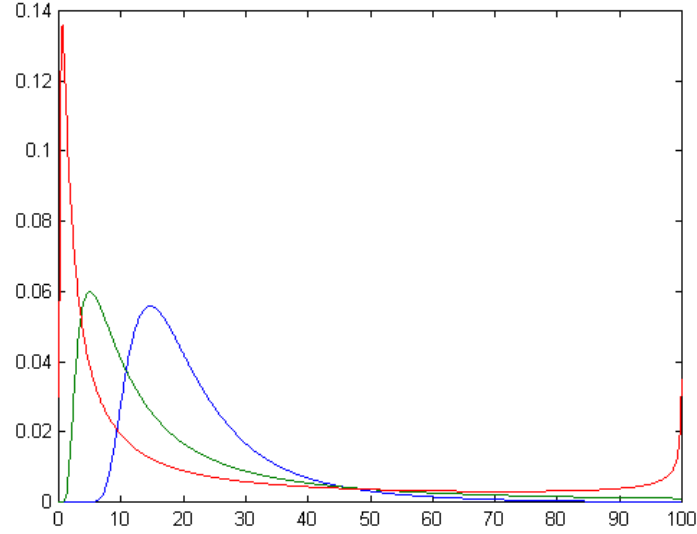
These properties can be verified from the plots in Figure (4.1).

FIGURE 4.1: pdf of hyperbolic transformed GPD family when $\xi = \dfrac{1}{2}, 1, 2$

$\square$

Example 4.4. *pdf of logarithmic transform of GPD distribution* $h_{T_2}(x)$

Suppose $(Y - u | Y > u)$ *follows a GPD distribution and the transform is*

$$Y = -\tau \ln\left(1 - \frac{X - a}{\tau}\right) + a$$

We obtained cdf of X in Example 4.2,

$$H_{T_2}(x) = 1 - \left[\frac{\ln \tau - \ln(\tau - \tilde{u} - x)}{\ln \tau - \ln(\tau - \tilde{u})}\right]^{-1/\xi}$$

Taking derivative, the pdf is:

$$h_{T_2}(x) = \frac{[\ln \tau - \ln(\tau - \tilde{u})]^{1/\xi}}{\xi(\tau - \tilde{u} - x)}[\ln \tau - \ln(\tau - \tilde{u} - x)]^{-1/\xi - 1}$$

58

where

$$0 \leq \tilde{u} < \tau, \qquad \tilde{0} \leq x < \tau - \tilde{u}$$

Similar to the approaches in Example 4.3, we obtain:

$$\lim_{x \downarrow 0} \frac{h_{T_2}(x)}{h(x)} = \frac{\tilde{u}}{\tau - \tilde{u}} \cdot \frac{1}{\ln \tau - \ln(\tau - \tilde{u})} \qquad and \qquad \lim_{x \uparrow (\tau - \tilde{u})} h_{T_2}(x) = \infty$$

The following are the plots when $\xi = \dfrac{1}{2}, 1, 2$ and $\tilde{u} = 10, 20, 30$ respectively.
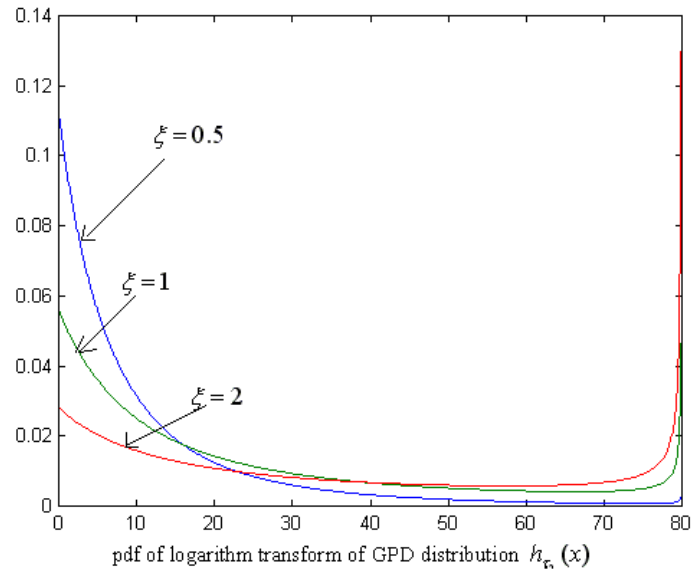

FIGURE 4.2: pdf of logarithmic transformed GPD family when $\xi = \dfrac{1}{2}, 1, 2$
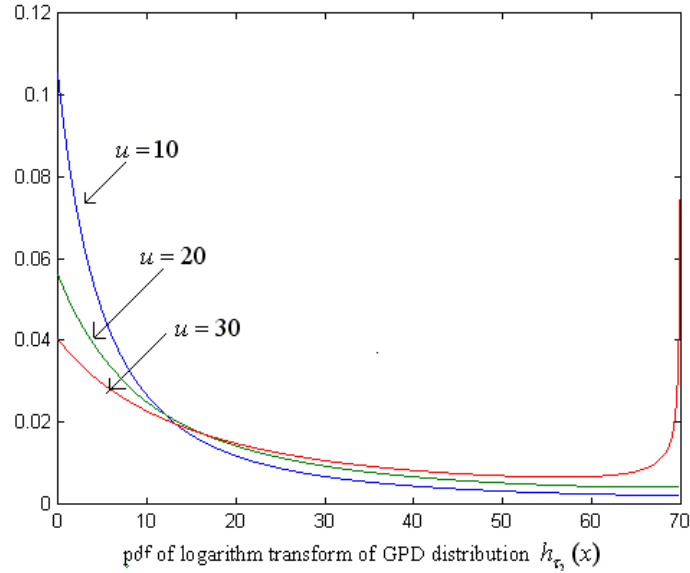
FIGURE 4.3: pdf of logarithmic transformed GPD family when $\tilde{u} = 10, 20, 30$

Comments 4.2.

(1) Indicated by Theorem 4.1 (d) and checked above, the pdfs of transformed GEV and GPD are very close to the original pdfs when x is close to the left end point.

(2) It is interesting to look at the result at the right end point. For pdf under transform $T_1(x)$, the end behavior depends on ξ. If the ξ is larger than index p (corresponds to heavy tail), the pdf blows up at the right end point which suggests 'point mass' behavior. For transform $T_2(x)$, this 'point mass' behavior will always occur for any positive ξ. We can interpret $T_2(x)$ as the asymptotic behavior of $T_1(x)$ as $p \downarrow 0$. Pdf using transform $T_3(x)$ shows behavior similar to the case of $p = 1$ in $T_1(x)$.

Table 4.1–4.2 consists of cdfs and pdfs of transformed GEV and GPD for reference:

Function	Expression $\left(\text{where } a = \mu - \dfrac{\sigma}{\xi} \right)$
$G_{T_1}(x)$	$\exp\left\{ - \left[\dfrac{\xi}{\sigma} \dfrac{x-a}{\left(1 - \frac{x-a}{\tau}\right)^p} \right]^{-1/\xi} \right\}$
$g_{\mathrm{T}_1}(x)$	$G_{T_1}(x) \cdot \dfrac{1}{\sigma} \left[\dfrac{\xi}{\sigma}(x-a) \right]^{-1/\xi-1} \cdot \left(1 - \dfrac{x-a}{\tau} \right)^{p/\xi-1} \cdot \left[1 + (p-1)\dfrac{x-a}{\tau} \right]$
$G_{T_2}(x)$	$\exp\left\{ - \left[-\dfrac{\xi\tau}{\sigma} \ln\left(1 - \dfrac{x-a}{\tau} \right) \right]^{-1/\xi} \right\}$
$g_{\mathrm{T}_2}(x)$	$G_{T_2}(x) \cdot \dfrac{1}{\sigma} \cdot \dfrac{\tau}{\tau - (x-a)} \cdot \left[-\dfrac{\xi\tau}{\sigma} \ln\left(1 - \dfrac{x-a}{\tau} \right) \right]^{-1/\xi-1}$
$G_{T_3}(x)$	$\exp\left\{ - \left[\dfrac{2\xi\tau}{\pi\sigma} \tan\left(\dfrac{\pi}{2} \dfrac{x-a}{\tau} \right) \right]^{-1/\xi} \right\}$
$g_{\mathrm{T}_3}(x)$	$G_{T_3}(x) \cdot \dfrac{1}{\sigma} \sec^2\left(\dfrac{\pi}{2} \dfrac{x-a}{\tau} \right) \left[\dfrac{2\xi\tau}{\pi\sigma} \tan\left(\dfrac{\pi}{2} \dfrac{x-a}{\tau} \right) \right]^{-1/\xi-1}$

TABLE 4.1 cdfs and pdfs of transformed GEV distributions.

Function	Expression $\left(\text{where } \tilde{u} = u - a = u - \mu + \dfrac{\sigma}{\xi} \right)$
$H_{T_1}(x)$	$1 - \tilde{u}^{p/\xi}(\tau - \tilde{u})^{-p/\xi}(x + \tilde{u})^{-p/\xi}(\tau - \tilde{u} - x)^{p/\xi}$
$h_{\mathrm{T}_1}(x)$	$\dfrac{p\tau}{\xi} \tilde{u}^{p/\xi}(\tau - \tilde{u})^{-p/\xi}(x + \tilde{u})^{-p/\xi-1}(\tau - \tilde{u} - x)^{p/\xi-1}$
$H_{T_2}(x)$	$1 - \left[\dfrac{\ln\tau - \ln(\tau - \tilde{u} - x)}{\ln\tau - \ln(\tau - \tilde{u})} \right]^{-1/\xi}$
$h_{\mathrm{T}_2}(x)$	$\dfrac{[\ln\tau - \ln(\tau - \tilde{u})]^{1/\xi}}{\xi(\tau - \tilde{u} - x)} [\ln\tau - \ln(\tau - \tilde{u} - x)]^{-1/\xi-1}$
$H_{T_3}(x)$	$1 - \left[\tan\left(\dfrac{\pi\tilde{u}}{2\tau} \right) \right]^{1/\xi} \left[\tan\left(\dfrac{\pi}{2} \dfrac{\tilde{u} + x}{\tau} \right) \right]^{-1/\xi}$
$h_{\mathrm{T}_3}(x)$	$\dfrac{\pi}{2\tau\xi} \left[\tan\left(\dfrac{\pi\tilde{u}}{2\tau} \right) \right]^{1/\xi} \left[\tan\left(\dfrac{\pi}{2} \dfrac{\tilde{u} + x}{\tau} \right) \right]^{-1/\xi-1} \sec^2\left(\dfrac{\pi}{2} \dfrac{\tilde{u} + x}{\tau} \right)$

TABLE 4.2 cdfs and pdfs of transformed GPD distributions.

4.3 Conditional Expectation of Excess Over Threshold

As we discussed in earlier chapters, the conditional expectation of excess over threshold is helpful in determining the choice of threshold u. There are a few reasons that the corresponding results have certain restrictions in usage. Recall the formula

$$E(X - u | X > u) = \frac{\sigma_{u_0} + \xi u}{1 - \xi}$$

provided $\xi < 1$. When $\xi \geq 1$, the mean is infinite. In practice, since the data collected will always produce a finite mean residual life plot, an interpretation is impossible in case $\xi \geq 1$. Also in the above formula, the conditional expectation is a linear curve independent of the threshold u. Both the mean residual life plot of the rainfall data (Figure(5.1)) from [6] and the mean residual life plot of the fire data from (Figure (5.2)) from [9] indicate a forced damping at the higher value end. Figures and more discussions can be found in Chapter 5.

Moreover, the method should be used very cautiously due to the lack of stability of the plot even when the underlying distribution is exactly Pareto distribution. We refer to [9] (page 294-303) for the discussion.

Things are very different if transformed distributions are used. We concentrate our result on the transformed GPD distributions. First, the conditional expectation is finite for all $\xi > 0$ since the support of the transformed distribution is finite. And when the threshold approaches the higher end of the support, the conditional expectation converges to zero. Explicit formulas are not possible for all transformed distributions. However, in case that we are able to solve, the illustration is clear as in the next example.

Example 4.5. *Conditional Expectation of Excess Over Threshold for $H_{T_1}(x)$*

Notice that the distribution depends solely through $\dfrac{p}{\xi}$, firstly we study the case when

$\dfrac{p}{\xi} = 1$. *Correspondingly, $h_{T_1}(x) = \dfrac{\tau \tilde{u}}{\tau - \tilde{u}}(x + \tilde{u})^{-2}$ Thus for $\tilde{u} > \tilde{u}_0$,*

$$
\begin{aligned}
E(X - u | X > u) &= \int_0^{\tau - \tilde{u}} x h_{T_1}(x) dx \\
&= \int_0^{\tau - \tilde{u}} \frac{\tau \tilde{u}}{\tau - \tilde{u}} \frac{x}{(x + \tilde{u})^2} dx \qquad (substitute \quad y = x + \tilde{u}) \\
&= \frac{\tau \tilde{u}}{\tau - \tilde{u}} \int_{\tilde{u}}^{\tau} \frac{y - \tilde{u}}{y^2} dy \\
&= \frac{\tau \tilde{u}}{\tau - \tilde{u}} \left[\ln y + \frac{\tilde{u}}{y} \right] |_{y = \tilde{u}}^{y = \tau} \\
&= \frac{\tau \tilde{u}}{\tau - \tilde{u}} \left[\ln \tau + \frac{\tilde{u}}{\tau} - \ln \tilde{u} - 1 \right] \\
&= \frac{\tau \tilde{u}}{\tau - \tilde{u}} \ln \frac{\tau}{\tilde{u}} - \tilde{u}
\end{aligned}
$$

Let $\tilde{u} = \lambda \tau$, then

$$
E(X - u | X > u) = \tau \left(\frac{-\lambda \ln \lambda}{1 - \lambda} - \lambda \right) \quad for \quad \lambda_0 = \frac{u_0}{\tau} < \lambda < 1
$$

A plot of $m(\lambda) := \dfrac{-\lambda \ln \lambda}{1 - \lambda} - \lambda$ against λ can be generated by Matlab.

From Figure)(4.4), we conclude the following properties:

- *The plot of $m(\lambda)$ is a bell-shaped curve which skewed to the left.*

- *The maximum occurs at about $\lambda_0 = 0.3162$ or $\tilde{u} = 0.3162\tau$.*

- *The curve is asymptotic to a line with slope $-\dfrac{1}{2}$ at the right end point $\lambda = 1$.*
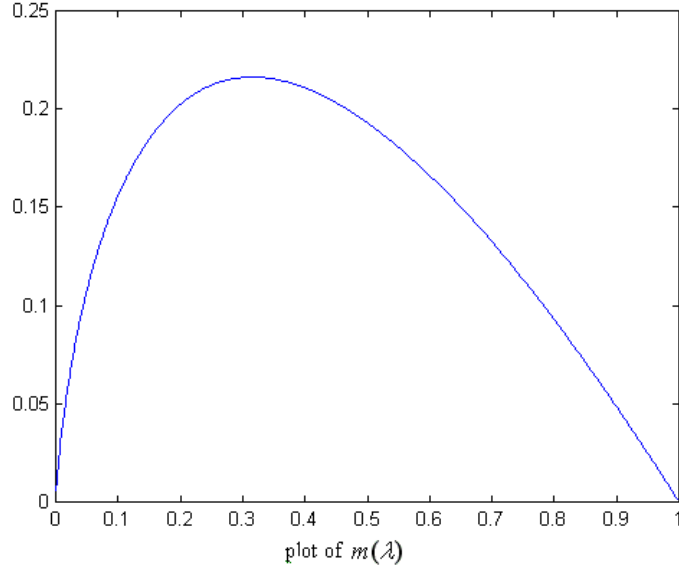
63

FIGURE 4.4: Plot of $m(\lambda)$

To obtain these properties, we simply look at

$$m'(\lambda) = \frac{(1-\lambda)(-2+\lambda) - \ln\lambda}{(1-\lambda)^2}$$

$$m''(\lambda) = \frac{(1-\lambda)(-1-\lambda) + \ln\lambda}{(1-\lambda)^3} < 0 \quad for \quad 0 < \lambda < 1$$

Then

$$\lim_{\lambda \to 0} m'(\lambda) = \infty \qquad and \qquad \lim_{\lambda \to 1} m'(\lambda) = -\frac{1}{2}$$

And the only solution to $m'(\lambda) = 0$ is $\lambda_0 = 0.3162$.

Next we carry out similar computation in general case $p \neq \xi$. Define $r := \dfrac{p}{\xi} > 0$, so the distribution function of the hyperbolic transformed GPD distribution is given by

$$H_{T_1}(x) = 1 - \left(\frac{\tilde{u}}{\tau - \tilde{u}}\right)^r \left(\frac{\tau - \tilde{u} - x}{x + \tilde{u}}\right)^r$$

64

where $\tilde{u} > \tilde{u}_0$, then

$$
\begin{aligned}
E(X - u | X > u) &= \int_0^{\tau - \tilde{u}} x \, dH_{T_1}(x) \\
&= xH_{T_1}(x)\big|_{x=0}^{x=\tau-\tilde{u}} - \int_0^{\tau-\tilde{u}} H_{T_1}(x) \, dx \\
&= (\tau - \tilde{u}) - \int_0^{\tau-\tilde{u}} \left[1 - \left(\frac{\tilde{u}}{\tau - \tilde{u}} \right)^r \left(\frac{\tau - \tilde{u} - x}{x + \tilde{u}} \right)^r \right] dx \\
&= \left(\frac{\tilde{u}}{\tau - \tilde{u}} \right)^r \int_0^{\tau-\tilde{u}} \left(\frac{\tau - \tilde{u} - x}{x + \tilde{u}} \right)^r dx \qquad (\text{substitute} \quad y = x + \tilde{u}) \\
&= \left(\frac{\tilde{u}}{\tau - \tilde{u}} \right)^r \int_{\tilde{u}}^{\tau} \left(\frac{\tau - y}{y} \right)^r dy \qquad (\text{substitute} \quad z = \tau y, \tilde{u} = \lambda \tau) \\
&= \tau \left(\frac{\lambda}{1 - \lambda} \right)^r \int_\lambda^1 \left(\frac{1 - z}{z} \right)^r dz
\end{aligned}
$$

Define $m_r(\lambda) := \left(\dfrac{\lambda}{1 - \lambda} \right)^r \displaystyle\int_\lambda^1 \left(\dfrac{1 - z}{z} \right)^r dz$ and it can be interpreted as the excess over threshold ratio. We use Matlab to generate plot of $m_r(\lambda)$ in Figure(4.5) for different choices of r and then we conclude the following properties of $m_r(\lambda)$:

- The plot of $m(\lambda)$ is a bell-shaped curve with one single mode.

- The maximum excess over threshold ratio decreases as r increases, and the corresponding λ_0 decreases.

- The curve is asymptotic to a line with slope $-\dfrac{1}{r + 1}$ at the right end point $\lambda = 1$.
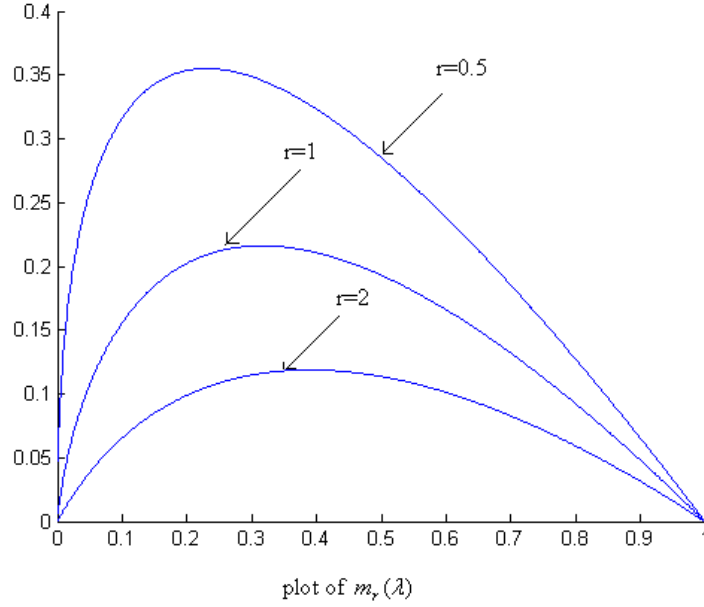
FIGURE 4.5: Plot of $m_r(\lambda)$

Some of the properties can be derived from the calculation of

$$
\begin{aligned}
m_r'(\lambda) &= \left[\left(\frac{\lambda}{1-\lambda}\right)^r\right]' \cdot \int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz + \left(\frac{\lambda}{1-\lambda}\right)^r \cdot \frac{d}{d\lambda} \int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz \\
&= r\left(\frac{\lambda}{1-\lambda}\right)^{r-1} \frac{1}{(1-\lambda)^2} \int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz - 1 \\
&= r\frac{\lambda^{r-1}}{(1-\lambda)^{r+1}} \int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz - 1
\end{aligned}
$$

$$
\begin{aligned}
\lim_{\lambda\to 1} m_r'(\lambda) &= r \cdot \lim_{\lambda\to 1} \frac{\int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz}{(1-\lambda)^{r+1}} - 1 \quad (\textit{apply L'Hopital's Rule}) \\
&= r \lim_{\lambda\to 1} \frac{-\left(\frac{1-\lambda}{\lambda}\right)^r}{(r+1)(1-\lambda)^r(-1)} - 1
\end{aligned}
$$

66

$$= -\frac{1}{r+1}$$

If $r < 1$, since $\frac{1-z}{z} \geq 1$ when $z \leq \frac{1}{2}$, we have

$$\int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz \geq \int_\lambda^{1/2} 1 dz = \frac{1}{2} - \lambda$$

Then

$$\lim_{\lambda \to 0} m_r'(\lambda) \geq \lim_{\lambda \to 0} r\lambda^{r-1}(\frac{1}{2} - \lambda) - 1 = \infty$$

If $r > 1$, $\lim_{\lambda \to 0} \int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz = \infty$, so

$$
\begin{aligned}
\lim_{\lambda \to 0} m_r'(\lambda) &= r \cdot \lim_{\lambda \to 0} \frac{\int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz}{\lambda^{1-r}} - 1 \quad (apply\ L'Hopital's\ Rule) \\
&= r \lim_{\lambda \to 1} \frac{-\left(\frac{1-\lambda}{\lambda}\right)^r}{(1-r)\lambda^{-r}} - 1 \\
&= \frac{1}{r-1} > 0
\end{aligned}
$$

In summary, we have

$$\lim_{\lambda \downarrow 0} m_r'(\lambda) = \begin{cases} \infty, & if\ r \leq 1 \\ \dfrac{1}{r-1} & if\ r > 1 \end{cases}$$

Thus, the mean excess over threshold function increases at the small threshold and decreases for the large threshold. From continuity of $E(X - u|X > u)$ with respect to u, there exists $\lambda = \lambda_0$ or $\tilde{u}_0 = \lambda\tau$ such that $E(X - u|X > u)$ reaches maximum.

CHAPTER 5

MODELING OVER TRANSFORMED DISTRIBUTIONS

5.1   Model Selection: Transformed or Not?

As we discussed in Chapter 4, modeling over the transformed distributions is a technique that extends the original models to be applied to the cases when natural limit exists. It enables us to choose models from broader ranges with the additional range parameter τ. We approach the model selection in the following steps.

The first thing we need to consider is whether the natural limit exists. In most cases the answer is yes, since all of our measurements and readings are finite. However, our major concern is whether the introduction of this limit will affect the modeling. For example, suppose the age distribution of the United States population is our interest. Then currently people that are at least 100 years old is only a negligible portion in the study. Any error caused by inaccurate estimation of this age period may only have a very slight effect on the whole model. Things are different if our study object changes to the population that are at least 80 years old. A more accurate model is desirable for the population with age over 100. In detail, since the general Pareto distribution has linear mean excess function and there is a natural force toward mortality for very old people, we see already that the mean excess function changes to

fit this situation when we apply transforms. Thus a transformed model is preferable when an elderly population is studied.

For the specific transforms to be used in the model, we recommend the adoption of proper transforms from the tradition of different fields. For many models in hydrology and climatology, the hyperbolic transform (particularly when p = 1) is often a good choice. For distributions that are known to have log-normal distributions, stock returns, for instance, the logarithmic transform may perform well. The trigonometric transform might be attractive to engineers. The only criterion is to choose the simplest model that fits the data reasonably well.

The mean residual life plot is suggested to be used in the model selection. Notice that the major difference between the original GPD distribution and the transformed one is that for the transform, the mean excess function increases first and then decreases and the whole curve is concave down. If we take a look at the mean residual life plot of the hurricane data in chapter 3 (figure(3.2)), we can see that only one observation represents the decreasing part in the plot. Also in the hurricane data example, the natural limit of the total loss is the total wealth of people who live close to the southeast coastline. This limit is a huge number compared with the recorded damage losses. Hence we conclude that the range parameter ξ may not be necessary in the model of hurricane data.

We examine two examples where a transform is quite important. The first example is the daily rainfall data in Coles [9] (page 80), and the mean residual life plot is given in figure(5.1).
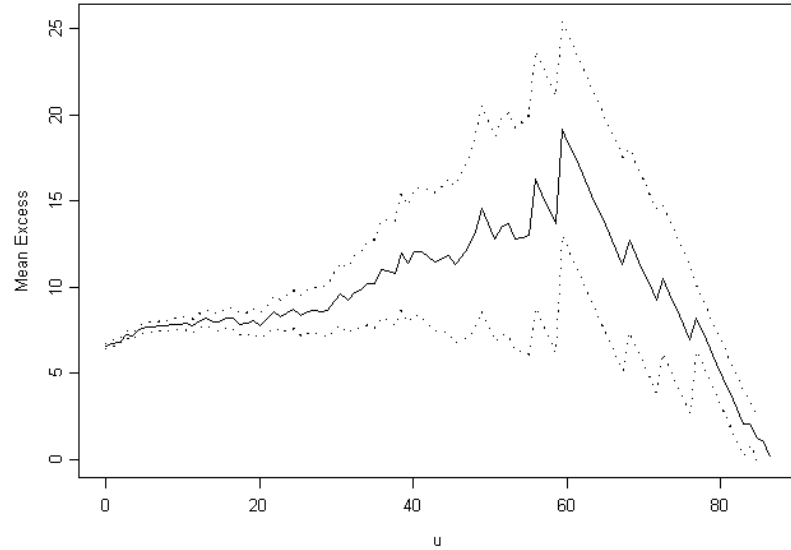
FIGURE 5.1: Mean residual life plot for daily rainfall data, Source: Coles[6]

The complete data set has 17,531 observations and there are 152 exceedances of the threshold $u = 30$. We can see clearly the maximum point in the plot and there are about 10 observations representing the decreasing part of the plot. Even though the evidence is not convincingly strong, we should give fully consideration of using transformed distributions in the model.

The next example is from Embrechts, Kluppelberg and Mikosch [6] (page 320). The mean residual life plot is attached. About 50 observations in the range of above 6000 that compose the decreasing part in the plot. The evidence is quite strong that the data seem to cluster towards some specific upper value. Using a transformed model becomes very appealing here.
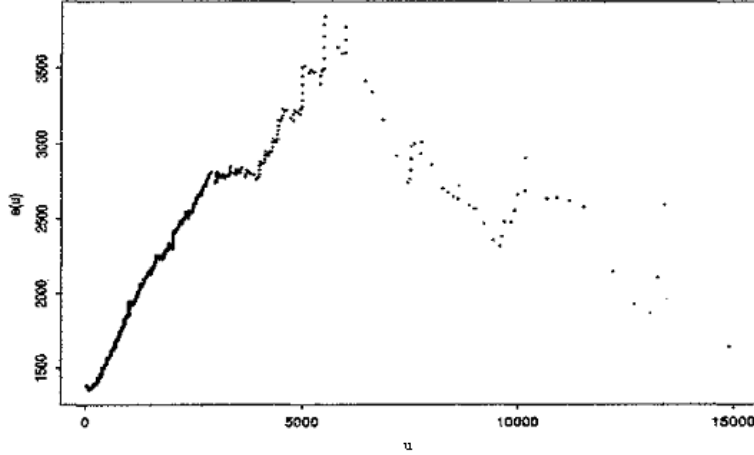
70

FIGURE 5.2: Mean residual life plot for fire data, Source: Embretchs[9]

Based on what we have discussed, we can go one step further to get an estimation
on the shape parameter ξ. Suppose we fit the data with a transformed GPD model
using the EOT method. Fix $p = 1$ in the hyperbolic transform and follow what we
have defined in chapter 4:

$$m_r(\lambda) := \left(\frac{\lambda}{1-\lambda}\right)^r \int_\lambda^1 \left(\frac{1-z}{z}\right)^r dz$$

where $r = \dfrac{1}{\xi} = \alpha$. Then we generate plots of $m_r(\lambda)$ for $\xi = 0.1, 0.2, ..., 2.0$ as in
figure(5.3). Keep in mind that $m_r(\lambda)$ does not take the range parameter τ into
account. Since in general τ is unknown, we cannot get any information about ξ
directly from τ. Also notice that the relative position of the rescaled maximum point
(the ratio of the y and x coordinates) stays constant for fixed ξ. Table(5.1), consisting
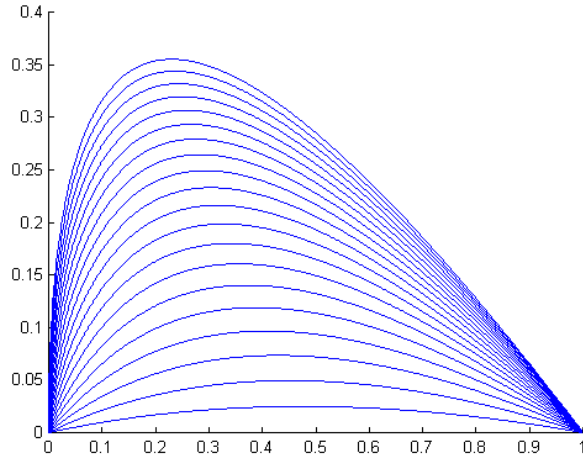rescaled maximum points and y/x ratios for various ξ, is generated for reference.

71

FIGURE 5.3: Mean excess function of transformed GPD for $\xi = 0.1, 0.2, ..., 2.0$

ξ	maximum point	y/x ratio	ξ	maximum point	y/x ratio
0.1	(0.475, 0.0249)	0.0524	1.1	(0.305, 0.2330)	0.7640
0.2	(0.451, 0.0495)	0.1097	1.2	(0.294, 0.2491)	0.8473
0.3	(0.429, 0.0735)	0.1712	1.3	(0.284, 0.2645)	0.9312
0.4	(0.409, 0.0966)	0.2362	1.4	(0.275, 0.2792)	1.0151
0.5	(0.390, 0.1189)	0.3048	1.5	(0.267, 0.2932)	1.0981
0.6	(0.372, 0.1402)	0.3768	1.6	(0.259, 0.3066)	1.1839
0.7	(0.356. 0.1605)	0.4509	1.7	(0.251. 0.3195)	1.2729
0.8	(0.342, 0.1800)	0.5262	1.8	(0.244, 0.3318)	1.3600
0.9	(0.329, 0.1985)	0.6033	1.9	(0.237, 0.3437)	1.4502
1.0	(0.316, 0.2162)	0.6841	2.0	(0.231, 0.3551)	1.5370

TABLE 5.1 Maximum points of the mean excess function for $\xi = 0.1, 0.2, ..., 2.0$.

The y/x ratio of the maximum point is monotonic as we increases ξ. Therefore we can get an estimate of ξ from the valuation of y/x ratio in the mean residual plot. We apply it to the two examples we mentioned. From figure(5.1) of the rainfall data, the y/x ratio of the maximum point is between 0.25 and 0.30. Check Table(5.1) we obtain an estimate of ξ between 0.4 and 0.5. Similarly, the y/x ratio of the maximum point in figure(5.2) for the fire data is about 0.7 and thus gives an estimate of ξ between 1.0 and 1.1.

Once ξ is estimated, we can combine Table(5.1) with the mean residual life plot to give references of the range parameter τ. Suppose we take $\xi = 0.4$ in the rainfall example. The observed x-coordinates of the maximum point is 62. Then $\frac{62}{0.409} = 151.6$ is the estimate of the natural limit τ in the model. Similarly in the fire data, we take $\xi = 1.0$ and then the estimate of the range parameter τ is $\frac{5600}{0.316} = 17,721$.

Even though the mean excess function of a transformed distribution, in general, has better stability than before, the lack of stability is still the major concern for it to be used as precise reference. We recommend the use of the standard model selection criteria described in section 2.4 for the addition of range parameter τ. For more accurate parameter estimation, we prefer the MLE method in the next section.

5.2 Parameter Estimations of Transformed Distributions

We adopt the MLE principle in the parameter estimation of transformed distributions. Suppose $\mathbf{x} = x_1, x_2, ..., x_n$ are realizations of independent variables from a common distribution F. The parameter estimation is given by the values of the parameters

that maximize the log-likelihood function in the parameter space. We list some of the calculation results for reference.

(1) Hyperbolic transform of the GEV family: case p = 1

Suppose $F(x) = G_{T_1}(x) = \exp\left\{ -\left[\frac{\xi\tau}{\sigma}\frac{x-a}{\tau-(x-a)}\right]^{-1/\xi}\right\}$ and $\theta = (\xi; a, \sigma, \tau)$, then

$$L(\theta|\mathbf{x}) = \frac{1}{\sigma^n}\cdot\exp\left\{-\sum_{i=1}^{n}\left[\frac{\xi\tau}{\sigma}\cdot\frac{x-a}{\tau-(x-a)}\right]^{-1/\xi}\right\}$$
$$\cdot\prod_{i=1}^{n}\left[\frac{\xi}{\sigma}(x_i-a)\right]^{-1/\xi-1}\cdot\prod_{i=1}^{n}\left(1-\frac{x_i-a}{\tau}\right)^{1/\xi-1}$$

and

$$l(\theta|\mathbf{x}) = -n\ln\sigma - \sum_{i=1}^{n}\left[\frac{\xi\tau}{\sigma}\cdot\frac{x-a}{\tau-(x-a)}\right]^{-1/\xi}$$
$$-(\frac{1}{\xi}+1)\sum_{i=1}^{n}\ln\left[\frac{\xi}{\sigma}(x_i-a)\right]+(\frac{1}{\xi}-1)\sum_{i=1}^{n}\ln\left(1-\frac{x_i-a}{\tau}\right)$$

The log-likelihood equations obtained by taking partial derivative of $l(\theta|\mathbf{x})$ are too complicated to display here. Again we call for numerical procedures for the maximization.

(2) Hyperbolic transform of the GPD family: case p = 1 and a = 0

Suppose $F(x) = H_{T_1}(x) = 1 - \left(\frac{u}{\tau-u}\cdot\frac{x+u}{\tau-x-u}\right)^{-1/\xi}$ and $\theta = (\xi; \tau)$, then

$$L(\theta|\mathbf{x}) = \tau^n\cdot\xi^{-n}\cdot\left(\frac{u}{\tau-u}\right)^{n/\xi}\cdot\prod_{i=1}^{n}(x_i+u)^{-1/\xi-1}\cdot\prod_{i=1}^{n}(\tau-x_i-u)^{1/\xi-1}$$

and

$$l(\theta|\mathbf{x}) = n\ln\frac{\tau}{\xi} + \frac{n}{\xi}\ln\frac{u}{\tau-u} - (\frac{1}{\xi}+1)\sum_{i=1}^{n}\ln(x_i+u)+(\frac{1}{\xi}-1)\sum_{i=1}^{n}\ln(\tau-x_i-u)$$

74

Taking the partial derivative of $l(\theta|\mathbf{x})$ with respect to ξ and τ, we obtain the following log-likelihood equations:

$$
\begin{cases}
-n\xi - n\ln\dfrac{u}{\tau - u} + \displaystyle\sum_{i=1}^{n} \ln\dfrac{x_i + u}{\tau - x_i - u} &= 0 \\
-n\tau + \dfrac{n}{\xi(\tau - u)} + (\dfrac{1}{\xi} - 1)\displaystyle\sum_{i=1}^{n} \dfrac{1}{\tau - x_i - u} &= 0
\end{cases}
\tag{5.1}
$$

Solving the first equation in (5.1) we obtain

$$
\xi = -\ln\frac{u}{\tau - u} + \frac{1}{n}\sum_{i=1}^{n} \ln\frac{x_i + u}{\tau - x_i - u} > 0
$$

Replacing ξ by this result in the second equation, we can obtain an implicit equation in τ. Solutions that satisfy $\tau > \max\{x_1, ..., x_n\}$ can be found by numerical procedures.

(3) Logarithmic transform of the GPD family: case a $= 0$

Suppose $F(x) = H_{T_2}(x) = 1 - \left[\dfrac{\ln\tau - \ln(\tau - x - u)}{\ln\tau - \ln(\tau - u)}\right]^{-1/\xi}$ and $\theta = (\xi; \tau)$, then

$$
L(\theta|\mathbf{x}) = \xi^{-n}\cdot\left(\ln\frac{\tau}{\tau - u}\right)^{n/\xi}\cdot\prod_{i=1}^{n}\frac{1}{\tau - x_i - u}\cdot\prod_{i=1}^{n}\left(\ln\frac{\tau}{\tau - x_i - u}\right)^{-1/\xi - 1}
$$

and

$$
l(\theta|\mathbf{x}) = -n\ln\xi + \frac{n}{\xi}\ln\left(\ln\frac{\tau}{\tau - u}\right) - \sum_{i=1}^{n}\ln(\tau - x_i - u) - (\frac{1}{\xi} + 1)\sum_{i=1}^{n}\ln\left(\ln\frac{\tau}{\tau - x_i - u}\right)
$$

Taking the partial derivative of $l(\theta|\mathbf{x})$ with respect to ξ and τ, we obtain the following log-likelihood equations:

$$
\begin{cases}
-n\xi - n\ln\left(\ln\dfrac{\tau}{\tau - u}\right) + \displaystyle\sum_{i=1}^{n}\ln\left(\ln\dfrac{\tau}{\tau - x_i - u}\right) &= 0 \\
(\dfrac{1}{\xi} + 1)\displaystyle\sum_{i=1}^{n}\left[\dfrac{u}{\tau(\tau - x_i - u)}\cdot\left(\ln\dfrac{\tau}{\tau - x_i - u}\right)^{-1}\right] &= \dfrac{nu}{\xi\tau(\tau - u)} + \displaystyle\sum_{i=1}^{n}\dfrac{1}{\tau - x_i - u}
\end{cases}
\tag{5.2}
$$

75

Solving the first equation in (5.2) we obtain

$$\xi = -\ln\left(\ln\frac{\tau}{\tau - u}\right) + \frac{1}{n}\sum_{i=1}^{n}\ln\left(\ln\frac{\tau}{\tau - x_i - u}\right) > 0$$

Replacing ξ by this result in the second equation, we can obtain an implicit equation in τ. Solutions that satisfy $\tau > \max\{x_1, ..., x_n\}$ can be found by numerical procedures.

(4) Trigonometric transform of the GPD family: case a $= 0$

Suppose $F(x) = H_{T_3}(x) = 1 - \left[\tan\left(\frac{\pi u}{2\tau}\right)\right]^{1/\xi}\left[\tan\left(\frac{\pi}{2}\frac{u+x}{\tau}\right)\right]^{-1/\xi}$ and $\theta = (\xi; \tau)$, then

$$L(\theta|\mathbf{x}) = n\ln\frac{\pi}{2\tau\xi}\cdot\left(\tan\frac{\pi u}{2\tau}\right)^{n/\xi}\cdot\prod_{i=1}^{n}\sec^2\frac{\pi(u+x_i)}{2\tau}\cdot\prod_{i=1}^{n}\left(\tan\frac{\pi(u+x_i)}{2\tau}\right)^{-1/\xi-1}$$

and

$$l(\theta|\mathbf{x}) = n\ln\frac{\pi}{2\tau\xi} + \frac{n}{\xi}\ln\left(\tan\frac{\pi u}{2\tau}\right) + 2\sum_{i=1}^{n}\ln\left(\sec\frac{\pi(u+x_i)}{2\tau}\right) - (\frac{1}{\xi}+1)\sum_{i=1}^{n}\ln\left(\tan\frac{\pi(u+x_i)}{2\tau}\right)$$

Taking the partial derivative of $l(\theta|\mathbf{x})$ with respect to ξ and τ, we obtain the following log-likelihood equations:

$$\begin{cases} -n\xi - n\ln\left(\tan\frac{\pi u}{2\tau}\right) + \sum_{i=1}^{n}\ln\left(\tan\frac{\pi(u+x_i)}{2\tau}\right) & = \quad 0 \\ n\tau + \frac{\pi u n}{\xi\sin(\pi u/\tau)} + \sum_{i=1}^{n}\left[\pi(u+x_i)\cdot\tan\frac{\pi(u+x_i)}{2\tau}\right] & = \quad (\frac{1}{\xi}+1)\sum_{i=1}^{n}\left[\frac{\pi(u+x_i)}{\sin(\pi(u+x_i)/\tau)}\right] \end{cases}$$

$$(5.3)$$

Solving the first equation in (5.3) we obtain

$$\xi = -\ln\left(\tan\frac{\pi u}{2\tau}\right) + \frac{1}{n}\sum_{i=1}^{n}\ln\left(\tan\frac{\pi(u+x_i)}{2\tau}\right) > 0$$

Replacing ξ by this result in the second equation, we can obtain an implicit equation in τ. Solutions that satisfy $\tau > \max\{x_1, ..., x_n\}$ can be found by numerical procedures.

76

The results we obtain here give very simple connection between the the shape parameter ξ and the range parameter τ. It looks very similar to the format of the known Hill estimator, which provides inference of the shape parameter ξ under the general maximum domain of attraction condition. For more detail, see Hill [14].

5.3 Future Research Directions

In general, transforming bounded data so that it can be fitted into unbounded distribution models provides a useful tool in case a natural bound exists. This technique can be easily combined with many other techniques in the research frontier of extreme values.

- For most problems involved with actuarial modeling and value at risk in finance, ξ is positive and the transforms work fine. The very intuitive problem is to extend the model to the cases when the shape parameter $\xi \leq 0$. When $\xi < 0$, one possible way is to find appropriate transforms $(-\infty, b] \to (a, b]$ and take similar steps as in previous chapters to obtain corresponding results. The transforms in chapter 4 all have little effect at the vicinity of the fixed point b. Thus the effect on the extreme (maximum) value is also small, may even negligible in this situation. And we need either to seek solutions that improves the model or give a reasonable explanation why transforms are not necessary. For the Gumbel distribution, transforms can only be performed on the half interval that includes the interesting end.

- As we have seen in section 4.2, the probability density functions(pdf) of the transformed GEV or GPD distributions may have an infinite limit at the upper bound $a + \tau$. This indicates that a mixture of continuous and discrete distribution may fit the model as well. Suppose the point mass taken at the upper bound is ϵ, then the pdf can be written as

$$f(x) = (1 - \epsilon)h(x) + \epsilon\delta(x - a - \tau)$$

  where h(x) is the pdf of transformed distribution and $\delta(x)$ is the unit point mass density.

- One of the most active areas in this field is extreme value theory applied to time-dependent sequences and certain time series models. By adopting the general theory of point processes, there is a natural way explaining the independence between parameters. In particular, the range parameter τ added to the GEV and GPD models can also depend on the time variable t. We can apply all techniques that are used to other parameters to fit a reasonable function for τ.

  For example, one of the interesting topics is the limits of human's athletic ability. Taking the record time of 100-meter dash as an index, we certainly know that there should be a natural limit for this record, for instance, 9.5 seconds. Our interest may be at the improvement of this limit as time evolves. Some approaches are possible here: one way is to collect the top 10 finishes of each year and apply order statistics techniques; an alternative way is to collect all runs with record time less than 10.2 seconds and then apply the EOT method.

- So far we have studied the extreme value theory of the univariate case. If there is an inter-relation between two variables, it is beneficial to study the joint behavior of the combined data. We list some examples of bivariate models: the distance between the moon and the earth vs. sea level; extreme weather vs. the El Nino effect; the exchange rate of US dollar against both UK sterling and Japanese yen. Our interest is in the effect of the introduction of the range variable τ. If the two variates in a bivariate model are positively correlated, and the range parameter τ is a significant factor for one of the variables, then probably the other variable can be fitted with a transformed model. Further research also needs to explore the possibility of constructing transforms that map the unbounded multidimensional regions to simple regions.

- Out last approach is to apply the technique of proper transforms to extreme models using Bayesian inferences and/or Markov chains. A Bayesian analysis of extreme value data is desirable if we acquire other source of information through a prior distribution. It is specially helpful when the number of available observations is not large enough to apply large sample approximation in extreme value theory. The byproduct of Bayesian analysis is a more complete inference through the posterior distribution than the MLE method. In case the extreme values seem to cluster, we can model the successive observations $x_1, x_2, ..., x_n$ with a first-order Markov chain process. The MLE method works well in this setting, and the introduction of the range τ only increases the dimension of the parameter space by 1.

BIBLIOGRAPHY

[1] Arnold, B.C. & Balakrishnan, N. and Nagaraja, H.N. *A First Course in Order Statistics.* Wiley, New York. (1992), 119,186,195.

[2] Azzalini, A. *Statistical Inference Based on the Likelihood.* Chapman and Hall, London. (1996).

[3] Bassi, F. & Embrechts, P. & Kafetzaki, M. "A survival kit on quantile estimation". *Environmetrics* **5** (1994), 221–239.

[4] Casella, G. & Berger, J.O. *Statistical Inference.* Duxbury, Belmont, California. (1990).

[5] Coles, S.G. & Tawn, J.A. "A seasonal Markov model for extremely low temperatures". *Environmetrics* **5** (1994), 221–239.

[6] Coles, S.G. *An Introduction to Statistical Modeling of Extreme Values.* Springer-Verlag, London. (2001)

[7] Coles, S.G. & Powell, E. A. "Bayesian methods in extreme value modeling: a review and new developments". *International Statistical Review* **64** (1996), 119–136.

[8] David, D.A. *Order Statistics.* Wiley, New York. (1981), 195,293.

[9] Embrechs, P., Kluppelberg, C. and Mikosch, T. *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin. (1997)

[10] Fisher, R.A. & Tippet, L.H.C. "Limiting forms of the frequency distribution of the largest or smallest number of the sample". *Proc. Cambridge Philos.* Soc. **24** (1928), 180–190.

[11] K. A. Froot, K. A. & O'Connell P. G., "The Pricing of US Catastrophe Reinsurance", NBER Working Paper No. 6043, May 1997, in The Financing of Catastrophe Risk, K. A. Froot, ed. Chicago: University of Chicago Press, 1999.

[12] Gnedenko, B.V. & Kolmogorov, A.N. *Limit Theorems for Sums of Independent Random Variables.* Addison-Wesley, Campbridge, Mass. (1954), 80–81.

[13] de Haan, L. "Fighting the arch-enemy with mathematics". *Statistica Neerlandica* **44** (1990), 45–68.

[14] Hill, B.M. "A simple general approach to inference about the tail of a distribution". *Ann. Statist* **3** (1975), 1163–1174.

[15] Jaffe, D.M. & Russell, T. "Catastrophe insurance, capital markets and uninsurable risk". (1996), Financial Institutions Center, The Wharton School, 96–12.

[16] Mishkin, F.S. & White, E.N. "U.S. stock market crashes and their aftermath: implications for monetary policy", (2002), asset price bubbles conference federal reserve bank of chicago and the world bank, Chicago, Illinois.

[17] Pielke, Jr., R. A. & C. W. Landsea, "normalized Hurricane Damages in the United States: 1925-1995". *Weather and Forecasting* **13** (1998), 621–631.

[18] Resnick, S.I. *Extreme Values, Regular Variation, and Point Processes.* Springer, New York. (1987), Proposition 0.3.

[19] Smith, R. L. "Maximum likelihood estimation in a class of non-regular cases". *Biometrika* **72** (1985), 67–90.