

# The Application of Extreme Value Theory in Payment of Electrical Engineering

Jingbo Liu

Economic Research Institute of Liaoning Electric  
Power Company Limited  
Liaoning, China  
ml3381322690@163.com

Ou Zhang

Economic Research Institute of Liaoning Electric  
Power Company Limited  
Liaoning, China  
745164085@qq.com

Xiangyu Zong \*

North China Electric Power University (Beijing)  
Beijing, China  
zongxiangyu@163.com  
\* Corresponding Author

Di Wang

Economic Research Institute of Liaoning Electric  
Power Company Limited  
Liaoning, China  
1056310314@qq.com

**Abstract**—Due to the owner and contractor's psychology of making a profit, the risks caused by enormous differences between actual payment and budget in electrical engineering indeed exist. However, it has not raised enough attention, and the relevant management of investment risk cannot solve it appropriately. A lot of problems are formed such as the rupture of capital chain, the delay of engineering, excessive investment and so on. Extreme value theory is a theory that studies the abnormal phenomenon in statistics, and it has been maturely used in risk analysis of financial field. The big differences between actual payment and budget can be treated as extremes. In this paper, researchers attempt to use the model of peaks over threshold based on extreme value theory to establish an early warning model, and the idea is to analyze the distribution of the extreme values, especially the tail which abnormal data assembles. This application notices the potential risk in electrical engineering and the method has the virtues of timeliness, veracity, easy operated that others do not have.

**Keywords**—*Electrical Engineering; Early Warning System; Extreme Value Theory; POT; Risk*

## I. INTRODUCTION

The risk management in electrical engineering investment is not easy as imagined. For the owner and contractor's psychology of making a profit, the owners do not pay fully the advanced fund of contractors, and more claims for compensation are frequently occurred. The further impact is the problems of wage arrears of migrant rural workers, beyond budgeting, engineering delay et al, which arouse wide concern in society without exception. But at present, there are still no effective ways to solve it well [1]. The traditional ways almost focus on the type of management, assessment of imputation or contracts to evaluated risks, and all these methods cannot give a precise description of risk. Indeed, the progress of engineering payment is decided by many factors which belongs to abnormal risks and usually cannot be foreseen [2].

The study of application of extreme theory began from 1930s, and Gumble firstly applied the extreme theory systematically to practice, which explains the flood statistical distribution, such as abnormal weather

observation statistical problems. In 1970s Pickands found the generalized Pareto distribution, and based on this find, peaks over threshold model (POT) was proposed, which simulates data's distribution of the sample that excesses the threshold and makes a clear description of tail extremes in original data[3]. Nowadays, the extreme value theory has widely used in the analysis of nature disaster, financial risk and so on [4]. In this paper researchers will analyze the difference between actual engineering payment and budget, establish a risk early warning model based on the processed statistical data, finally describe and control the risk in the investment.

## II. METHODOLOGY

Extreme events mean something that has lower probability to happen, but once occurred, it will do a great harm even bring a disaster. The extreme events cannot be early estimated, and they belong to the abnormal risk. In the process of engineering payment, the investment could be affected by a lot of factors, such as engineering changes, vibration of price, and the cut off of cash flow etc[5]. All these factors could lead the inconformity between budget and actual payment. Extreme value theory is based on the statistical analysis of data, especially for the distribution with "heavy tail and sharp peak" [6]. It can provide a clear description of data that exceeds threshold.

### A. Data Processing

In order to avoid the bigger error and saving costs, this paper studies the extreme value asymptotic distribution of the sample. In terms of an electrical engineering, it is easy to get the actual payment and budget, the sample data then could be produced after data processing [7].

The establishment of the model is based on the following assumptions:

- (1) Sample data meets the statistical characteristic of skewness and non-autocorrelation;
- (2) Sample data has the same nature, that is it comes from the same owner and contractor;
- (3) Sample data is sufficient.

During the process of engineering investment, the quantities are keeping changing as time limit for a project propelling, and there exist strong differences among each investment. Only taking the difference between actual payment  $c_i$  and budget  $b_i$  into account cannot obtain statistical regulation. In order to receive fine sample data, it is better to consider the absolute difference and relative difference between the actual payment and budget. In this paper, considering the practical condition of engineering and the limitation of the model, create statistics  $d_i$  :

$$d_i = \frac{c_i - b_i}{a_i} \quad (1)$$

$a_i$  is basic figure of an interval which is determined by the interval of  $c_i$  and  $b_i$ ; Engineering quantities have the characteristic of large sample statistics. So  $a_i$  also has corresponding characteristics of large sample. According to the central limit theorem, to a sample with the size of  $n$ , when  $n$  is sufficiently large, the average of sample obeys the average of  $\mu$  normal distribution. The payment of engineering can be thought as equal distributed among intervals, so it is reasonable to take  $a_i$  as the basic figure.  $d_i$  has the characteristic of similar normal distribution with the data processing [8].

For the electric power project construction period is long, usually with large quantities and capital-intensive, etc. in most cases the project funds would be paid monthly. Almost all of the payments are within 50 million CNY, for easy to count and weaken the influence of difference among every payment that caused the vibration of  $d_i$ , this paper takes 1 million CNY as interval length, divides into 5 ranges within 50 million CNY. Shown table 1.

TABLE I. INTERVAL DIFFERENTIATING AND BASIC FIGURE (10 THOUSANDS CNY)

| Financial intervals | [0,100] | [100,200] | [200,300] | [300,400] | [400,500] |
|---------------------|---------|-----------|-----------|-----------|-----------|
| $a_i$               | 50      | 150       | 250       | 350       | 450       |

### B. Generalized Pareto Distribution (GPD)

In order to describe the tail distribution of sample statistic precisely, this paper proposes to use the model of POT which is based on the GPD to simulate tail distribution GPD was firstly putted forward by Pickhands and Dumouchel gave its distribution model at the same year [9], that is:

$X_1, X_2, X_3, \dots, X_n$  is a sequence of positive independent and identically distributed random variables with the same distribution function, and there exists a positive function  $\beta(u)$  with a threshold of  $u$ . The excessive distribution of  $(X_i - u)$  can be expressed as:

$$G(y; \xi, \beta(u)) = 1 - \left(1 + \frac{\xi y}{\beta(u)}\right)^{-\frac{1}{\xi}}, \xi \neq 0 \quad (2)$$

If there is a threshold  $u_0$ , in terms of  $d_i$ , the excessive distribution is approximate GPD with parameter  $\sigma(u_0)$  and  $\xi$ , the probability of  $d_i$  can be expressed as,

$$p(d_i; \mu_0, \sigma, \xi) = \left(1 + \xi \frac{d_i - \mu_0}{\sigma}\right)^{-\frac{1}{\xi}} \quad (3)$$

In (3),  $\mu \in R$  is position parameter,  $\xi \in R$  is a shape parameter,  $\sigma > 0$  is a scale parameter.  $p(d_i > \mu_0) = 1 - \bar{p}$  means probability that  $d_i$  greater than  $u_0$ . The position parameter  $\mu$  can be calculated by sample data,  $\sigma, \xi$  can be estimated by maximum likelihood estimation (MLE) [10].

### C. The Method of Threshold Calculation Based on Hill Estimation

In fact, the model of GPD is a method that simulates the order statistics that exceed threshold  $u$ . The key is to make sure the appropriate threshold  $u$ . However, there is still no a uniform way about its selection at present. At present the method about the selection of  $u$  can be classified into graphical method and calculation method. The former is more vivid but the theory is not sufficient, while calculation method is limited by the sample statistics [11].

Hill estimation using model of exponential representations is used to estimate the optimal threshold by the optimal  $\hat{k}_n^{opt}$ . There is a positive parameter  $\gamma$  with the following form:

$$1 - F(x) = x^{-\gamma} l_F(x) \quad (4)$$

$l_F(x)$  is a slow change function, when  $x \rightarrow \infty$ ,  $\forall \lambda > 0, \frac{l_F(\lambda x)}{l_F(x)} \rightarrow 1, U(x) = x^\gamma l(x)$ .

Parameter  $\gamma$  in the above formulation is tail extreme indicator, based on the referred theory, Bierlant has created upper order statistic of logarithmic interval index regression model [12]:

$$Z_j = j(\log X_{n-j+1,n} - \log X_{n-j,n}) = (\gamma + b_{n,k} \left(\frac{j}{k+1}\right)^{-\rho}) f_j, 1 \leq j \leq k \quad (5)$$

$f_1, f_2, \dots, f_k$  is a series of independent random variables of standard index, and  $b_{n,k} = b \left(\frac{n+1}{k+1}\right), 2 \leq k \leq n-1$ .

The asymptotic variance of error (AMSE) can be written as:

$$AMSE(\hat{\gamma}_{k,n}^H) = AVar \hat{\gamma}_{k,n}^H + (ABias \hat{\gamma}_{k,n}^H)^2 = \frac{b_{n,k}^2}{4} + \frac{\gamma^2}{k} \quad (6)$$

From the above formulation it can be concluded that Avar and Abias are changing reversely, so there must exists an optimal  $k$  letting AMSE be the minimum, and  $k$  is the optimal  $k_n^{opt}$

$$k_n^{opt} = \arg \min_{3 \leq k \leq n} \left( \frac{b_{n,k}^2}{4} + \frac{\gamma^2}{k} \right) \quad (7)$$

#### D. Steps:

(1) Using maximum likelihood estimation estimates the parameter  $\hat{\gamma}, \hat{b}_{n,k}$  in the Eq. (5) when  $k \in \{3, \dots, n\}$ ;

(2) Working out AMSE for  $k \in \{3, \dots, n\}$ ;

(3) Selecting  $\hat{k}_n^{opt}$  that makes AMSE the minimum, and making sure the optimal threshold with the tool of MATLAB;

(4) Calculating the Hill estimation  $\gamma_n(\hat{k}_n^{opt})$ :

$$\gamma_n(\hat{k}_n^{opt}) = \frac{1}{\hat{k}_n^{opt}} \sum_{j=1}^{\hat{k}_n^{opt}} (\log X_{n-j+1,n} - \log X_{n-\hat{k}_n^{opt},n})$$

$\gamma_n(\hat{k}_n^{opt})$  is same to shape parameter  $\hat{\xi}$  in GPD;

(5) Simulating the excessive distribution and analyzing the fitting effects.

### III. ANALYSIS OF EXAMPLES

A grid project company provides 200 set of engineering settlement records as well as the corresponding budget amounts. According to the data processing mode, to deal with the engineering data, researchers can get the statistics  $d_i$  as shown in Fig. 1, using the data of interval probability density to fitting the image of probability density function, as shown in Fig. 2.

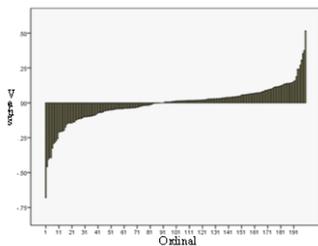


Figure 1. Sampling statistics

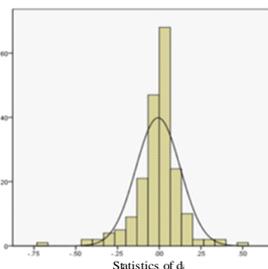


Figure 2. Statistics di frequency distribution histogram

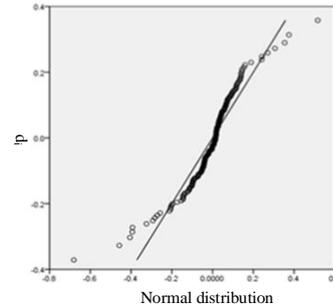


Figure 3. QQ plot

#### A. Skewness test:

If the second order and fourth order center moment of a random variable are existing, define the ratio of the fourth order center moment and the second order central moment of random variable's square as the kurtosis coefficients,

$$K = \frac{\mu_4}{\mu_2^2} = \frac{E(X - E(X))^4}{D(X)^2}$$

that is  $K = \frac{\mu_4}{\mu_2^2} = \frac{E(X - E(X))^4}{D(X)^2}$ ; coefficient of skewness  $S$  is a feature of describing the division of the deviation degree of symmetry, bilateral symmetry  $S = 0$ ,  $S > 0$  thick-tailed on the right side, distribution of the right;  $S < 0$  thick-tailed on the left side, distribution of the left.

Skewness is the third-order central moment of variable and the ratio of standard deviation, that is to say,  $S = \frac{\mu_3}{\sigma^3}$ ;  $JB = \frac{n(S^2 + K^2/4)}{6}$ , statistic  $JB$  is almost submissive the  $\chi^2$  distribution that degree of freedom is 2, that is  $JB \approx \chi^2(2)$ , distribution  $JB > \chi_a^2(2)$  is not conform to the normal distribution[13].

Taking advantage of the eigenvalues of the sample data which can get from the data analysis software SPSS, just as shown in table 2.

TABLE II. CHARACTERISTIC VALUE OF SKEWNESS TEST

| eigenvalue         | Eigenvalue |
|--------------------|------------|
| average            | -0.007     |
| median             | 0.012      |
| maximum            | 0.520      |
| minimum            | -0.68      |
| Standard deviation | 0.133      |
| skewness           | -0.785     |
| kurtosis           | 5.274      |
| JB                 | 252        |

Researchers can see from the Table 2. Sample distribution of skewness is less than 0, and kurtosis and standard deviation are greater than 0. Appearing a right and peak distribution, statistic JB is obviously greater than the critical value of chi-squared statistic which has two degrees of freedom, and it can be proved that the skewness's character of sequence  $d_i$ .

#### B. Autocorrelation Test:

In the actual engineering, intermediate settlement money and budget project payment are all occurring according to certain time node. Sequence  $d_i$  is a time series in strict significance, if there is autocorrelation among samples, it will affect the precision of the model output. The main possible causes of the autocorrelation in this case as following:

1. Payments of hysteresis cause the change of the random error term;
2. The interference of random factors;

3. Data processing cause serial correlation of random error term.

If the data has autocorrelation, it needs further processing of the data.

TABLE III. AUTOCORRELATION TEST OF  $d_i$

| lag | autocorr<br>elation | Standard<br>error | Box-ljung |    |       |
|-----|---------------------|-------------------|-----------|----|-------|
|     |                     |                   | values    | df | Sig.b |
| 1   | .945                | .070              | 788.210   | 1  | .000  |
| 2   | .905                | .070              | 921.648   | 2  | .000  |
| 3   | .865                | .070              | 1048.62   | 3  | .000  |
| 4   | .844                | .069              | 1169.22   | 4  | .000  |
| 5   | .822                | .069              | 1284.02   | 5  | .000  |
| 6   | .800                | .069              | 1394.11   | 6  | .000  |
| 7   | .779                | .069              | 1500.44   | 7  | .000  |
| 8   | .757                | .069              | 1603.02   | 8  | .000  |
| 9   | .737                | .069              | 1702.01   | 9  | .000  |
| 10  | .720                | .069              | 1394.14   | 10 | .000  |
| 11  | .705                | .068              | 1889.13   | 11 | .000  |
| 12  | .691                | .068              | 1169.22   | 12 | .000  |
| 13  | .677                | .068              | 1284.03   | 13 | .000  |

It can be seen from the data in the table that sequence  $d_i$  has non-autocorrelation. In the process of mathematical modeling analysis was carried out on the sample data, in this paper, using mathematical analysis tool MATLAB's powerful processing functions in data processing, researchers can calculate the tail up and down according to the threshold of the original sample[14].

After the test of skewness and conformity, selecting  $(\beta(u), \xi)$  that makes (8) the maximal and using the profile likelihood function can also estimate more accurate confidence intervals of  $\xi$  and  $\sigma$ [15].

$$L(y; \beta(u), \xi) = -k \ln(\beta(u)) - (1 + \frac{1}{\xi}) \quad (8)$$

In this example, the relevant parameters are shown in table4.

TABLE IV. RELEVANT PARAMETERS

| parameters  | $\xi$ | $\sigma$ | $u$   | $N_u$ |
|-------------|-------|----------|-------|-------|
| top tail    | 0.37  | 0.13     | 0.24  | 12    |
| bottom tail | 0.17  | 0.13     | -0.21 | 7     |

Changing (3) and adding the corresponding parameters, researchers can get the statistic range under different confidence level. Calculating under confidence level of 80%, 50% and 30% confidence interval and the difference interval respectively, researchers can get the results as table 5 shows:

TABLE V. THE INTERVALS OF CONFIDENCE AND DIFFERENCE

| confidence level | tail   | confidence interval | difference interval |
|------------------|--------|---------------------|---------------------|
| 80%              | top    | [0.24, 0.52]        | [12, 26]            |
|                  | bottom | [-0.21, -0.45]      | [-10.5, -22.5]      |
| 50%              | top    | [0.24, 0.34]        | [12, 17]            |
|                  | bottom | [-0.21, -0.30]      | [-10.5, -15]        |
| 30%              | top    | [0.24, 0.29]        | [12, 14.5]          |
|                  | bottom | [-0.21, -0.25]      | [-10.5, -12.5]      |

Taking the above three confidence levels as calculation basis to calculate the threshold of the early warning system, the results are reflected as table6.

TABLE VI. ALERT-SITUATION DISTRIBUTION INTERVALS (TEN THOUSANDS CNY)

| Alert-situation | situation                | intervals         |
|-----------------|--------------------------|-------------------|
| no alert        | payment exceeds budget   | [0, 12]           |
|                 | payment less than budget | [0, 10.5]         |
| light alert     | payment exceeds budget   | [12, 14.5]        |
|                 | payment less than budget | [10.5, 12.5]      |
| moderate alert  | payment exceeds budget   | [14.5, 17]        |
|                 | payment less than budget | [12, 14.5]        |
| severe alert    | payment exceeds budget   | [17, $\infty$ ]   |
|                 | payment less than budget | [14.5, $\infty$ ] |

#### IV. CONCLUSIONS

(1) The small probability of extreme risk that is deviated from budget in electrical engineering payment indeed exists, and it both comes from the owner and contractor. Researchers can get a precise description of the risk with extreme value theory, and the Risk early warning system based on POT model will give different warning react with the bias between actual payment and budget, which engages the relevant personnel taking corrective action in time;

(2) From the model data and output, it is not difficult to find that mostly actual payment is more than budget, but the extremes from actual payment less than budget are more commonly seen, which should be noticed in actual electrical engineering investment;

Though extremes are of small probability, it may cause a chain reaction of amplified if it happened. Therefore, it is necessary to control the occurrence of extremes. In this paper, based on the statistical law of threshold setting method has been proposed, it can sound warnings when there is an enormous difference between actual payment and budget. Taking the changes of actual engineering data into account, relevant parameters and threshold should keep pace, and some auxiliary for the corresponding risk plans should also be further analyzed and implemented.

#### REFERENCES

- [1] Karmakar M, Shukla G K. Managing extreme risk in some major stock markets: An extreme value approach[J]. International Review of Economics & Finance, 2015, 35: 1-25.
- [2] Beirlant J, Dierckx G, Guillou A, et al. On exponential representations of log-spacings of extreme order statistics[J]. Extremes, 2002, 5(2): 157-180.
- [3] Beggan C D, Beamish D, Richards A, et al. Prediction of extreme geomagnetically induced currents in the UK high - voltage network[J]. Space Weather, 2013, 11(7): 407-419.
- [4] Carter D J T, Challenor P G. Methods of fitting the Fisher-Tippett type 1 extreme value distribution[J]. Ocean Engineering, 1928, 10(3):191-199.
- [5] Gnedenko, B. Sur la distribution limite du terme maximum d'une série aléatoire[J]. Annals of Mathematics, 1943:423-453.
- [6] Gumbel E J. Statistical Study of Materials Fatigue[J]. OR, 1958, (2):148.
- [7] Balkema A A, Haan L D. Residual Life Time at Great Age[J]. Annals of Probability, 1974, (5):792-804.

- [8] Iii J P. Statistical Inference Using Extreme Order Statistics[J]. *Annals of Statistics*, 1975, (1):119-131.
- [9] Embrechts P, Klüppelberg C, Mikosch T. Modelling extremal events, *Applications of Mathematics (New)*[J]. MR-1458613 MR-1458613 EJP 19, 1997.
- [10] Reiss R D, Thomas M. *Statistical Analysis of Extreme Values-with applications to insurance*[J]. Birkhuser Verlag Basel, 1997, 106(A6):10915–10921.
- [11] Gilli M, K05llezi E. An Application of Extreme Value Theory for Measuring Financial Risk[J]. *Computational Economics*, 2006, 27(2-3):207-228.
- [12] Zhang Y, Cao Y. A fuzzy quantification approach of uncertainties in an extreme wave height modeling[J]. *Journal of Ocean*, 2015, 34(3):90-98.
- [13] Mainik G, Mitov G, Rüschendorf L. Portfolio optimization for heavy-tailed assets: Extreme Risk Index vs. Markowitz[J]. *Journal of Empirical Finance*, 2015.
- [14] Chen J. Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in zhejiang, china.[J]. *Plos One*, 2015, 10(2).
- [15] Brodin E, Rootzén H. Univariate and bivariate CPD methods for predicting extreme wind storm losses[J]. *Insurance: Mathematics and Economics*, 2009, 44(3): 345-356.