iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: III Month of publication: March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67620

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

Model Selection and Diagnostics in Extreme Value Theory

Mrs. K. Thavamani[1], Mrs. M. Valli[2], Mrs. S. Sudha[3]
[1]*Assistant Professor of Statistics, Department of Mathematics, K.S.R College of Arts and Science for Women, Tiruchengode, Namakkal*
[2, 3]*Assistant Professor of Statistics, Department of Mathematics and Statistics, Mahendra Arts and Science College, (Autonomous), Kalippatti, Namakkal, TamilNadu*

Abstract: Model selection and diagnostic testing are critical components in applying Extreme Value Theory (EVT) for real-world data analysis. EVT models, such as the Generalized Extreme Value (GEV) and Generalized Pareto Distribution (GPD), are used to estimate the behavior of extreme events, but selecting the appropriate model and evaluating its fit to the data are essential steps for reliable predictions. This paper discusses various model selection methods, including goodness-of-fit tests, cross-validation techniques, and model comparison criteria like Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). Additionally, diagnostic tools such as quantile-quantile (Q-Q) plots, probability-probability (P-P) plots, and residual analysis are explored to assess the adequacy of EVT models. The article emphasizes the importance of robust diagnostic procedures to ensure that the chosen model accurately captures the tail behavior of the data.
Keywords: Extreme Value Theory (EVT), Akaike's Information Criterion (AIC), quantile, robustdiagnostic.

I. INTRODUCTION

Extreme value theory deals with the stochastic behavior of the extreme values in a process. For a single process, the behavior of the maxima can be described by the three extreme value distributions–Gumbel, Frechet and negativeWeibullas suggested by Fisher and Tippett (1928). Kotz and Nadarajah (2000) indicated that the extreme value distributions could be traced back to the work done by Bernoulli in 1709. The first application of extreme value distributions was probably made by Fuller in 1914. Thereafter, several researchers have provided useful applications of extreme value distributions particularly to climate data from different regions of the world. Extreme Value Theory (EVT) is a statistical framework used to model the behavior of extreme events, such as maximum daily temperatures, financial crashes, or natural disasters. The core of EVT involves selecting and fitting appropriate models, such as the Generalized Extreme Value (GEV) distribution for block maxima and the Generalized Pareto Distribution (GPD) for peak over threshold (POT) data. However, model selection in EVT is not straightforward and requires careful consideration of the underlying data characteristics, model assumptions, and the goal of the analysis.

In EVT, model diagnostics are essential for validating the assumptions of the fitted model, particularly for extreme events that might not conform to standard distributions. Evaluating model fit can be challenging because extreme events are rare and the tail behavior of distributions may not be well understood. Various statistical tools have been developed to assess model selection and diagnostics in EVT, which include graphical methods like Q-Q plots and P-P plots, as well as numerical measures such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These techniques help in comparing the adequacy of different EVT models, and in determining the model that best represents the extreme behavior of the dataset.

II. BASIC DEFINITION OF EXTREME VALUE THEORY

1) Definition: Extreme Value Theory (EVT) is a statistical approach that focuses on analyzing extreme events or values in data, rather than assuming a normal or symmetric distribution. It provides a framework for understanding and modeling the behavior of extreme values that exceed a certain threshold.
2) Basic Definition: EVT is a statistical framework used to model and predict extreme events by analyzing the tails of probability distributions. It helps estimate the likelihood of rare but impactful occurrences.
3) Mathematical Definition: EVT studies the asymptotic behavior of the maximum (or minimum) of a large sample of independent and identically distributed (i.i.d.) random variables. It is characterized by two key approaches: the Block Maxima method (which follows the Generalized Extreme Value (GEV) distribution) and the Peaks Over Threshold (POT) method (which follows the Generalized Pareto Distribution (GPD)).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

4) Applied Definition: EVT is a statistical tool used in fields such as finance, meteorology, engineering, and insurance to assess and manage risks associated with extreme events, such as market crashes, floods, hurricanes, or structural failures. It helps organizations make informed decisions by quantifying the probability of rare but severe outcomes.

III. MODEL SELECTION IN EVT

The process of model selection in EVT typically starts with selecting a distribution, either GEV or GPD, based on the data structure (block maxima). Several statistical tools are employed to determine which model is most appropriate:

Model selection in EVT involves choosing an appropriate distribution to describe extreme values. There are two important methods used in modeling extreme value analysis and they are threshold selection and block maxima techniques. The threshold selection is important in many aspects of statistical inference of extreme or rare events because they use data more effectively than block maxima techniques.

A. Block Maxima Approach (GEV Distribution)

Uses the Generalized Extreme Value (GEV) distribution, which encompasses:

1) Gumbel (Type I): Light-tailed distributions (Example- Normal, Exponential)
2) Frechet (Type II): Heavy-tailed distributions (Example- Pareto)
3) Weibull (Type III): Bounded distributions (Example- Beta)

Selection involves assessing which GEV type best fits the block maxima.

B. Peaks Over Threshold (POT) Approach

Uses the Generalized Pareto Distribution (GPD) for values exceeding a high threshold.Requires threshold selection methods such as:

1) Mean residual life plots
2) Stability of parameter estimates
3) Threshold diagnostic plots

C. Model Selection Criteria

To determine the best EVT model, statistical techniques include:

1) Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC): These criteria are used to compare competing models. Lower AIC and BIC values indicate a better model fit, penalizing overly complex models.
2) Likelihood-based methods: Maximum likelihood estimation (MLE) is commonly used to fit EVT models, but it requires careful consideration of the number of parameters and potential overfitting.
3) Cross-validation: In some cases, a cross-validation approach can be applied to assess the predictive performance of EVT models.
4) Kolmogorov-Smirnov (KS) Test & Anderson-Darling Test: Measure fit to empirical data.
5) Graphical tools: QQ plots, PP plots, and probability-weighted moments.

IV. MODEL DIAGNOSTICS IN EVT

Once a model is selected, it is crucial to assess its fit using diagnostic tools. These tools help determine whether the model adequately represents the extreme tail of the data.

1) Quantile-Quantile (Q-Q) plots: Q-Q plots compare the quantiles of the fitted model's distribution to the observed quantiles of the data. If the points follow a straight line, the model fits well.
2) Probability-Probability (P-P) plots: Similar to Q-Q plots, P-P plots compare the cumulative distribution of the observed data to that of the fitted model.
3) Residual Analysis: Studying the residuals (differences between observed and fitted values) can indicate whether the chosen model fits well. Large residuals in the tail can suggest that the model is not capturing the extreme events accurately.
4) Goodness-of-Fit Tests: Statistical tests like the Anderson-Darling test and Kolmogorov-Smirnov test can be used to assess whether the data follow the fitted EVT distribution.

A. Parameter Stability Checks

1) Hill Plot (for heavy-tailed distributions): Determines stability of the tail index.
2) Threshold Stability Plot: Ensures parameters remain stable over different thresholds.

B. Residual Analysis

1) Examines residuals to check if they follow expected distributions.
2) Can use autocorrelation functions (ACF) to detect dependencies.

C. Return Level Estimation

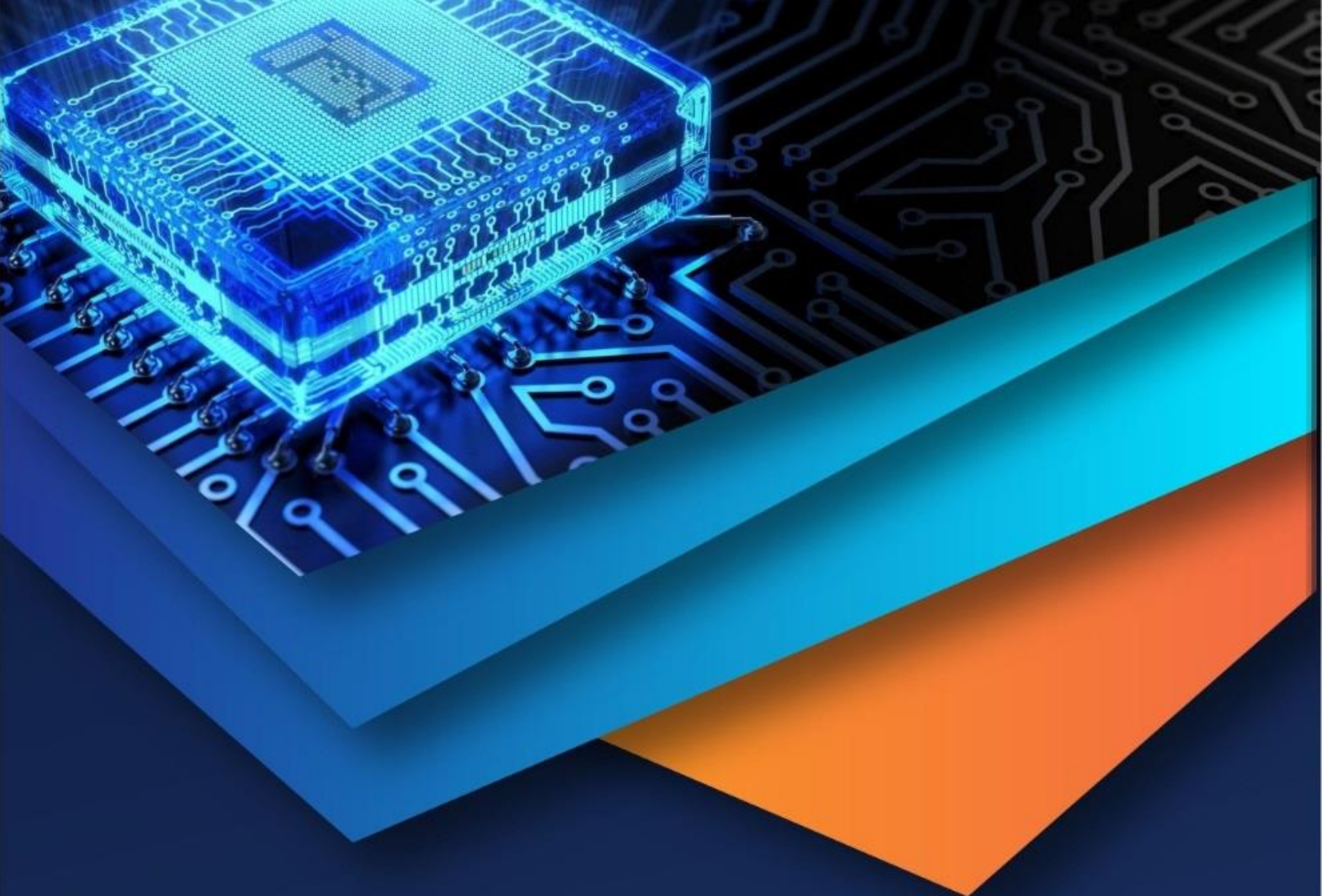
1) Assesses predicted extreme quantiles for different return periods.
2) Uses confidence intervals and bootstrap resampling to validate predictions.

V. CONCLUSION

Model selection and diagnostics are essential steps in the application of Extreme Value Theory. While EVT provides powerful tools for modeling extreme events, it is crucial to ensure that the chosen model adequately fits the tail behavior of the data. Goodness-of-fit tests, graphical diagnostics, and model comparison criteria like AIC and BIC all play significant roles in evaluating and selecting appropriate EVT models. Proper model diagnostics help to avoid misinterpretation of extreme events, which can lead to poor decision-making, particularly in risk management, finance, and environmental sciences. Ultimately, the reliability and accuracy of predictions for extreme events depend on robust model selection procedures and diagnostic checks that ensure the model is a good representation of the underlying data.

REFERENCES

[1] Coles, S. (2001).An Introduction to Statistical Modeling of Extreme Values. Springer.
[2] Beirlant, J., Goegebeur, Y., Teugels, J. L., & Segers, J. (2004).Statistics of Extremes: Theory and Applications. Wiley.
[3] Pickands, J. (1975)."Statistical Inference Using Extreme Order Statistics."The Annals of Statistics, 3(1), 119–131.
[4] Ghosh, S., & Resnick, S. I. (2010)."Modeling Extreme Risks in Finance and Insurance."Handbook of Heavy-Tailed Distributions in Finance. Elsevier.
[5] Rootzén, H., & Kluppelberg, C. (2015)."Statistical Methods for Extreme Value Analysis: Theoretical Results and Practical Applications."Springer.
[6] Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997).Modelling Extremal Events for Insurance and Finance. Springer.
[7] Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004).Statistics of Extremes: Theory and Applications. Wiley.
[8] Davison, A. C., & Smith, R. L. (1990).Models for Exceedances over High Thresholds. Journal of the Royal Statistical Society: Series B, 52(3), 393–442.
[9] McNeil, A. J., & Frey, R. (2000).Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. Journal of Empirical Finance, 7(3-4), 271–300.
[10] Gilleland, E., & Katz, R. W. (2016).extRemes 2.0: An Extreme Value Analysis Package in R. Journal of Statistical Software, 72(8), 1–39.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY