

## Modeling of Extreme Crude Oil Price using the Generalized Pareto Distribution: Brent and West Texas Benchmark Price

Ngozi J. Amachukwu<sup>1,\*</sup>, Happiness O. Obiora-Ilouno<sup>2</sup> and Edwin I. Obisue<sup>3</sup>

<sup>1</sup>Department of Statistics, Nnamdi Azikwe University, Awka, Nigeria  
e-mail: ngoziamachukwu@gmail.com

<sup>2</sup>Department of Statistics, Nnamdi Azikwe University, Awka, Nigeria  
e-mail: ho.oborailouno@unizik.edu.ng

<sup>3</sup>Department of Statistics, Delta State Polytechnic Oghara, Delta State, Nigeria

### Abstract

**Background and objective:** Crude oil is an essential commodity in many countries of the world. This work studies the risk involved in the extreme crude oil price, using the daily crude oil price of the Brent and the West Texas benchmark from year 1990 to 2019.

**Materials and methods:** The Peak Over Threshold (POT) approach of the Generalized Pareto Distribution (GPD) was used to model the extreme crude oil price while the value at risk and the expected shortfall was used to quantify the risk involved in extreme price of crude oil. The GPD, using the Q-Q plot was found to be a good model for the extreme values of the crude oil price.

**Results:** The Value at Risk (VaR) and the Expected Shortfall (ES) calculated at 90%, 95% and 99% with the Maximum Likelihood estimators of GPD parameters and the threshold values were found to decrease with increase in quantile for both benchmark. This shows that risk involved in extreme crude oil price will be borne only by the investors and public.

**Conclusion:** It was also found that the VaR and ES of the Brent are higher than that of West Texas. This implies that it is safer to invest in West Texas crude oil.

### 1. Introduction

Crude oil is a vital commodity widely traded in the world today. Crude oil is a liquid

Received: February 18, 2021; Accepted: April 20, 2021

2010 Mathematics Subject Classification: 62P20.

Keywords and phrases: modeling, extreme, volatility, threshold, generalized Pareto distribution.

\*Corresponding author

Copyright © 2021 the Authors

made up of mixture of hydrocarbon that exists in natural underground reservoirs. The importance of crude oil to economic growth cannot be over emphasized for more detail see [10] and [5]. Many factors have contributed to the volatility of crude oil, according to [3] the world economy, oil inventories, futures markets and political stability are the factor that affect the volatility of oil price. Price of crude oil has been modeled using many methods. [9] applied Extreme Value Theory (EVT) and focusing on the peak over threshold method to the daily returns of crude oil prices in the Canadian spot market between 1998 and 2006. [8] studied the stochastic process of the prices of crude oil, coal and natural gas from 1870 to 1996. [7] studied the relationship between crude oil prices and the difference between the price of crude oil and the price of its products.

Crude oil over the years has been the key drive of global economy growth. There are periods of time when the price of crude oil is relatively stable and other periods when the price can become volatile, changing quickly and by a significant amount leading to price volatility of other energy commodities and as a result understanding the volatility of crude oil price is important. Again, global economy become riskier as a result of high volatility of crude oil price with high Value at Risk (VaR) if extreme behavior of these prices of crude oil are not investigated. This work seeks to model the extreme crude oil price and quantify the risk associated with the extreme of WTI and BRENT crude oil prices from 1990 to 2019.

Extreme Value Theory (EVT) is a statistical analysis used in assessing, modeling and managing catastrophic events. EVT offers a framework for assessing the uncertainties inherent to describing the loss distribution associated with these extreme events [4]. According to [6], EVT plays a significant role in modeling extreme behaviors of various financial risk factors. EVT has show to be a strong tool in extreme event distributions and it is widely used in multi disciplinary area such as Value at Risk estimation [12].

This research concentrates on using the peak over threshold analysis to EVT to modeling and quantifying the risk involved in extreme crude oil price. This will be done by choosing a threshold using the mean residual plot and getting the extreme distribution above the chosen threshold. In addition, fitting the Generalized Pareto Distribution to the extreme distribution and testing the GPD fit. Finally, estimate the Value at Risk and the Expected Shortfall of the GPD model to quantify the risk.

## **2. Research Methodology**

This is a thesis work conducted in August 2019 at Awka and was presented to the Department of Statistics Nnamdi Azikwe University, Awka.

Data Description

The data used is a secondary data retrieved from [www.eia.gov.com](http://www.eia.gov.com). It is the crude oil price of West Texas and Brent from 1990-2019.

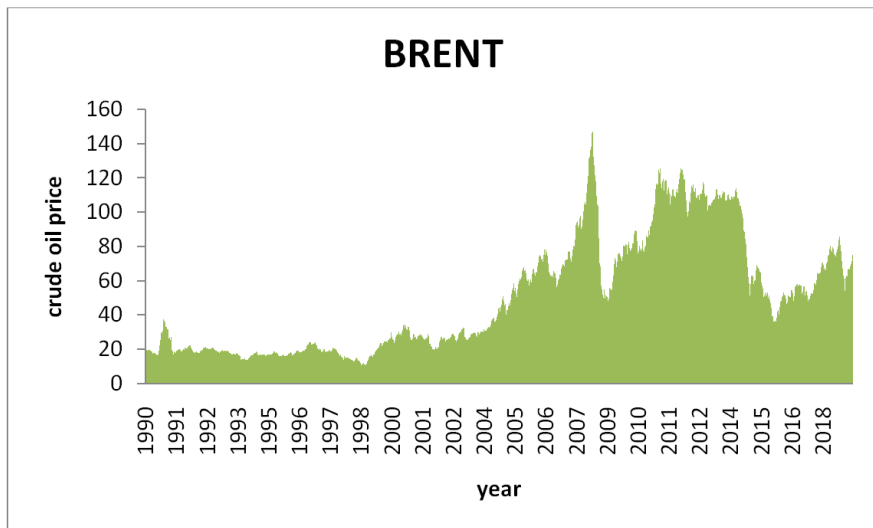


Figure 1. A graphical presentation of Brent crude oil price showing the volatility from 1990-2019.

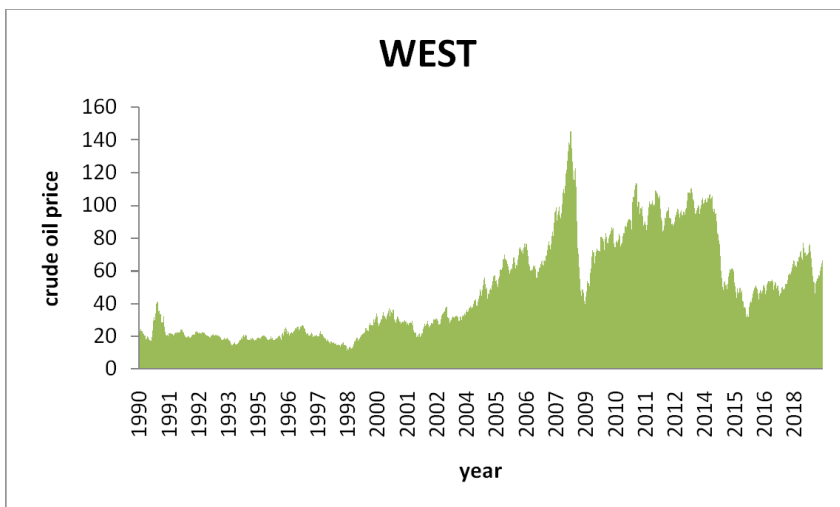


Figure 2. A graphical presentation of West Texas crude oil price showing the volatility from 1990-2019.

From graphical presentation in Figure 1 and Figure 2, it can be observed that there are extreme prices and as a result, the Extreme Value Theory analysis can be applied.

### Method of estimation (POT model)

The EVT is a strong and robust framework for analyzing the behavior of extreme distribution. There are two EVT approaches to sampling extreme events. The first is generalized extreme value (GEV) used in Block Maxima (BM) approach that involves grouping the data into samples/blocks, calculating the maximum observation in each block, fitting the Generalized Extreme Value (GEV) distribution to the maxima of the blocks and estimating the risk measure from the fitted Generalized Extreme Value (GEV) distribution. The other is the generalized Pareto distribution (GPD) used in peak over threshold approach, which involves selecting a threshold, calculating the excess of the observations over the threshold, fitting the Generalized Pareto Distribution (GPD) to the exceedances and computing the measure of risk. According to studies by [11], the estimators from Peak over Threshold approach perform better than the BM approach. Therefore, in this study we considered the POT model to EVT.

Suppose  $Y_1, Y_2, \dots, Y_n$  are independent observations respecting the prices of crude oil. If we assume the length of the selection interval to be  $m$ , then the daily prices can be divided into non-overlapping time intervals of length  $m$ .

The sequence of the extreme maxima will be

$$X_1 = \max(Y_1, Y_2, \dots, Y_m), X_2 = \max(Y_{m+1}, Y_{m+2}, \dots, Y_{2m}), \dots, X_{n/m} = \max(Y_{n-m}, \dots, Y_n).$$

Thus we assume that  $X_1, X_2, \dots, X_{n/m}$  follows the GPD with cumulative distribution function given as follows:

$$F(x) = \begin{cases} 1 - \left(1 - \frac{s}{\sigma}(x - \mu)\right)^{-\frac{1}{s}} & \text{for } s \neq 0, \sigma > 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{for } s = 0, \sigma > 0 \end{cases} \quad (2.1)$$

where  $s > 0$  shape parameter,  $\mu$  threshold and  $\sigma > 0$  scale parameter.

### Selection of Threshold

The choice of threshold is important because a lower threshold results in a large model error with more data and smaller parameter error while high threshold results in smaller model error and a less data with larger parameter error.

An appropriate threshold is required in fitting the GPD model and this can be achieved using the mean residual life plots.

### Mean Residual Plot

The threshold stability property of the GPD means that if the GPD is a valid model for excesses over some threshold  $\mu_0$ , then it is valid for excesses over all thresholds  $\mu_0$ , the expected value of your threshold excesses conditional on being greater than the threshold is

$$E(X - \mu / X > \mu) = \frac{\sigma_{\mu_0 + S\mu}}{1-S}. \tag{2.2}$$

Thus, for all  $\mu > \mu_0$ ,  $E(X - \mu / X > \mu)$  is a linear function of threshold  $\mu$ .

Furthermore,  $E(X - \mu / X > \mu)$  is simply the mean of the excesses of the threshold  $\mu$ , for which the sample mean of the threshold excesses of  $\mu$  provides an estimate. This estimate is obtained by Mean Residual Plot (MRL), a graphical procedure for identifying a suitable high threshold for modeling extremes via GPD. In this plot, a range of candidate values for  $\mu$  the corresponding mean threshold excesses is identified. This mean threshold excess is plotted against  $\mu$ , we then look for the value  $\mu_0$  above which we can see linearity in the plot [1].

### Data Pre-Processing

This is the process of extracting the set threshold exceedances for modeling of the Generalized Pareto Distribution using the selected threshold.

### Fitting the Generalized Pareto Distribution

The most used approach is the maximum likelihood method. It involves estimation of parameters of each of the sampled probability distributions. Once the parameters of GPD are estimated then the GPD fitted distribution is available for further analysis. The maximum likelihood estimates are used because they have several desirable properties, which include consistency, efficiency, asymptotic normality and invariance. This method would be used in this work.

The steps involved in finding the maximum likelihood estimators are as follows.

(1) Write down the likelihood function for the available data. If the likelihood is based on a set of known values  $x_n, 1, 2, \dots, n$ , then the likelihood function will take the form  $f(x_1)f(x_2) \dots f(x_n)$ , where  $f(x)$  is the PDF of the distribution that is to be fitted.

(2) Simplify the algebra by taking the logs

(3) Maximize the log-likelihood function by differentiating the Log-likelihood

function with respect to each of the unknown parameters, and setting the resulting expression(s) equal to zero.

(4) Solve the resulting equation(s) to find the MLEs of the parameters.

Let  $x_1 \dots x_n$  be a random variable of  $n_u$  exceedances of threshold for  $s \neq 0$  and using the above steps, the GPD log likelihood function MLE is gotten as

$$l(\sigma, s, x) = -n \log \sigma - \left(1 + \frac{1}{s}\right) \sum_{i=1}^n \log \left(1 + \frac{s x_i}{\sigma}\right). \quad (2.3)$$

### Model Adequacy

Usually, a statistical model is fitted to data to draw conclusion about some aspect of the population from which the data were obtained. The more accurate the fitted model is, the more reliable these conclusions are likely to be. Since the inference are sensitive to the accuracy of the fitted model, it is important to check that the model fits well. The Probability plots, Quantile plots and a simple histogram plot of the data against the fitted density are used to check the suitability of the fitted GPD to the set of extracted threshold exceedances.

A *probability plot* is a plot of the points

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

where  $x_{(i)}$ ,  $i = 1, \dots, n$  is an ordered sample of the independent observation and  $\hat{F}$  is a candidate model for the true probability distribution  $F$ . the quantity  $\frac{i}{n+1}$  corresponds to the empirical distribution function evaluated at  $x_{(i)}$ . If  $\hat{F}$  is a reasonable model for the true distribution, the points in the probability plot will lie close to the unit diagonal.

A *quantile-quantile plot* is a plot of the points

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

The quantity  $\hat{F}^{-1} \left( \frac{i}{n+1} \right)$  gives a model-based estimate of  $\frac{i}{n+1}$  quantile provided by the candidate distribution  $\hat{F}$  whilst  $x_{(i)}$  itself provides an empirical estimate of the quantile. Again, a well-fitted model would provide points on this plot lying close to the unit diagonal.

Risk Measure

Some of the most frequently used measure of risk in extreme quantile estimation includes value at risk (VaR) and Expected shortfall (ES) and return level. This corresponds to the determination of the value at a given variable exceed with a given probability.

Value at Risk (VaR) of the Generalized Pareto Distribution

Value at Risk (VaR) is used for measuring and assessing risk. VaR is defined as a quantile of the distribution of returns (losses) for an asset or portfolio. This can also be defined as the predicted worst-case loss at a specific confidence level over a certain period. Suppose a random variable  $X$  with a distribution function  $F$  describe negative returns on a certain financial instrument over a certain time horizon. Then VaR can be defined as the  $q$ th quantile of the distribution  $F$ ,

$$VaR_q = F^{-1}(q),$$

where  $F^{-1}$  is the inverse of the distribution ( $q$ ). The inverse of the distribution at a particular probability level is the called quantile.

The Value at Risk is given as

$$VaR_q = \mu + \frac{\sigma}{s} \left( \left( \frac{qN}{N_u} \right)^{-s} - 1 \right), \tag{2.4}$$

where  $N$  is the number of observation in the left tail and  $N_u$  is the number of excess beyond the threshold  $\mu$ .

Expected Shortfall

This is another measure of risk also known as the Conditional Value at Risk. This measures the average value of a loss in an investment portfolio that exceeds the given confidence level. The Expected Shortfall gives the average of the excesses of  $X$  over varying values of the threshold values

The Expected Shortfall is given as

$$\begin{aligned} ES_q &= E[X/X > VaR_q] \\ ES_q &= VaR_q + E[X - VaR_q/X > VaR_q] \\ ES_q &= \frac{VaR_q + \sigma - s\mu}{1-s}. \end{aligned} \tag{2.5}$$

### 3. Results

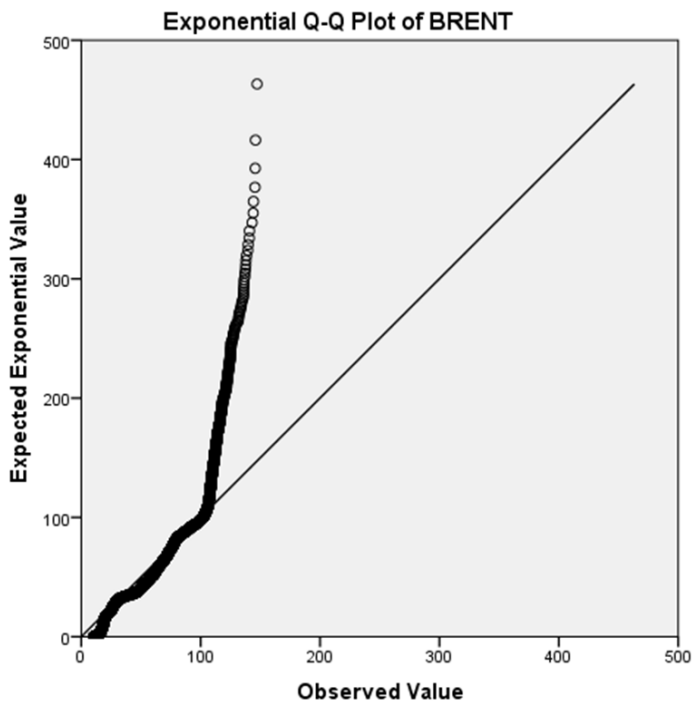
The peak over threshold model of the extreme value analysis was carried on the crude oil price of West Texas and Brent crude oil.

#### Descriptive Analysis

The descriptive analysis of the data were done using SPSS and the result were obtained as follows: for Brent-mean (49.2), standard deviation (33.2), skewness (0.74) and kurtosis (-0.70). West Texas-mean (47.5), standard deviation (29.6), skewness (0.74) and kurtosis (-0.59). The skewness indicates that the data distribution is thin-tailed distribution.

#### Preliminary Data Analysis

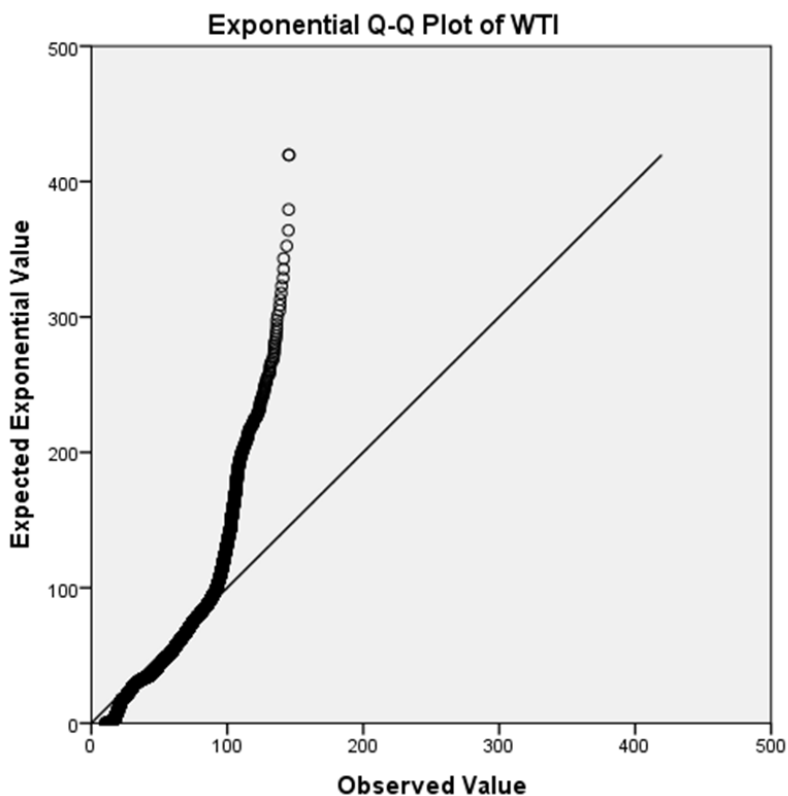
In EVT and application, the Q-Q plots are used to measure the tailedness of a distribution. Q-Q plot is plotted against the exponential distribution and if the sample comes from hypothesized distribution then the Q-Q plot is linear.



**Figure 3.** A Q-Q plot of the Brent daily crude oil price against standard exponential values indicating a thin tailed distribution.



Notice that there is a convex departure from the straight line in Figure 3 indicating that the crude oil price of Brent oil is a thin-tail distribution and as such good for Extreme value analysis.

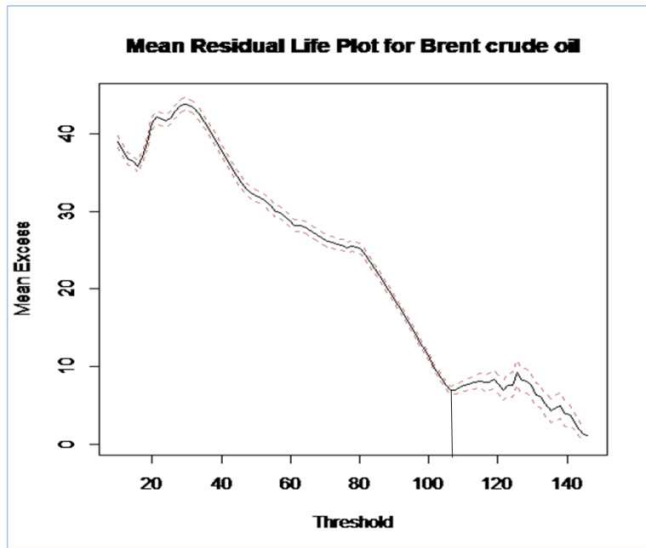


**Figure 4.** A Q-Q plot of the West Texas daily crude oil price against standard exponential values indicating a thin-tailed distribution.

Notice that there is a convex departure from the straight line in Figure 4 indicating that the crude oil price of Brent oil is a thin-tail distribution and as such good for Extreme value analysis.

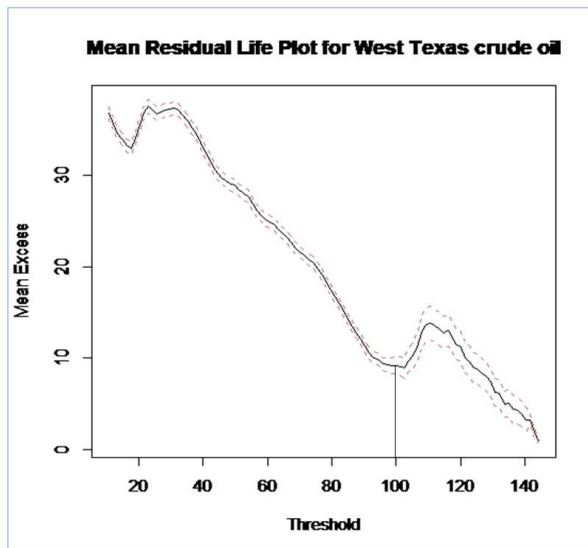
Determination of appropriate threshold

The size of the sample in the POT model is determined by the choice of threshold. Too high threshold will leads to few observations over the threshold and a large variance while too low threshold leads to more observation as extreme there by making the asymptotic assumption of EVT less valid. The mean residual life plot proposed by [2] was adopted to choose the threshold.



**Figure 5.** A mean residual plot of the Brent daily crude oil price, plotted against various thresholds.

From Figure 5, \$106 is taken as the threshold of the Brent crude oil price in order to have linearity above the threshold selected.

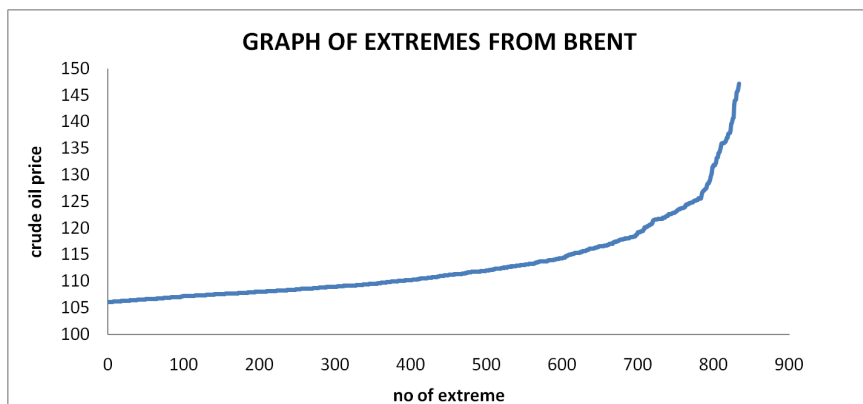


**Figure 6.** A mean residual plot of the West Texas crude oil price, plotted against various thresholds.

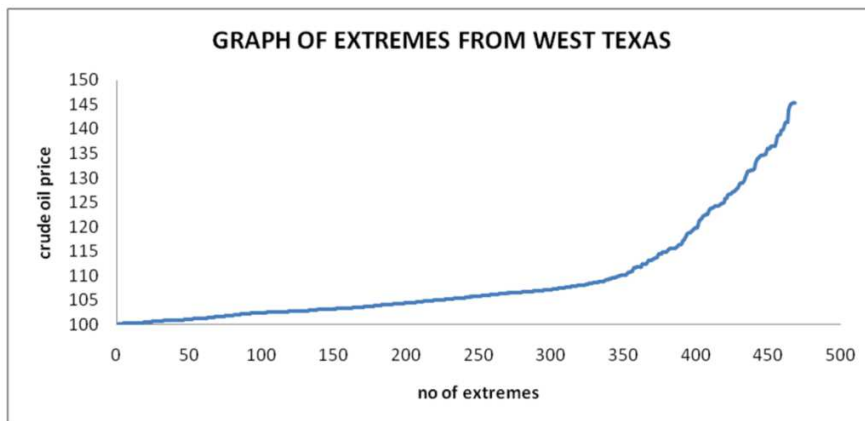
From Figure 6, \$100 is taken as the threshold of the West Texas crude oil price in order to have linearity above the threshold selected. To show that linearity occurs, the graph of crude oil extreme against number of extreme where plotted in Figure 7 and Figure 8.

Data pre-processing

This is a straightforward process used in threshold-based analysis for identifying extremes. Using the threshold \$100 and \$106 for West Texas and Brent respectively we have identified 833 observations and 468 observations as being the extreme for Brent and West Texas respectively.



**Figure 7.** A plot showing the linearity of the extreme above the threshold \$106 for the Brent crude oil price.



**Figure 8.** A plot showing the linearity of the extreme above the threshold of the West Texas crude oil price.

### Fitting the Generalized Pareto Distribution

In this work, the maximum likelihood was used to obtain the parameters of the GPD. Using the `gpd.fit` in R package, we got the results as follows:

**Table 1.** Result of ML Estimators.

Estimated	Brent	West Texas
Log likelihood value	24466.828	1500.529
Scale parameter $\sigma$	6.796	7.755
Standard error	0.352	0.580
Shape parameter $s$	0.045	0.158
Standard error	0.039	0.060

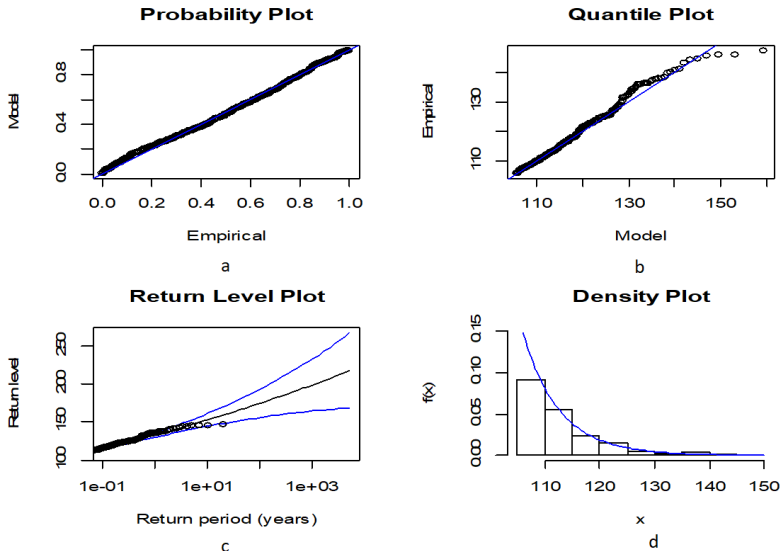
From Table 1,  $\hat{s} = 0.045$ ;  $\hat{\sigma} = 6.796$  for Brent and  $\hat{s} = 0.045$ ;  $\hat{\sigma} = 6.796$ .

Notice that the scale parameter of both crude oil is greater than one, indicating that the distribution was stretched (spread out). Also the shape parameters greater than zero shows that the distributions are polynomial tail decreasing distributions.

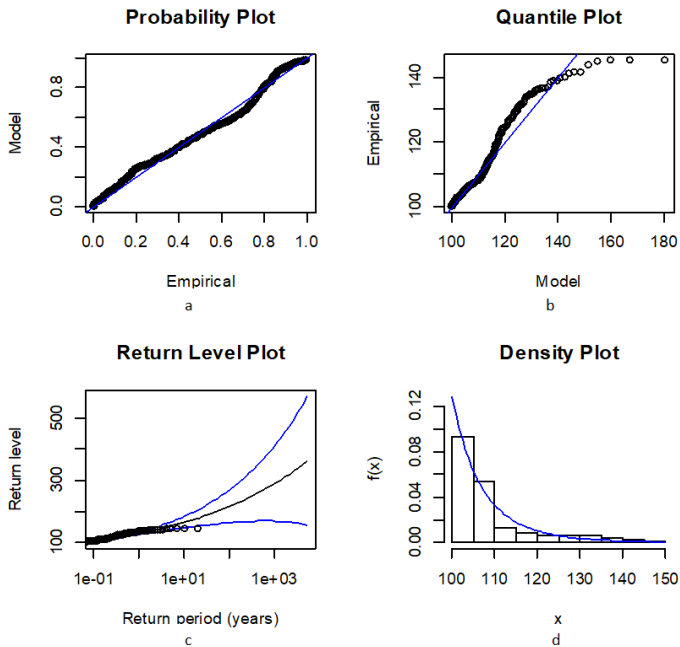
### Model Adequacy

The Probability plots, Quantile plots and simple histogram plot of the data against the fitted density are used to check the suitability of the fitted GPD to the set of extracted threshold exceedances.

Notice, the points in Figure 9a and b lies close to the unit diagonal and the curve in Figure 9c touches the bars. These show that Generalized Pareto Distribution model is a reasonable model for the extreme distribution of the Brent crude oil.



**Figure 9.** Diagnostic plots indicating the goodness-of-fit of the GPD to the West Texas crude oil price extremes.



**Figure 10.** Diagnostic plots indicating the goodness-of-fit of the GPD to the West Texa crude oil price extremes.

Notice, the points in Figure 10a and b lies close to the unit diagonal and the curve in Figure 10c touches the bars. These show that Generalized Pareto Distribution model is a reasonable model for the extremes of the West Texas crude oil.

#### Determination of Value at Risk and Expected Shortfall

The Value at Risk and Expected Shortfall were calculated using equation (2.4) and (2.5) respectively at 90%, 95% and 99% quantiles and the result as follows in Table 2 and Table 3.

**Table 2.** Value at risk and Expected shortfall estimates for Brent crude oil.

QUANTILE	VaR	ES
90%	92.30	98.77
95%	91.97	98.43
99%	91.72	98.16

**Table 3.** Value at risk and Expected shortfall estimates for West Texas.

QUANTILE	VaR	ES
90%	83.01	89.03
95%	82.74	88.71
99%	82.53	88.46

Notice from Table 2 and Table 3 that the Value at Risk decreases with increase in quantile. Again, the Expected Shortfall decreases with increase in quantile. This means that the predicted worst-case loss at a specific confidence level over a certain period decreases as the quantile increases. This can also imply that the risks are borne by the investors and the general public at large. Notice that VaR and ES values in Table 2 are high than that of Table 3, it indicates that it is riskier to invest in Brent crude oil market than West Texas crude oil market.

## **4. Discussion, Conclusion and Recommendation**

### Discussion

From the threshold-based analysis of EVT carried out on crude oil price obtained from the Brent and West Texas crude oil market from the year 1990 to 2019, the

following result were obtained: The descriptive analysis shows that the mean of Brent and West Texas are \$49.2 and \$47.5 respectively and skewness are the same with value of 0.73. This shows that the distribution is tail distribution thereby making application of Extreme Value Theory valid. In determination of the threshold, the mean residual life plot over various thresholds was done. The mean residual life shows that thresholds of \$106 and \$100 for Brent and West Texas respectively will cause linearity in the extremes above the threshold. In fitting the GPD model, the maximum likelihood estimate of the scale parameters for both crude oil was found to be greater than one, indicating that the distribution was stretched (spread out). Also the shape parameters greater than zero shows that the distributions are polynomial tail decreasing distributions. The GPD, using the probability plot, Q-Q plot and Histogram was found to be a good and reasonable fit for the distribution extreme crude oil price. The estimated VaR and ES values at 90%, 95% and 99% quantile were found to decrease with increase in quantiles. This indicates that the predicted loss decrease with increase confidence interval. The estimated VaR and ES values of the Brent were found also to be higher than the West Texas, indicating its more risk for investor to invest in Brent than in West Texas. More crude oil benchmark would have been better for our analysis but due to the challenges involved in obtaining the data, we were limited to West Texas and Brent benchmark.

### Conclusion

From the result of the analysis, it has been shown that GPD is a reasonable fit for the distribution of extreme crude oil prices and it is a good statistical distribution for fitting the tails of unknown distribution. Based on the estimated VaR and ES values, it shown that the risk involved in crude oil price volatility is borne by the investors and public and it is safer to invest in crude oil in West Texas crude oil.

### Significance statement

This study discovers and quantifies the risk involved in the volatility behaviour of the crude oil price of WTI and Brent crude oil. This study will help the researcher to provide to energy risk managers and investors exact value at risk when investing in the crude oil of Brent and West Texas oil. It provides researcher with insight for further study on more application of Extreme Value Theory.

### **References**

- [1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, London: Springer-Verlag, 2001. <https://doi.org/10.1007/978-1-4471-3675-0>

- 
- [2] A.C. Davison and R.L. Smith, Models for exceedances over high thresholds, *J. R. Stat. Soc. B* 52(3) (1990), 393-442. <https://doi.org/10.1111/j.2517-6161.1990.tb01796.x>
- [3] S. Huang, H. An, S. Wen and F. An, Revisiting driving factors of oil price shocks across time scales, *Energy* 139 (2017), 617-629. <https://doi.org/10.1016/j.energy.2017.07.158>
- [4] R.J. Karobia, Modeling extreme claims using generalized Pareto distributions family in an insurance company, *Thesis paper*, Presented to the School of Mathematics, University of Nairobi, 2015.
- [5] G. Lalude, Importance of oil to the global community, *Global Journal of Human-Social Science* 15(1) Version 1.0 (2015). Retrieved from <https://socialscienceresearch.org/index.php/GJHSS/article/view/1524>
- [6] A.J. McNeil, R. Frey and P. Embrechts, *Quantitative Risk Management*, New Jersey: Princeton University Press, 2005.
- [7] A. Murat and E. Tokat, Forecasting oil price movements with crack spread futures, *Energy Economics* 31 (2009), 85-90. <https://doi.org/10.1016/j.eneco.2008.07.008>
- [8] R.S. Pindyck, The long-run forecasting of energy prices, *The Energy Journal* 20 (1999), 1-27. <https://doi.org/10.5547/ISSN0195-6574-EJ-Vol20-No2-1>
- [9] F. Ren and D.E. Giles, Extreme value analysis of daily Canadian crude oil prices, *Applied Financial Economics* 20(12) (2010), 941-954. <https://doi.org/10.1080/09603101003724323>
- [10] T.M. Rybezyński, *The Economics of Oil Crisis*, London: Macmillan Press, 1996.
- [11] L. Xu and Q. Zhang, Modeling agricultural catastrophic risk, *Agriculture and Agricultural Science Procedia* 1 (2010), 251-257. <https://doi.org/10.1016/j.aaspro.2010.09.031>
- [12] X. Zhao, Extreme value modeling with application in finance and neonatal research, *PhD thesis*, The University of Canterbury, 2010. Retrieved from <http://ir.canterbury.ac.nz/bitstream/10092/4024/1/thesisfulltext.pdf>
- 

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted, use, distribution and reproduction in any medium, or format for any purpose, even commercially provided the work is properly cited.

---