Modelling Extreme Changes in Crude Oil Prices using Time Series and Extreme Value Theory

Brian Niall MacCarvill

Supervisors

Dr Zak Varty and Dr Euan McGonigle

IMPERIAL

Department of Mathematics
Imperial College London

September 11, 2024

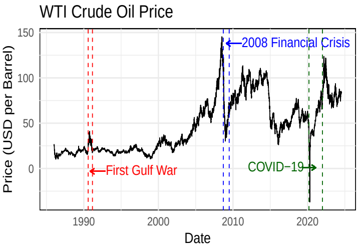
Presentation Outline

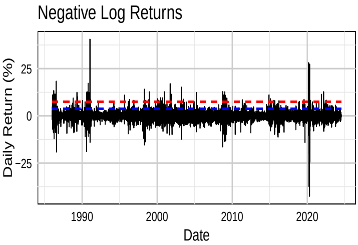
Introduction, Motivation and Data

Motivation

- Changes in crude oil prices can have far-reaching effects on the global economy (Bollerslev, 1986).
- The cost of goods and services can fluctuate in response to changes in oil prices, and higher inflation is linked to increased oil price volatility (Oyuna and Yaobin, 2021).
- Given its significance, analysing extreme changes in crude oil movement can be crucial in future planning.
- Accurate predictions of extreme price movements enables more informed decisions about investments, production levels and supply contracts.

The Data





- The 2 datasets used were the West Texas Intermediate (WTI) and Brent crude oil weekday closing price between 1986 and 2024.
- WTI crude oil is found in North America while Brent crude oil is found in the North Sea in Europe.
- Both WTI and Brent crude oil are used as key benchmarks for oil prices.
- There is volatility clustering, in which periods of high volatility tend to be followed by periods of low volatility in crude oil price returns (Zavadska et al., 2020).

Terminology

Assume that losses are the positive values from a series X with distribution function F.

Value at Risk (VaR): The VaR, $\text{VaR}_\alpha(X)$, is defined as:

$$\text{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

Accurate predictions of VaRs enable more informed decisions about investments, production levels and supply contracts.

Return Levels (z_p): Is the value that satisfies $F(z_p) = 1 - p$. The return level z_p is equivalent to VaR_p and so they are used interchangeably.

Expected Shortfall (ES): ES, also known as Conditional VaR, measures the expected loss given that the loss exceeds the VaR. It is defined as:

$$\text{ES}_\alpha(X) = \mathbb{E}[X | X > \text{VaR}_\alpha(X)]$$

ES can be calculated as a way of testing how well assumption for the distribution of the values over the VaR threshold are.

Methods Overview

This project seeks uses time series models and Extreme Value Theory (EVT) to predict the VaR and ES forecasts at future time steps greater than 1 in a novel way.

The fundamental assumptions used in the project were that:

1. The log daily returns of crude oil can be modelled and forecasted using a time series with autocorrelated variance.
2. The largest values of the estimated residuals, from the time series, can be modelled using an extreme value distribution called a Generalised Pareto Distribution (GPD)

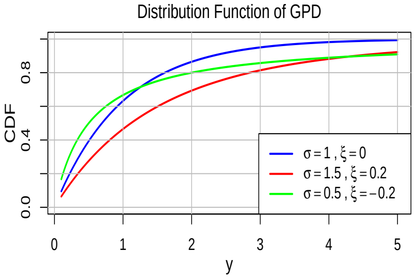
These assumptions are used to estimate and test the VaR and ES estimates for future values.

Generalised Pareto Distribution (GPD)

Generalised Pareto Distribution (GPD)

The Generalised Pareto Distribution (GPD) has the distribution function:

$$G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\beta}\right)_+^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta}\right)_+, & \text{if } \xi = 0 \end{cases}$$



Pickands (1975) proved that for a large class of distributions F, there exist parameters $\xi \in \mathbb{R}$ and $\beta \in \mathbb{R}^+$ such that:

$$\lim_{u \to x_\infty} \inf_{0<\beta<\infty} \sup_{0 \leq y \leq \infty} \left| F_u(y) - G_{\xi,\beta}(y) \right| = 0$$

where $x_\infty = \sup\{x : F(x) < 1\}$ & $F_u(y) = \frac{F(y+u) - F(u)}{1 - F(u)}$. β is the "scale" parameter and ξ the "shape" parameter.

Upper Quantiles and Conditional Expected values

For an ordered data set, $z_n \leq z_{n-1} \leq \ldots \leq z_1$ and $k < n$. An estimate of the return level z_q can be calculated using a GPD, by fitting a subset $\{z_1, \ldots, z_k\}$, to obtain the MLEs for parameters β and ξ.

For $q > 1 - k/n$, the return level estimate, $\hat{z}_q$, can be estimated as:

$$\hat{z}_q = z_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{1-q}{k/n} \right)^{-\hat{\xi}} - 1 \right).$$

For $q > 1 - k/n$, the ES_q, can also be estimated as:

$$\begin{aligned} \hat{S}_q &= \mathbb{E}[Z \mid Z > \hat{z}_q] \\ &= \frac{\hat{z}_q}{1-\hat{\xi}} + \frac{\hat{\beta} + \hat{\xi} z_{k+1}}{1-\hat{\xi}}. \end{aligned}$$

Time Series Modelling

GARCH models are used in time series analysis to model and forecast financial volatility. Introduced by Bollerslev (1986), GARCH models allow for long-memory in the volatility of a time series. In this project GARCH models are used to estimate the residuals Z and forecast future means and variances.

Define a AR(p)-GARCH(q,r) time series:

$$X_t = \mu_t + \sigma_t Z_t$$

in which $Z_t \sim \mathcal{N}(0,1)$ is an i.i.d white noise process,

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} \quad \text{and} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1, \ldots, q$ and $\beta_i > 0$ for $i = 1, \ldots, r$ are the parameters and $\varepsilon_t = X_t - \mu_t$.

Forecasting

The mean value forecast, denoted as $\hat{\mu}_t(h)$, for μ_{t+h}, given time series values $X_1, \ldots, X_t$, can be calculated recursively as follows:

$$\hat{\mu}_t(h) = \hat{\phi}_0 + \sum_{i=1}^{p} \hat{\phi}_i \hat{\mu}_t(h-i)$$

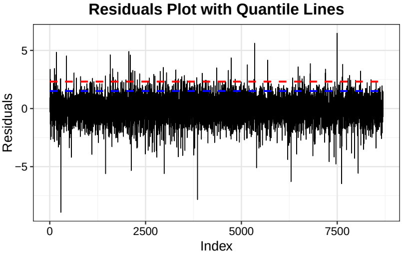
where $\hat{\mu}_t(h-i) = X_{t+h-i}$ if $h \leq i$.

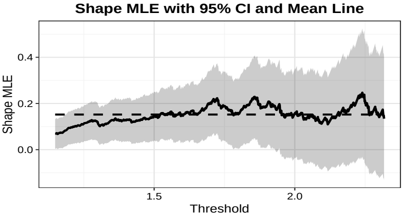
The sigma forecast, denoted as $\hat{\sigma}_t(h)$, for σ_{t+h}, can be calculated recursively as follows:

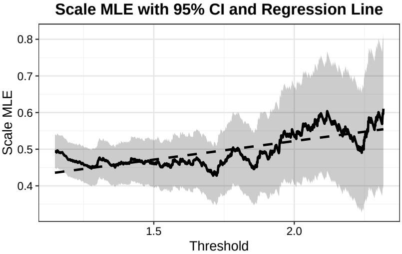
$$\hat{\sigma}_t^2(h) = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i \hat{\varepsilon}_{t+h-i}^2 + \sum_{j=1}^{r} \hat{\beta}_j \hat{\sigma}_t^2(h-j)$$

where $\hat{\sigma}_t^2(h-j) = \hat{\sigma}_{t+h-j}$ if $h \leq j$ and $\hat{\varepsilon}_\tau^2 = \begin{cases} (X_\tau - \hat{\mu}_\tau)^2, & \text{if } \tau \leq t \\ \hat{\sigma}_t^2(\tau - t), & \text{otherwise.} \end{cases}$

WTI Crude Oil Residuals and Diagnostic Plots Example







$$\{\hat{z}_m,\ldots,\hat{z}_{n+1}\} = \left\{\frac{x_m-\hat{\mu}_{m-1}(1)}{\hat{\sigma}_{m-1}(1)},\ldots,\frac{x_{n+1}-\hat{\mu}_n(1)}{\hat{\sigma}_n(1)}\right\}$$

- The MLEs for the time series parameters was calculated using the entire data set.
- $\hat{\mu}_{p-1}(1)$ and $\hat{\sigma}_{p-1}(1)$ are the 1-step forecasts given values from time $p-n$ to time $p-1$.
- The Scale (β) and Shape (ξ) parameters should be linear and constant respectively with increasing u, if the Residuals follow a GPD.

Novel Methodology

Equation to estimate VaRs and ESs

The future forecast for the upper quantiles using EVT and time series modelling is calculated as:

$$\hat{x}_q^t(h) = \hat{\mu}_t(h) + \hat{\sigma}_t(h)\hat{z}_q$$
$$= \hat{\mu}_t(h) + \hat{\sigma}_t(h)\left(z_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-q}{k/n}\right)^{-\hat{\xi}} - 1\right)\right)$$

for $q > 1 - \frac{k}{n}$.

The future forecast for the expected value for points over a quantile using EVT and time series modelling is calculated as:

$$\hat{S}_q^t(h) = \hat{\mu}_t(h) + \hat{\sigma}_t(h)E[Z|Z > \hat{z}_q]$$
$$= \hat{\mu}_t(h) + \hat{\sigma}_t(h)\left(\frac{\hat{z}_q}{1-\hat{\xi}} + \frac{\hat{\beta} + \hat{\xi}z_{k+1}}{1-\hat{\xi}}\right)$$

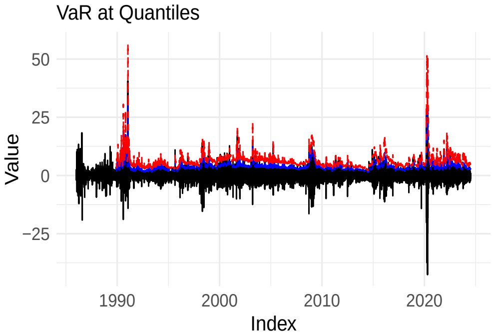
for $q > 1 - \frac{k}{n}$.

Rolling Back testing Upper Quantiles Estimates

- **Input:**
  - Dataset: $x_1, x_2, \ldots, x_m$
  - Look-back period: n days
  - Number of large values: k values
  - Future day for testing: h days
  - Time Series Model Specifications: p,q and r
  - Quantile: q
- **Procedure:**
  - For each time step t from n to $m-h$:
    - Define the data window $X_{(t-n+1):(t)}$ as $\{x_{t-n+1}, x_{t-n+2}, \ldots, x_t\}$
    - Calculate MLEs, $\hat{\theta} = (\hat{\phi}_0, \ldots, \hat{\phi}_p, \hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_q, \hat{\beta}_1, \ldots, \hat{\beta}_r)$, for $X_{(t-n+1):(t)}$
    - Let $S = \left\{ \frac{x_{t-n+i} - \hat{\mu}_{t-n+i}}{\hat{\sigma}_{t-n+i}} : i = 1, \ldots, n \right\}$
    - Let S_k be the set of the k largest elements in S.
    - Fit S_k to a GPD to get MLEs for $\hat{\beta}$ and $\hat{\xi}$
    - Calculate $\hat{\mu}_t(h)$,$\hat{\sigma}_t(h)$, $\hat{z}_\mathrm{q}$
    - Calculate $\hat{x}_\mathrm{q}^t(h) = \hat{\mu}_t(h) + \hat{\sigma}_t(h)\hat{z}_\mathrm{q}$
- **Output:**
  - Return the set of predicted VaR quantiles $\{\hat{x}_\mathrm{q}^t(h) : t = n, \ldots, m-h\}$

Results

AR(1)-GARCH(1,1) Model for WTI Log Daily Returns ($n = 1000$, $k = 100$, $h = 1$) VaR estimate

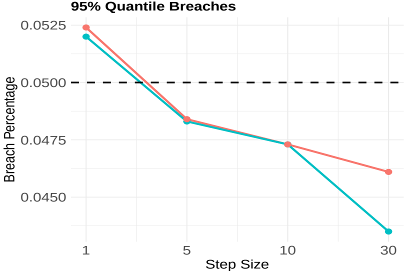


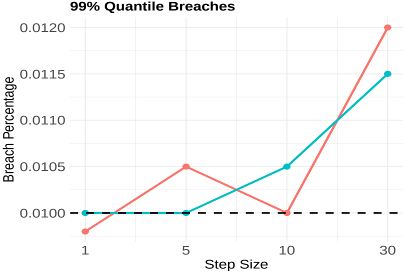
Quantiles: Blue line = 95% estimate, Red line = 99% estimate.

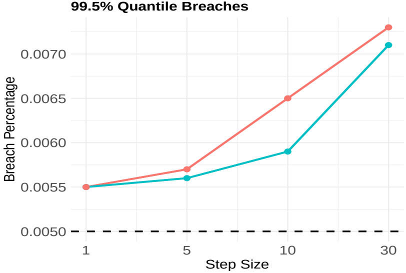
Observations:

- The estimates change quickly with immediate volatility.
- The 95% and 99% quantile estimates were breached 5.2% and 0.98% of the time respectively.
- For increasing steps ahead predictions the immediate volatility is less influential on the estimates.

Value at Risk Results ($n = 1000$, $k = 100$)

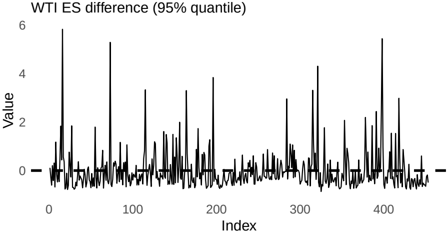


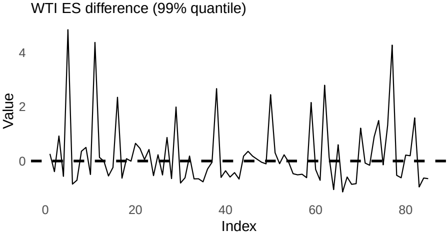




- The red line represents Brent Crude Oil, while the blue line represents WTI Crude Oil.
- The 95% and 99.5% quantile estimate tends to overestimate and underestimate respectively the actual values for increasing step size..
- Results were significantly worse for $n = 500$ and $k = 50$.
- Similar results were found for an AR(0)-GARCH(1,1) model.

AR(1)-GARCH(1,1) Model for WTI Log Daily Returns ($n = 1000$, $k = 100$, $h = 1$) ES estimate





The set used to test the conditional expected values estimates is:

$$\left\{ \frac{x_{t+h} - \hat{S}_q^t(h)}{\hat{\sigma}_t(h)} : t \in \{n, \dots, m-h\}, x_{t+h} > \hat{x}_q^t(h) \right\}$$

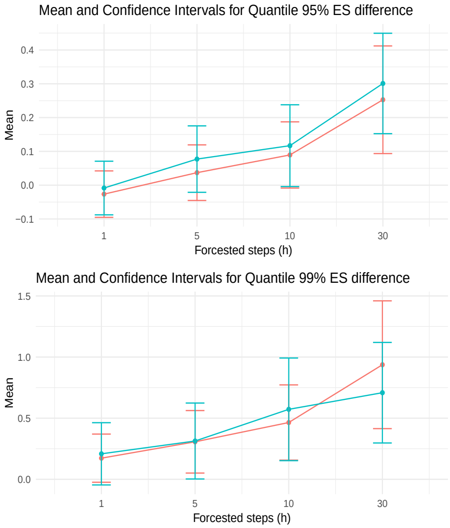
in which $\{\hat{S}_q^t(h) : t = n, \dots, m-h\}$ is calculated using the same rolling back testing.

- **(Top plot)** mean $= -0.01$ and var $= 0.74$.
- **(Bottom plot)** mean $= 0.21$ and var $= 1.43$.

Expected Shortfall Results

Plots showing the mean and 95% confidence intervals for the difference in ES estimates and residuals at the 95% and 99% quantiles and step sizes using an AR(1)-GARCH(1,1) model, with $n = 1000$ and $k = 100$:

- The means and confidence intervals grow further away from 0 as the step size increases.
- For time steps of 10 and 30 the data are not within a 95% confidence interval of 0.



Conclusion and Discussion

Conclusion

- No significant evidence that incorporating an AR(1) mean component significantly enhances model accuracy over a constant mean (AR(0)).
- VaR estimates remain relatively accurate across both datasets, although accuracy declines as the forecasting horizon extends.
- ES estimates consistently fall below empirical shortfalls, particularly at the 99% and 99.5% quantiles and for larger step sizes, suggesting potential underestimation of risk in extreme scenarios.
- Future research could explore alternative models or additional components to improve long-term forecasting accuracy and better capture extreme risk.

Discussion

- McNeil and Frey (2000) recommended a Monte-Carlo method to obtain VaR estimates for longer time horizons than 1.
- This project found that the novel method used to obtain VaR estimates for longer time horizons than 1 preformed well empirically compared to the results in McNeil and Frey (2000).
- The average difference between breaches and ES estimates were near zero for small steps, supporting McNeil and Frey (2000)'s conclusion that ES is a robust risk measure.
- For larger steps, ES estimates were consistently smaller than actual values, suggesting that adjustments could improve ES accuracy.
- Further research on the theoretical properties and empirical comparisons with existing methods using diverse data could provide deeper insights.

References

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3):271–300. Special issue on Risk Management.

Oyuna, D. and Yaobin, L. (2021). Forecasting the crude oil prices volatility with stochastic volatility models. *Sage Open*, 11(3):21582440211026269.

Pickands, J. I. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Zavadska, M., Morales, L., and Coughlan, J. (2020). Brent crude oil prices volatility during major crises. *Finance Research Letters*, 32:101078.

Thank You