An Extreme Value Analysis of Bid-Ask Spreads

Jake Morrison Mc Leroth

24913693

Project presented in the partial fulfilment
of the requirement for the course
Extreme Value Theory
November 2025

Lecturer: Prof. T. De Wet

PLAGIARISM DECLARATION

1. Plagiarism is the use of ideas, material and other intellectual property of another's work and to present it as my own.
2. I agree that plagiarism is a punishable offence because it constitutes theft.
3. Accordingly, all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.
4. I also understand that direct translations are plagiarism.
5. I declare that the work contained in this project, except otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this project or another project.

24913693	
Student number	Signature
J.M. Mc Leroth	28 November 2025
Initials and surname	Date

TABLE OF CONTENTS

SECTION 1

INTRODUCTION

Bid-ask spreads play a central role in financial markets. They are a proxy measurement for liqudity, and thus large increases in spreads impacts on traders. Understanding the magnitude and frequency of such shocks is therefore critical for effective liquidity risk management.

Extreme Value Theory (EVT) provides a framework for modelling the behaviour of extreme observations. However, spread may not be stationary, violating the assumptions required for EVT. To address this, we analyse the log-returns of the relative spread, which exhibit stationarity, making them suitable for EVT modelling.

The goal of this project is to quantify the severity of extreme liquidity shocks by estimating high quantiles of spread changes. These results offer important insights into the liquidity risks faced by investors and highlight the importance of managing liquidity risk.

SECTION 2

BID-ASK SPREADS

2.1 WHAT ARE BID-ASK SPREADS?

In the world of investing, market participants are regularly buying and selling financial instruments. As explained in Maginn *et al.* (2007), this is executed by two main order types: market orders and limit orders. A market order is the order to buy or sell a security at the best available price. The idea here is to execute the trade as quickly as possible regardless of the price. A limit order executes the trade if the price is above or below a certain level. This ensures that the trade will be executed at a favourable price if it goes through; however, there is no guarantee that a buyer or seller will be found. As a result, limit orders often have an expiry date. Limit orders give rise to bid and ask prices. The market bid is the highest price that the security is willing to be bought at, and the marked ask is the lowest price the security is willing to be sold at. These are in the form of a buy limit and a sell limit order, respectively. Hence, if you are wanting to sell a security, the market bid price is the best price at which you can sell immediately. Mathematically, we define the bid price and the ask price at time t as P_t^b and P_t^a. The difference between the two values is an important measure, known as the bid-ask spread, and can be calculated as follows (Affleck-Graves *et al.*, 2000):

$$S_t = P_t^a - P_t^b$$

The spread is the difference between the best price you can buy at and the best price you can sell at. This value is always positive (Tsay, 2005), which intuitively makes sense since sellers will always want the highest price possible and buyers the lowest. It is an important measure, as it gives an indication of how liquid a security is. Liquidity is how easy an instrument is to trade (i.e., buy and sell quickly). As Harris (2002) puts it, the spread is what ”impatient traders pay for immediacy”. Liquid markets with tighter spreads make for easier and more efficient trading. When spreads are large, it is harder to execute trades, as you will have to settle for an unsatisfactory price. Hence, the modelling of bid-ask spreads is useful for traders to evaluate liquidity risk. Liquidity can dry up in times of financial distress. Hence, this is not just of interest to investors but also to banks and regulators (Muela *et al.*, 2017).

2.1.1 Calculating Spreads

The issue with the previous measure is that it is not comparable across securities due to each security having a different price. Consider the following example: Two stocks, A and B, trade at R100 and R2, respectively. The bid and ask prices for A are R99 and R101. The bid and ask prices for B are R1 and R3. In both cases, the spread is R2. This is misleading, as the spread of A is only 2% of the price, whereas for A it is 100% of the price. The liquidity picture is thus very different for A and B; however, you would not be able to tell that by looking at only the spread. Hence, we introduce the relative spread (Muela *et al.*, 2017):

$$\begin{aligned} S_t^{rel} &= \frac{P_t^a - P_t^b}{0.5(P_t^a + P_t^b)} \\ &= \frac{P_t^a - P_t^b}{P_t^{mid}} \end{aligned}$$

where $P_t^{mid} = 0.5(P_t^a + P_t^b)$ is the mid price. This measure is now robust to the different magnitudes of asset prices. The extreme value methodology used in this project requires data to be stationary. Hence, at the risk of the spread data not being stationary, we introduce another measure:

$$\begin{aligned} S_t^{log} &= \Delta ln(S_t^{rel}) \\ &= ln(S_t^{rel}) - ln(S_{t-1}^{rel}) \\ &= ln(\frac{S_t^{rel})}{S_{t-1}^{rel}}) \end{aligned}$$

which can be easily converted back to the original S_t^{rel} values:

$$S_t^{rel} = S_{t-1}^{rel} \exp(S_t^{log})$$

This is analogous to using log return when looking at a price series. Accordingly, we will refer to this measure as the log return spread.

2.2 BID-ASK DATA

2.2.1 Bid and Ask Prices

The data used in this analysis are the daily closing bid and ask prices of FirstRand LTD. The data ranges from 03/08/1999 to 12/08/2025. The closing bid, ask and market prices are shown:

Figure 2.1: FirsRand Prices and Quotes from 2023

Other than FirstRand being a large stock, liquid stock on the Johannesburg Stock Exchange (JSE), the choice is fairly arbitrary since analysing multiple data sets is out of the scope of this assignment. That is to say that this methodology can be used on any stock(s) on the JSE (or any other index), or even any financial security with bid and ask prices.

2.2.2 Data Preprocessing

The preprocessing was done in the following steps:

1. Assign rows with $P_t^a \leq P_t^b$ as NA (i.e. rows with non-positive spreads).
2. Assign rows with any clear price mistakes as NA.
3. Calculate spreads S_t and relative spreads S_t^{rel}.
4. Compute log-returns of spreads S_t^{log}.
5. Remove rows with NA values.

This resulted in dropping 25 rows. The reason for assigning NA values to the rows with mistakes was to ensure that any S_t^{log} values correspond to daily changes. Dropping the values first would result in some of the values corresponding to changes over two or more days. An example of a clear price mistake was a row where the price was 522.10 yet the ask and bid prices were 30.66 and 28.24 - this is clearly a mistake in the data as the quotes would not be this low.

2.3 BID-ASK SPREADS: AN EXPLORATORY DATA ANALYSIS

In this section we give an empirical analysis of the data. The summary statistics are given:

Table 2.1: Summary Statistics for Spreads and Relative Spreads

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Spreads	0.800	2.000	4.000	6.705	8.070	109.000
Relative Spreads (%)	0.01153	0.08304	0.20942	0.32696	0.43557	9.23130

Table 2.1 summarises the empirical distributions of the spread and relative spread data. As can be seen, the difference between the third quantile and the maximum is much larger than that of the minimum and the first quartile. Hence, there seems to be a heavy tail on the right.

2.3.1 Bid-Ask Spreads

The time series and autocorrelation functions (ACF) of the spread data is plotted:

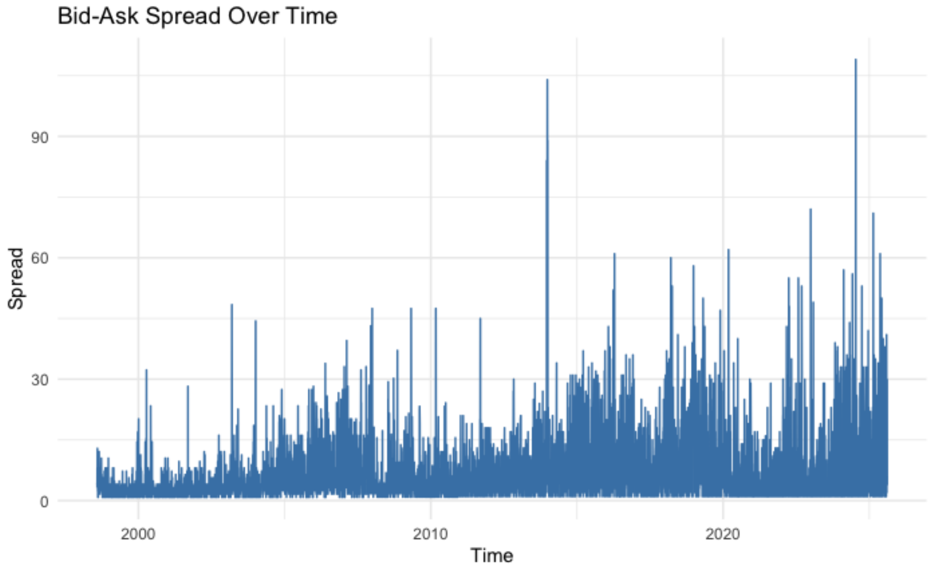
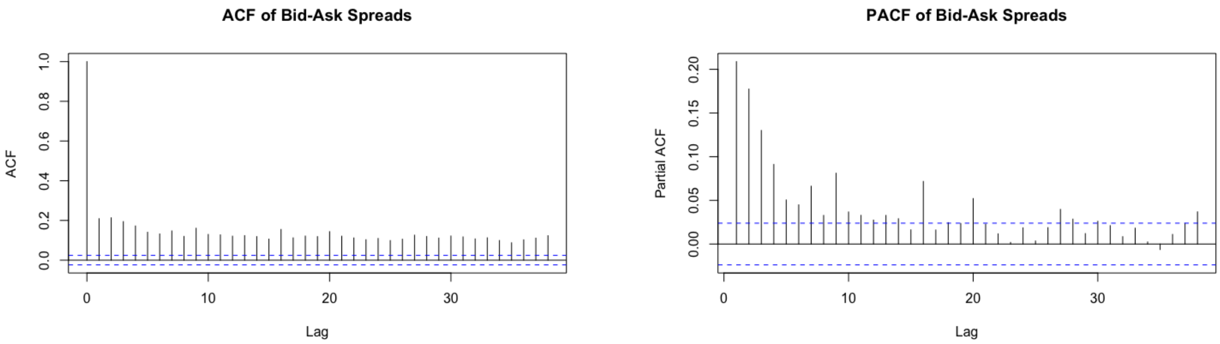


Figure 2.2: Bid-Ask Spreads Over Time



ACF of Spreads

PACF of Spreads

Figure 2.3: ACF and PACF of the Spread Series

The same is done for the relative spread data:

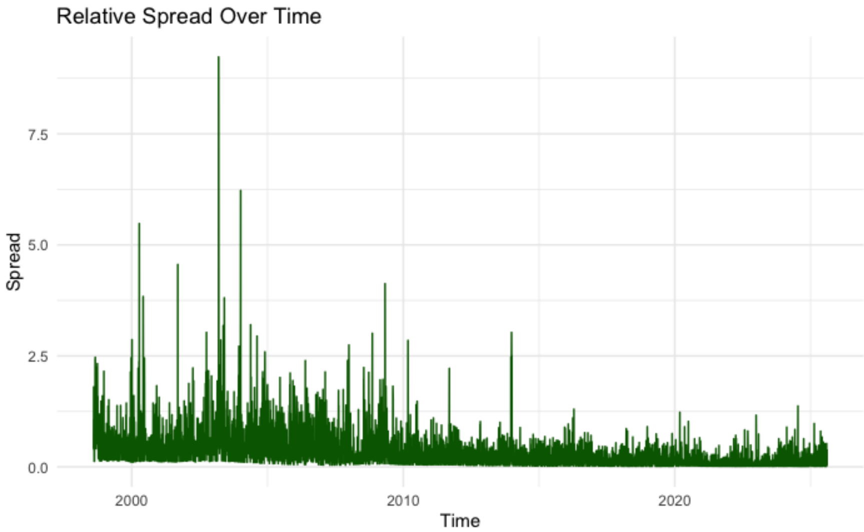
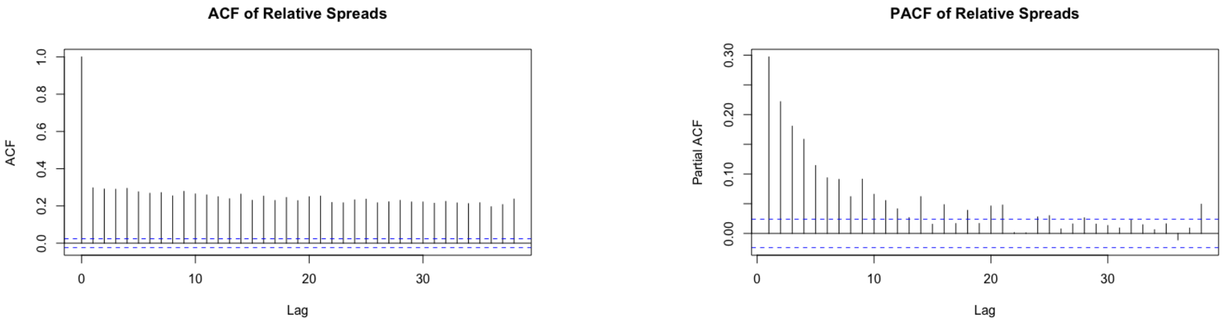


Figure 2.4: Relative Spreads Over Time



ACF of Relative Spreads

PACF of Relative Spreads

Figure 2.5: ACF and PACF of the Relative Spread Series

Both of the series seem to be extremely volatile, with extreme spikes present throughout the sample. The relative spread plot shows that generally the spreads are tight; however, large jumps take place fairly often. In other words, the stock exhibits liquid behaviour most of the time; however, there are often large liquidity shocks. With respect to the ACF plots, both exhibit strong long-term dependence behaviour with very slow decay – this is consistent with non-stationarity.

To test the stationarity of the data, we make use of the augmented Dickey-Fuller test, where the null hypothesis is that the series has a unit root (Tsay, 2005). If a series does not have a unit root, it can still be non-stationary, so we also use Kwiatkowski–Phillips–Schmidt–Shin (KPSS). Here the null hypothesis is that the series is level stationary (Kwiatkowski *et al.*, 1992). The results are as follows:

Table 2.2: ADF and KPSS Stationarity Tests for Spread Series

Series	Test	Statistic	Lag Order	p-value
Spread	ADF	-12.709	18	< 0.01
Spread	KPSS	11.753	11	< 0.01
Relative Spread	ADF	-13.505	18	< 0.01
Relative Spread	KPSS	33.236	11	< 0.01

In both cases the ADF test rejects that there is a unit root. However, the KPSS test null is also rejected, meaning the KPSS test says that both series are not stationary. Combining this evidence with the persistence of autocorrelation in the ACF plot, we conclude that both series' are non-stationary. This brings us to the next section.

2.3.2 Log Returns of Spreads

The log returns are plotted over time:

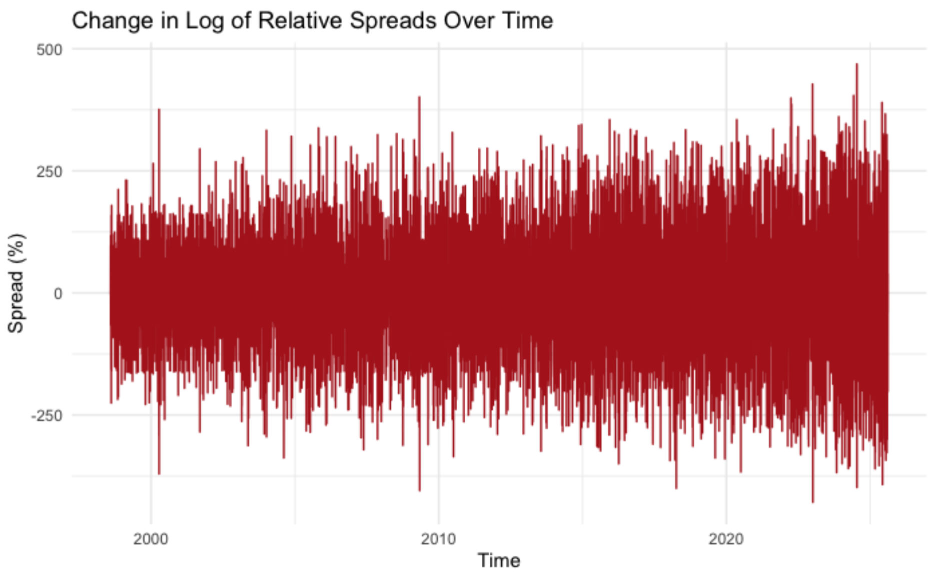
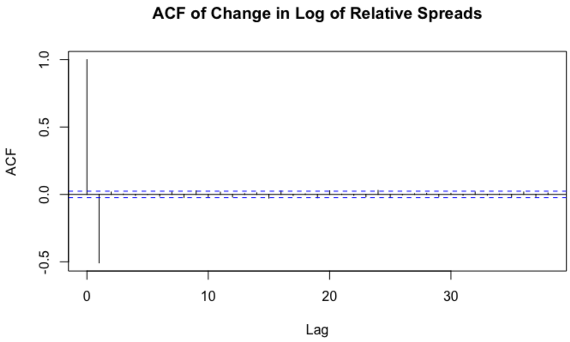
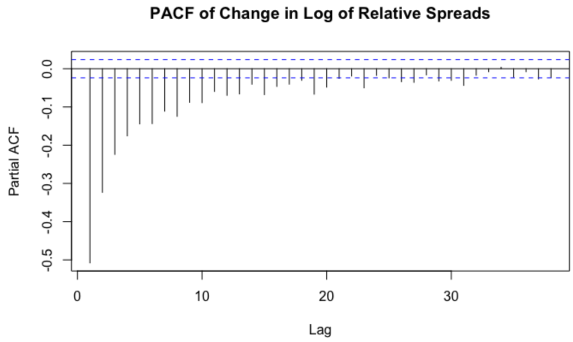


Figure 2.6: Log Returns of Spreads Over Time

Again, we see large spikes. However, the series does appear to look more stationary than the previous two. The ACF plots are given:



ACF of Log-Return Spreads



PACF of Log-Return Spreads

Figure 2.7: ACF and PACF of Log-Return Spread Series

The ACF has a strong negative correlation at lag 1 and afterwards drops off. The partial ACF trails off. The log return series therefore seems to exhibit the behaviour of a moving average process of order 1 (Tsay, 2005). The stationarity test results are shown:

Table 2.3: ADF and KPSS Tests for Log-Return Spread Series

Series	**Test**	**Statistic**	**Lag Order**	**p-value**
Log-Return Spread	ADF	-31.287	18	< 0.01
Log-Return Spread	KPSS	0.0037026	11	> 0.10

According to both tests, we can conclude the log-return spreads are stationary, and therefore can be used in the extreme value analysis.

2.3.3 Why an Extreme Value Analysis?

The log-return spread values S_t^{log} capture proportional increases or decreases in the underlying relative spreads and therefore reflect the tightening or widening of market liquidity. Large positive values correspond to sudden spread blowouts—precisely the liquidity shocks that traders, risk managers, and regulators are most concerned about, as they lead to abrupt increases in execution costs and make it harder to enter or exit positions. Modelling the extremes of S_t^{log} using extreme value theory therefore provides a direct way to quantify the tail risk of these liquidity shocks (i.e., extreme liquidity events), yielding informative estimates of extreme quantiles and return levels.

These metrics help assess the likelihood and severity of sudden liquidity deteriorations, contributing to a more robust understanding of liquidity risk.

SECTION 3

METHODOLOGY

The S_t^{log} values represent a percentage. As a result, they have been multiplied by 100 to make them more readable.

3.1 METHOD: BLOCK MAXIMA

3.1.1 Thoeretical Background: Generalised Extreme Value Distribution

The following theory is based on Beirlant *et al.* (2006).

3.1.1.1 Independent Case

For a set of independent and identically distributed (i.i.d) data $X_1, ..., X_n$, let $X_{n,n} = max\{X_1, ..., X_n\}$. From the Fisher-Tippet theorem, we know the limiting distribution of the normalized maximum converges to a Generalized Extreme Value distribution, which can be written as:

$$G(x; \sigma, \gamma, \mu) = \begin{cases} \exp\left(-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right), & 1+\gamma\frac{x-\mu}{\sigma} > 0, \gamma \neq 0, \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), & x \in R, \gamma = 0, \end{cases}$$

with $\sigma > 0$ and $\mu \in R$.

3.1.1.2 Stationary Data

Say now our data $X_1, ..., X_n$ is no longer i.i.d, but stationary. Then from Leadbetter 1974, we know that the normalized maximum of this series also converges to a GEV, if and only if the $D(u_n)$ condition hold. Here the $D(u_n)$ condition essentially limits the long range dependence, ensuring the block maxima are asymptotically i.i.d.

3.1.1.3 The Extremal Index

From Leadbetter 1983, if the distribution of the normalized maxima of stationary data converges, we get:

$$G(x) = \tilde{G}^{\theta}(x),$$

with $\theta \in [0, 1]$. Here G is the limiting distribution (GEV) for stationary data, $\tilde{G}$ is the limiting distribution (GEV) for i.i.d data, and θ is defined as the Extremal Index. It will be seen later that this parameter is extremely important in estimating quantiles.

3.1.2 Implementation: Method of Block Maxima

To start, we split our data, $S_1^{log}, ..., S_N^{log}$, into blocks of size n. We then take the maximum of each block, resulting in the block maxima data $Y_1, \ldots, Y_m$. There is a bias-variance trade-off with block size. Too small means the data is not close to GEV distributed. Too large means we don't have a lot of data to use in the estimation. A block size of 22 is chosen, as this corresponds to monthly blocks, which seems natural given the financial setting. Additionally, it is relatively large, but also results in a lot of data to use for estimation. Block sizes of 5 and 50 are also checked, with the results in the appendix.

3.1.2.1 Parameter Estimation

To estimate the GEV parameters (for $\gamma \neq 0$), we use maximum likelihood estimation (mle):

$$\begin{aligned}\log L(\sigma, \gamma, \mu) = & - m \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{m} \log \left(1 + \gamma \frac{Y_i - \mu}{\sigma}\right) \\ & - \sum_{i=1}^{m} \left(1 + \gamma \frac{Y_i - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\end{aligned}$$

The resulting parameters are such that they maximise the above equation. To estimate the Extremal Index, we use the block estimator and sliding block estimator as given in chapter 10 of Beirlant *et al.* (2006).

3.1.2.2 Quantile Estimation

We estimate the quantiles of the GEV distribution using the following equation:

$$q_{Y,p} = \begin{cases} \mu + \frac{\sigma}{\gamma}\left[(-\log(1-p))^{-\gamma} - 1\right], & \gamma \neq 0, \\ \mu - \sigma \log(-\log(1-p)), & \gamma = 0, \end{cases}$$

If we have $\gamma < 0$, the upper endpoint can be estimated:

$$q_{Y,0} = \mu - \frac{\sigma}{\gamma}$$

We know that $F_{X_{n,n}} = F^n \approx G$, hence, the quantiles of the original X data are calculated with:

$$q^*_{X,p} = \begin{cases} \mu + \sigma \frac{\{-n\theta \log(1-p)\}^{-\gamma} - 1}{\gamma}, & \text{if } \gamma \neq 0, \\ \mu - \sigma \log\{-n\theta \log(1-p)\}, & \text{if } \gamma = 0. \end{cases}$$

where n is the block length. Here the extremal index can be seen to have an effect on calculating the quantiles of the underlying distribution, and, can therefore be interpreted as a decreasing factor of the effective sample size (Coles *et al.*, 2001). To calculate the return level, we take $p = 1/m$.

Normal and profile-likelihood confidence intervals (CI) are calculated for parameter estimates, as described in chapter 5 of Beirlant *et al.* (2006). Additionally, the delta method is used to get the normal confidence intervals of the quantile estimates. The profile-likelihood CI is also used for the quantile estimates.

3.2 METHOD: PEAKS-OVER-THRESHOLD

3.2.1 Theoretical Background: Generalized Pareto Distribution

We know for the Pickands-Balkema-de Haan theorem that the limiting distribution of the conditional exceedances $Y = X - t | X > t$ is the Generalized Pareto Distribution:

$$\bar{F}_t(y) \sim \left(1 + \frac{\gamma y}{\sigma}\right)^{-1/\gamma}.$$

We write the GPD as:

$$H(y;\gamma,\sigma) = \begin{cases} 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-1/\gamma}, & y \in (0,\infty) & \text{if } \gamma > 0, \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & y \in (0,\infty) & \text{if } \gamma = 0, \\ 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-1/\gamma}, & y \in \left(0, -\frac{\sigma}{\gamma}\right) & \text{if } \gamma < 0, \end{cases}$$

3.2.2 Implementation: Peaks-Over-Threshold

We get our sample by taking all the values above the threshold and declustering. A cluster is defined by an imperical rule - we define a new cluster when an exceedance occurs after at least 5 consecutive non-exceedances. We then take the maximum of each cluster and subtracting the threshold. This results in $Y_1, \ldots, Y_{n_c}$, where n_c is the number of clusters. In the i.i.d case, we would have N_t instead of n_c, where N_t is the number of exceedances. Our choice of threshold is based on the mean-excess plot.

3.2.2.1 Parameter Estimation

Again, we use maximum likelihood estimation by finding the parameters that maximise:

$$\log L(\sigma,\gamma) = -n_c \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{n_c} \log\left(1 + \frac{\gamma Y_i}{\sigma}\right)$$

for $1 + \frac{\gamma Y_i}{\sigma} > 0, i = 1, \ldots, n_c$.

3.2.2.2 Quantile Estimation

We estimate the Quantiles of the GPD distribution using:

$$q^*_{Y,p} = \begin{cases} \frac{\sigma}{\gamma}\left(p^{-\gamma} - 1\right) & \gamma \neq 0, \\ -\sigma \log p & \gamma = 0, \end{cases}$$

Adapting the return level formula given in Coles *et al.* (2001), we get the quantile estimate for the underlying distribution

$$q^*_{X,p} = t + \frac{\hat{\sigma}}{\hat{\gamma}}\left(\left(\frac{p}{\hat{\zeta}_t\hat{\theta}}\right)^{-\hat{\gamma}} - 1\right)$$

where $\hat{\zeta}_t = \frac{N_t}{n}$ and $\hat{\theta} = \frac{n_c}{N_t}$. Again, we see the extremal index reduces the effective sample size. To estimate the upper bound (for $\gamma < 0$) we use:

$$\hat{x}_+ = t - \frac{\hat{\sigma}}{\hat{\gamma}}$$

Normal confidence intervals are again calculated as in Beirlant *et al.* (2006).

SECTION 4

RESULTS

4.1 BLOCK MAXIMA

4.1.1 Parameter Estimates

The parameter estimates for the GEV, along with their standard errors and normal confidence intervals are given:

Table 4.1: GEV Parameter Estimates (MLE), Standard Errors, and 95% Confidence Intervals

Parameter	**Estimate**	**SE**	**Lower 95% CI**	**Upper 95% CI**
γ	-0.16332	0.03734	-0.23652	-0.09013
μ (%)	210.53691	3.89817	202.89664	218.17717
σ (%)	61.38281	2.77634	55.94129	66.82433

We see that the γ value in negative, with the confidence interval also being negative. Therefore, we are in the extremal Weibull domain. This is interesting as it means there is an upper bound for which spreads can increase. Also note that since the data was scales, the μ and σ parameters and their standard errors are also expressed as percentages. To determine the quality of the fit, we look at a QQ plot:

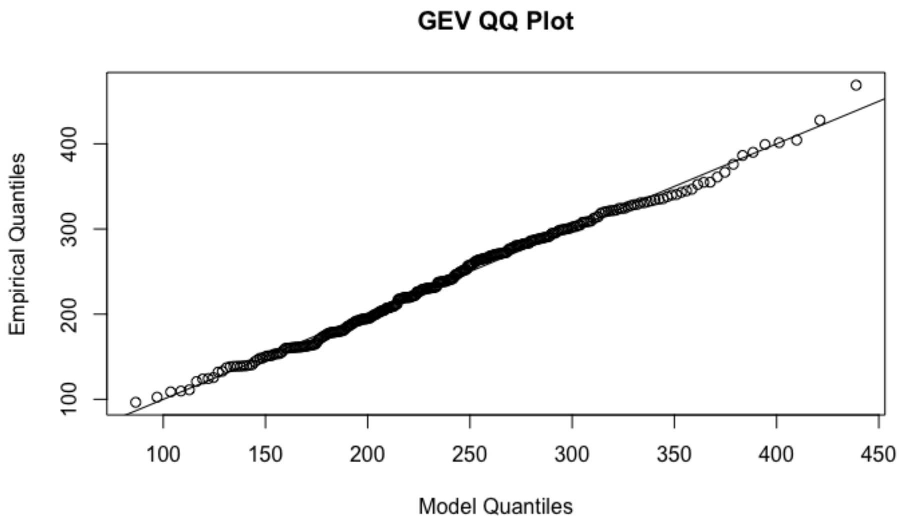


Figure 4.1: GEV QQ Plot

The model clearly fits the majority of the data very well. The two largest data points seem to be slightly underestimated, however, the overall plot indicates that the GEV models the extremes well. The results for the weekly and 50-day block maxima are also given in the appendix. The model fit seems to be the best overall for the monthly maxima, with the other two having more deviations away from the line in the QQ plot. We then look at the profile-likelihood confidence interval for γ:

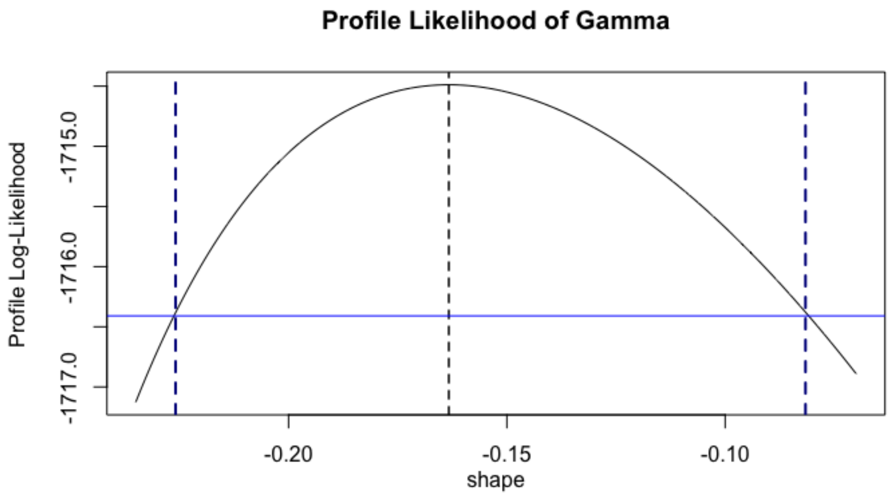


Figure 4.2: γ Profile Likelihood

Again, the entire confidence interval is below 0, therefore we are definitely in the extremal Weibull domain.

To estimate the Extremal Index, a threshold u is needed. The following table gives the $theta$ values for different threshold, chosen as quantiles in the data:

Quantile	**Threshold** u	$\hat{\theta}$ **(blocks)**	$\hat{\theta}$ **(sliding)**
0.980	270.1981	0.7481481	0.7319128
0.990	298.6159	0.8676471	0.8334823
0.995	324.2127	0.9411765	0.8918194
0.999	378.7496	1.0000000	0.8858803

Table 4.2: Extremal index estimates for various upper-tail thresholds using disjoint and sliding block methods.

As expected, the block estimate rises with the threshold. The most stable region seems to be $p = 0.99$ and $p = 0.995$ when looking at the two estimates together. For the sliding estimate,

$p = 0.999$ yields an estimate still in the same range as $p = 0.99$ and $p = 0.995$. Hence, it seems reasonable to use $\hat{\theta} = 0.88$.

4.1.2 Quantiles and Return Levels

4.1.2.1 Quantiles of The GEV

The monthly return level plot of the GEV is given:

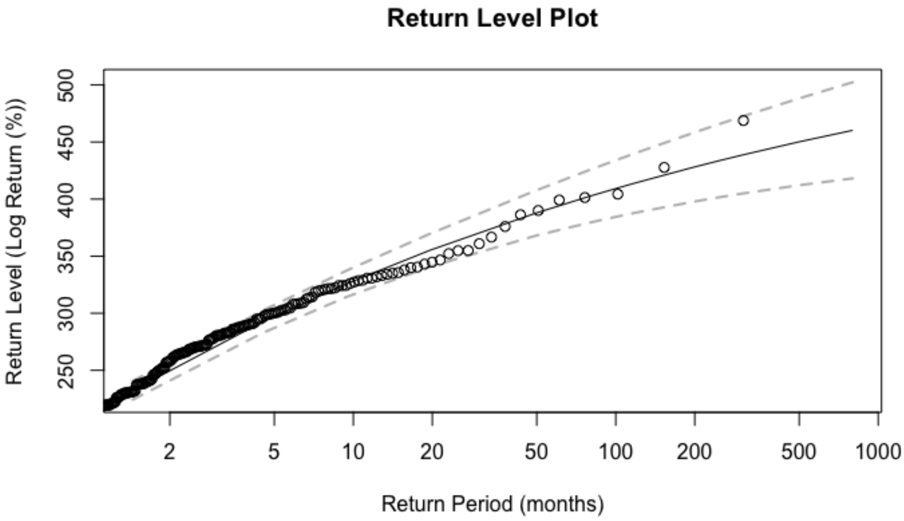


Figure 4.3: GEV Return Level

Several high quantile estimates of the GEV are given with their confidence intervals:

Tail Prob.	Estimate (%)	Delta-Method CI	Profile Likelihood CI
0.01	409.0731	(384.3318, 433.8144)	(390.0215, 441.8595)
0.001	464.7361	(420.6852, 508.7869)	(433.7068, 526.7414)
0.0001	502.8681	(438.8997, 566.8365)	(459.7253, 598.1753)

Table 4.3: GEV Return Level Estimates with Delta-Method and Profile Likelihood Confidence Intervals

The upper endpoint is calculated as 586.3716%.

4.1.2.2 Quantiles of the Underlying Distribution

We then estimate the high quantiles for the underlying distribution: These tail probabilities trans-

Tail Prob. p_{tail}	**Estimate (%)**	**95% Delta-Method CI**
0.01	298.7062	(288.6671, 308.7453)
0.001	389.0190	(368.9764, 409.0615)
0.0001	450.8879	(412.5958, 489.1800)

Table 4.4: Dependence-adjusted stationary return levels and 95% delta-method confidence intervals.

late into a one in 100 day event, a one in 1000 day event and a one in 10000 day event respectively. In years, that is a 0.27, 2.7 and 27 year return level. Financially, these results are quite scary. The one in a 100 day estimate is 298.7062%, or 2.98. This results in a $e^{2.98} \approx 20$ times increase in the current spread level. The theoretical upper bound under the GEV results in a $e^{5.86} \approx 352.02$ increase in the spread. The multiplier transformations are show for the probability levels:

Tail Prob. p_{tail}	**Multiplier** $\exp(\widehat{q}/100)$	**95% CI for Multiplier**
0.01	19.82	(17.98, 21.90)
0.001	48.87	(40.03, 59.70)
0.0001	90.82	(61.90, 133.20)

Table 4.5: Transformed dependence-adjusted return levels expressed as multiplicative factors.

The results imply that spreads can increase drastically and fairly often, which is troubling from a liquidity risk perspective.

4.2 PEAKS-OVER-THRESHOLDS

4.2.1 Parameter Estimates

Before estimating the GPD parameters, a threshold must be chosen. We do this by looking at the mean excess plot:

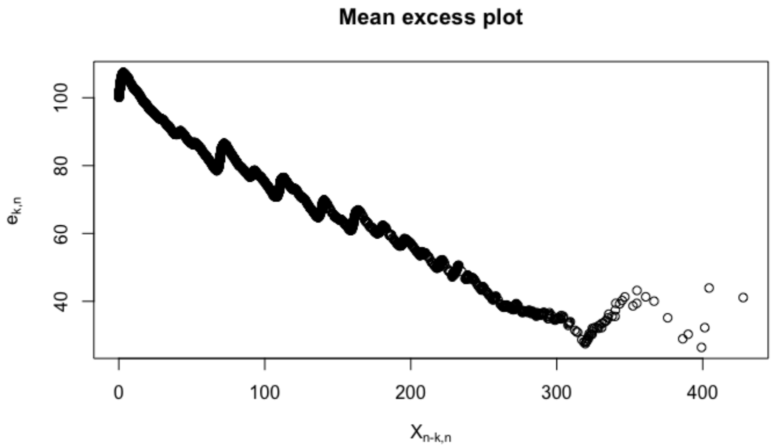


Figure 4.4: Mean Excess Plot

The plot exhibits linearity at a high value of X from about 230 onwards. This corresponds to k=250. Looking at γ estimates for different thresholds, this seems to be a good choice:

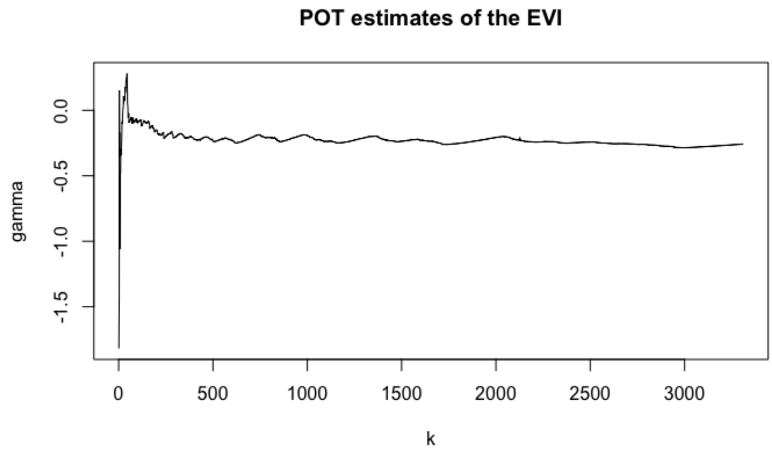


Figure 4.5: Gamma estimates as a function of thresholds

Usinge a threshold of 230, we decluster the data. This resulted in 209 clusters, and and extremal index of $\hat{\theta} = 0.7894754$. The parameter estimates are as follows:

Parameter	**Estimate**	**SE**	**Lower 95% CI**	**Upper 95% CI**
γ	-0.1958732	0.04712672	-0.2882399	-0.1035065
σ	60.6623559	5.01145032	50.8400938	70.4846181

Table 4.6: GPD parameter estimates and 95% confidence intervals (declustered exceedances, threshold $t = 230$).

We again check the fit using a QQ plot:

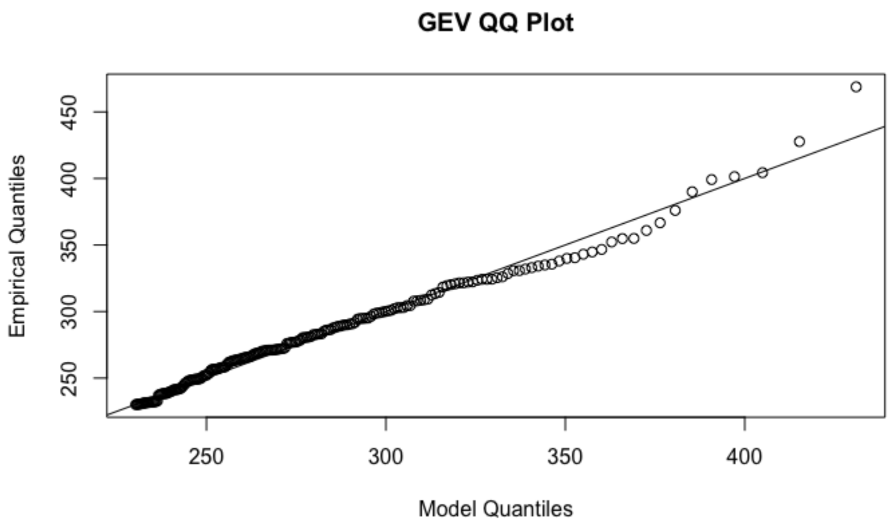


Figure 4.6: GPD QQ Plot

The data generally seems to be modelled well until about halfway through the quantiles, where the observations deviate a bit. Overall the fit is good, however, it does not seem to be as good as the GEV fit.

4.2.2 Quantiles and Return Levels

The high quantile estimates are given:

Tail Probability p	**GPD Quantile**	**Underlying Quantile (Unconditional)**
$p = 0.01$	414.0419	292.3936
$p = 0.001$	459.6589	382.1712
$p = 0.0001$	488.7161	439.3579

Table 4.7: Comparison of GPD exceedance quantiles (conditional on $X > t = 230$) and underlying-series unconditional return levels using ζ_t and extremal index θ.

The estimate for the endpoint is: 539.7022. These results reiterate the findings of the GEV case, where large liquidity shocks can occur fairly often.

4.3 DISCUSSION OF RESULTS

As mentioned earlier, bid-ask spreads are a commonly used proxy for market liquidity. Modelling the log-returns of the spreads allows us to quantify proportional changes in liquidity, since a log-return of size x corresponds to a multiplicative change of e^x in the spread level. The extreme-value results show that even at moderate tail probabilities, the spread can increase by a factor of approximately 20×, and such events have a return period of around 100 days, meaning that they may occur more than once per year. From a financial perspective, a twenty-fold widening of the spread represents a severe liquidity shock: execution costs rise sharply and trading can become incredibly expensive or even impossible. These findings highlight the magnitude of liquidity risk faced by traders and risk managers, and underscore the importance of monitoring and managing exposure to sudden liquidity deterioration.

SECTION 5

CONCLUSION

This study applied Extreme Value Theory to the log-returns of the relative bid-ask spread of FirstRand Ltd. in order to quantify the likelihood and magnitude of extreme liquidity events. After demonstrating that spread levels and relative spreads exhibit nonstationarity and strong persistence, the log-return transformation was shown to produce a stationary series suitable for EVT modelling. The resulting return-level and quantile estimates provide insight into the tail behaviour of liquidity shocks.

The empirical results reveal the large liquidity shocks take place fairly often. From a financial perspective, this analysis highlights the importance of modelling liquidity risk explicitly. Spreads, as a proxy for market liquidity, exhibit jump-like behaviour. Understanding the probability and scale of such jumps is crucial for risk managers, traders and regulators who need to prepare for extreme liquidity shocks. Overall, this project shows the usefulness of EVT from a liquidity risk management perspective and the importance of accounting for liquidity risk.

Further research in this area would be the inclusion of covariate, such as other financial or economic data, or non-stationary EVT modelling of the spreads where the parameters are a function of time.

REFERENCES

Affleck-Graves, J., Callahan, C.M. and Ramanan, R. (2000). Detecting abnormal bid-ask spread: a comparison of event study methods. *Review of Quantitative Finance and Accounting*, vol. 14, no. 1, pp. 45–65.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Coles, S., Bawa, J., Trenner, L. and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, vol. 208. Springer.

Harris, L. (2002). *Trading and exchanges: Market microstructure for practitioners*. Oxford university press.

Kwiatkowski, D., Phillips, P.C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, vol. 54, no. 1-3, pp. 159–178.

Maginn, J.L., Tuttle, D.L., McLeavey, D.W. and Pinto, J.E. (2007). *Managing investment portfolios: a dynamic process*, vol. 3. John Wiley & Sons.

Muela, S.B., Martín, C.L. and Sanz, R.A. (2017). An application of extreme value theory in estimating liquidity risk. *European Research on Management and Business Economics*, vol. 23, no. 3, pp. 157–164.

Tsay, R.S. (2005). *Analysis of financial time series*. John wiley & sons.

APPENDIX A

APPENDIX

Parameter	Estimate	95% CI (Lower)	95% CI (Upper)
γ	−0.1054902	−0.1451826	−0.0657977
μ	125.2231291	121.1712644	129.2749939
σ	67.3406267	64.4383868	70.2428667

Parameter	Standard Error
γ	0.02025162
μ	2.06731592
σ	1.48076187

Table A.1: GEV MLE parameter estimates, standard errors, and 95% confidence intervals based on **weekly maxima**.

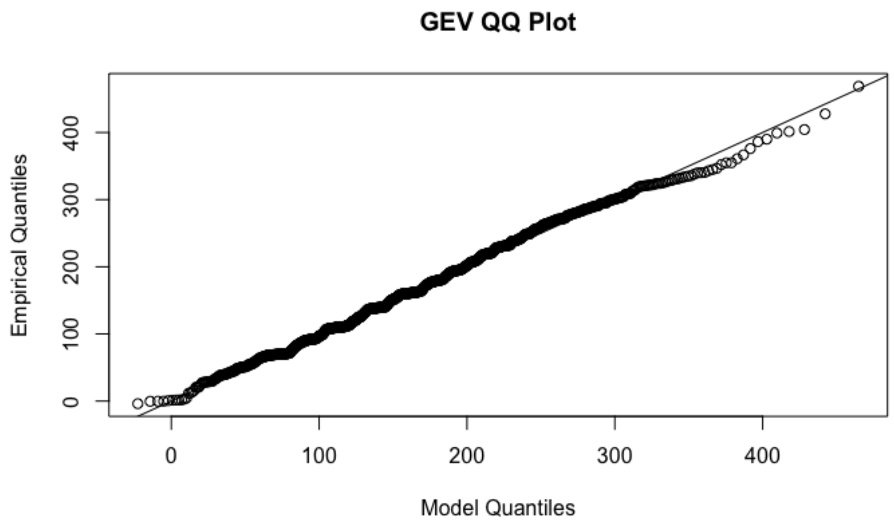


Figure A.1: Weekly Maxima

Parameter	Estimate	95% CI (Lower)	95% CI (Upper)
γ	-0.2102269	-0.2929282	-0.1275256
μ	252.2926266	241.5573291	263.0279241
σ	58.4855114	51.0660742	65.9049486

Parameter	Standard Error
γ	0.04219531
μ	5.47729326
σ	3.78549669

Table A.2: GEV MLE parameter estimates, 95% confidence intervals, and standard errors based on **50-day block maxima**.

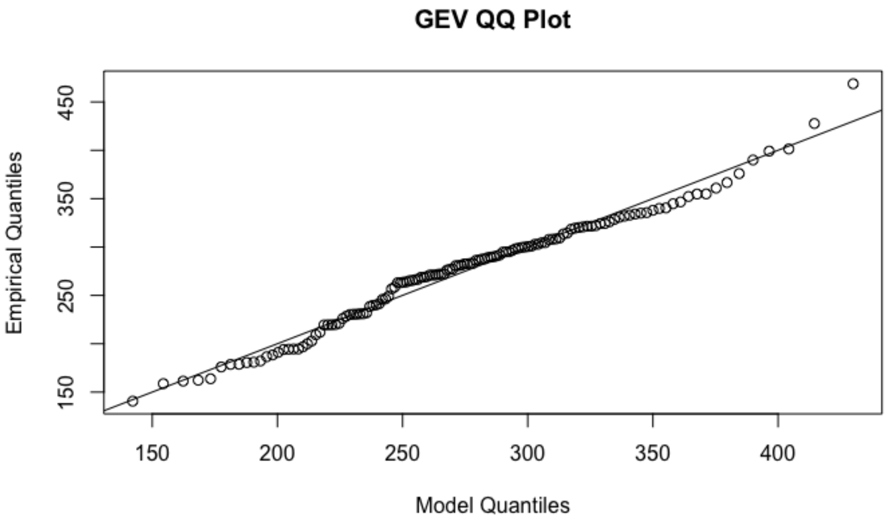


Figure A.2: 50 day maxima