AN EXTREME VALUE ASSESSMENT OF BITCOIN

by

Christopher Newmark

Project

Submitted to the College of Science and Mathematics

in partial fulfillment of the requirements for the degree of

Master of Arts in Mathematics

California State University, Fresno

December 2018

Advisor: Dr. Steve Chung

Approval of Committee

We, the undersigned, approve the project of the following student as meeting the standards of the Department of Mathematics.

Christopher Newmark

Dr. Steve Chung	Department
Dr. Comlan de Souza	Department
Dr. Ke Wu	Department

Acknowledgements

I would like to thank God, without Whose grace I would not have come to this program and made it through. Thank you to my family who have supported me emotionally and financially; with much love and even more food. Thank you to my professors at Fresno State, who have given me the great gift of a love for mathematics which will last a lifetime. Thank you to my advisers for your excellent feedback on my project, and especially Dr. Chung whose time and insight has been very valuable to me - not just for this project, but also during my time at Fresno State. Thank you to my friends for all of the encouragement, support, and help: especially Clayton, Nathan and John. Thank you to Dr. Balderama and Dr. Bishop for your time discussing and helping to improve my project. A special thank you to Dr. Markin for your inspiration to never give up and your effort to help keep me in the program. And thank you to my girlfriend, then fiancee, and now wife, Sydney, for your unwavering support and love, and for always believing in me.

Abstract

Bitcoin is the major currency in the emerging cryptocurrency market. Many institutions and private investors are now paying attention to this global phenomenon. Depending on which website one looks at, bitcoin's price began at $13.47 on January $6th, 2013$ and reached its peak price of $\$18,877.72$ on December $17th, 2017$. But the central question is: just how risky is this currency? In this paper, we use a recent procedure based on the Hill plot in order to better estimate the threshold of our bitcoin data set, which ranges from August $1st, 2013$ to August $1st, 2016$. We then compare these calculated thresholds to thresholds chosen by more subjective methods. And finally, we apply the calculated thresholds to the *generalized Pareto distribution* from extreme value theory. Our results show that the choice of threshold has a significant impact on the *shape* of the corresponding *Pareto distribution*, and that the predictions made by the corresponding models also differ significantly.

Contents

Chapter 1

Introduction

1.1 Currency, Cryptocurrency, and Bitcoin

In the last decade technology has become more prevalent. From credit cards to Apple Pay, we rely daily on technology for our digital transactions. For most people, their money is held at their bank, and they can withdraw paper money from their account at any time. The banks are controlled by a centralized government entity which results in inefficiency and fees. There is a general belief that a decentralized currency would remedy much of this inefficiency and create more economic opportunity and autonomy. Until the publication of the landmark white paper in 2008 [8] by an unknown computer scientist or group under the name of Satoshi Nakamoto, it was believed by many that this decentralized digital currency was impossible to create. Nakamoto's paper showed that it was possible, but how would it work? Figure 1.1 shows a screen shot from *blockgeeks.com* with a flowchart of the blockchain validation process for cryptocurrency.

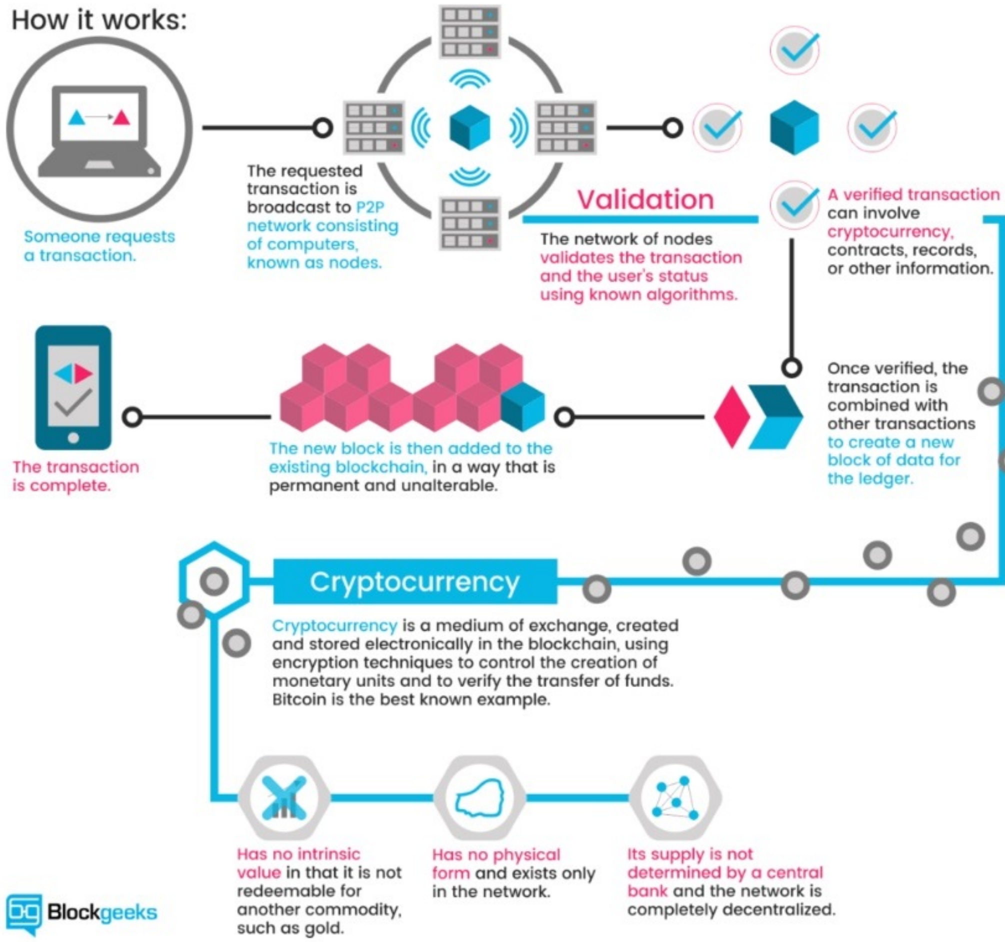

Figure 1.1: Blockchain and Cryptocurrency ideas.

Cryptocurrency's popularity has been growing. In 2016, there were 720 different cryptocurrencies [10]. Now in 2018, at the time of this report, there are $2,099$ digital cryptocurrencies. The largest and most popular cryptocurrency by far is Bitcoin. And it is this currency which we focus on in this work.

These cryptocurrencies are named as such because they are digitally encrypted and are thus extremely difficult (and some believe impossible) to trace. An individual can go online, and after verification which is usually a few weeks, transact their digital currency for cryptocurrency in the same manner that one would exchange say, dollars, for euros. There are many places one can spend cryptocurrency, and they are only increasing in amount and access. However, illegitimate places to spend cryptocurrency, primarily on the dark web, have been an issue of contention. This has garnered world-wide attention and calls for government regulation - the main entity from which it sought freedom in the first place. As can be seen, its use is highly controversial. However, this has not stopped many people from speculating on its worth. It is thought by many that cryptocurrency is the future for money, and if not in its current privatized form, then implemented (legitimately) by government entities - see for example Venezuela's Petro.

A number of banks are becoming interested and involved in the cryptocurrency market as well [3]. This is the beginning of the motivation for this paper. Many financial institutions rely on characterizing the risk of their investments, which now involves cryptocurrency. Previous statistical analysis from Osterreider and Lorenz (2017) sought to classify the probability distribution of bitcoin and compare its different statistical properties to those of well known currencies. Their papers showed that, like many other currencies, the returns of bitcoin are not normally distributed and instead have fat-tailed distributions, see [9] and [10]. Moreover, the volatility of bitcoin was shown to be $6-7$ times higher than the G10 currencies. Additionally, extreme events for bitcoin led to losses which were 8 times higher than G10 currencies. Using extreme value analysis, they also showed that the risk associated with bitcoin is much higher compared to traditional currencies with an expected loss of about 10% every 20 days [10]. In this work, we propose a new threshold for this extreme value analysis using the method from Zhou et. al. (2007). We also calculate three more thresholds, and then compare the effects that all of these different thresholds have on our understanding of bitcoin's risk.

1.2 Data

Our data set is the price of one bitcoin in USD taken from the dates August 2013 to August 2016. The data set was downloaded for free from *coindesk.com*. Figure 1.2 shows the price of bitcoin during this time interval. The summary statistics for bitcoin in this time interval are shown in Table 1.1.

Min.	1st Quartile	Median	Mean	3rd Quartile	Max.	St. Dev.
0.06	6.63	124.30	221.30	402.60	1147.00	241.69

Table 1.1: Summary of bitcoin daily price returns.

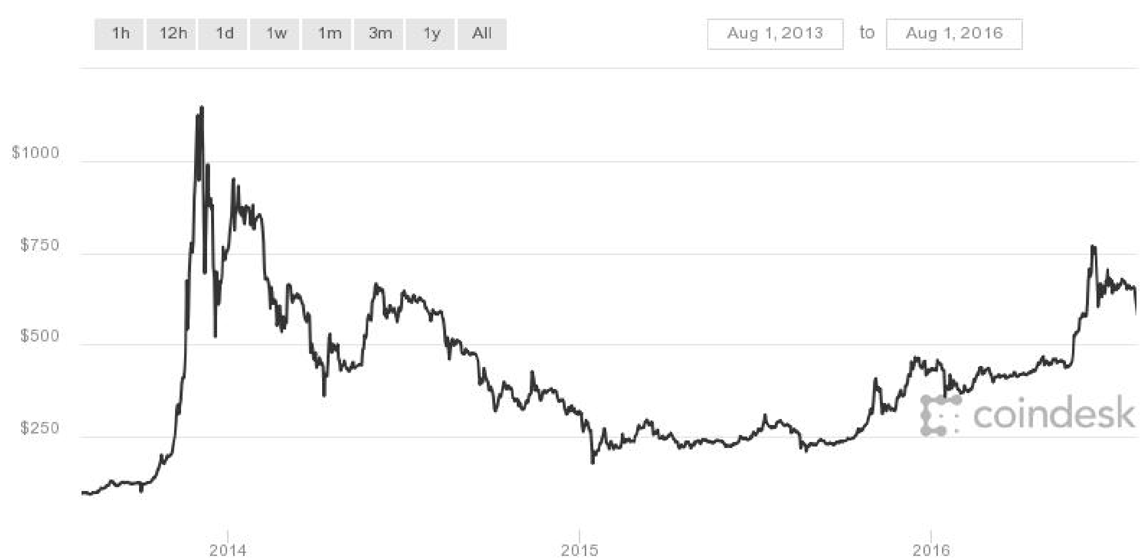

Figure 1.2: Bitcoin price chart from August 2013 to August 2016.

It is common in the financial world to model currency not based on the price, but based on the daily, monthly, or yearly returns. We can create this data set by taking the difference of time intervals, i.e., the return at time period i is $X_i - X_{i-1}$. Thus, if one bought a currency yesterday for \$100 and sold it today for \$110, then one's return today would be $\$110 - \$100 = \$10$. Table 1.2 shows the summary statistics for the bitcoin return series.

Min.	1st Quartile	Median	Mean	3rd Quartile	Max.	St. Dev.
-208.00	-0.79	0.01	0.26	1.42	198.00	17.50

Table 1.2: Summary statistics for the regular return series of the closing prices of bitcoin.

It is important to note that the summary statistics shown are for the regular return series. However, in practice, either the normalized returns (called the simple returns) or natural logarithmic returns are used in the analysis.

Hudson and Gregoriou (2015) show the effect that using logarithmic returns has on the results of financial data analysis. Although the logarithmic returns have attractive mathematical properties, their conclusion is that using the logarithmic return series does not allow for an accurate description of the statistical properties of the data. It is for this reason that we do not use the logarithmic return series for bitcoin in this work. Thus we have chosen here to use the simple return series which is created in the following way:

$$SimpleReturns_i = \frac{X_i - X_{i-1}}{X_{i-1}}$$

The simple return series is the percentage gain or loss - in this case, on a daily basis. It is desirable to consider percentages since they are dimensionless and are thus comparable to other percentage return series. For example, if we were comparing bitcoin to another currency and, on the same day, bitcoin dropped by \$100 and the other currency dropped by \$10, one might think that bitcoin changed quite drastically in comparison. However, if we then looked at the simple return and found out that that the previous day, bitcoin was at $\$6,000$ and the other currency was at \$20, then the percentage change from $\$6,000$ to $\$5,900$ would be about 17% whereas the percentage change from \$20 to \$10 would be 50%. This example shows how the relative price differences are accounted for by using the simple return series.

1.3 Testing for Normality

When beginning any modeling or regression problem, it is standard to check if the data is approximately normally distributed since the assumptions about the underlying distribution of the data will affect the analysis. Thus, we want to decide if the simple returns are normally distributed. We begin by looking at the $quantile - quantile$ or (QQ) plot of our return series. This involves creating a graph of (x, y) coordinates with the quantiles of the empirical data paired with the theoretical standard normal distribution quantiles. If the normal distribution is a good fit, then the graph should be approximately linear. Figure 1.3 shows the QQ plot of the return series. Note that the S-shape indicates that the bitcoin return series has fat-tails and is poorly fit by the normal distribution.

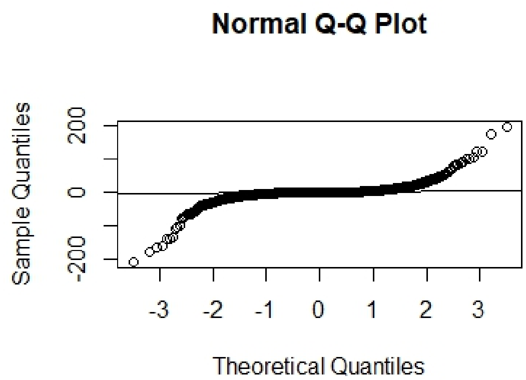

Figure 1.3: Plot of the data quantiles paired with the normal distribution quantiles.

Figure 1.4 shows the fitted normal curve in red over the normalized return series. This normal distribution has a mean of 0.006011241 and a standard deviation of 0.060935593.

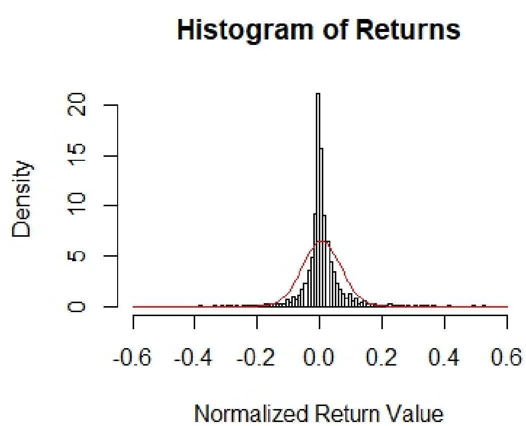

Figure 1.4: Histogram of simple returns with red normal distribution.

Figure 1.5 shows this same graph closer to the tail of the negative returns. Note that the normal curve does a poor job of estimating this return series in both the tail and the mode.

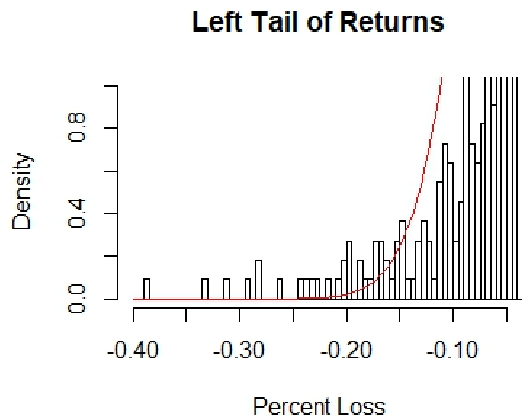

Figure 1.5: Close up of the negative simple returns with red normal distribution.

When considering the normality of a univariate data set, there are two main statistics which must be computed, the *skewness* and *kurtosis*. The former statistic tells us how symmetric our distribution is - is the mean value equivalent to the median value? The latter statistic gives a quantification for how thin or fat the spread of the distribution is - are the tails fat or skinny? We define these two statistics respectively as $S = E\big[\big(\frac{X-\mu}{\sigma}\big)^3\big]$ and $K = E\big[\big(\frac{X-\mu}{\sigma}\big)^4\big]$. For a normal distribution with a mean of zero and a standard deviation of one, the *skewness* is zero and the *kurtosis* is three. In addition to the standard definition, there is an alternate definition which looks at how far away the *kurtosis* is from the assumed value of three. Thus the *excess kurtosis* is defined as $EK = E\big[\big(\frac{X-\mu}{\sigma}\big)^4\big] - 3$. To decide if a difference in *skewness* or *kurtosis* of our data is significant compared to the normal distribution, we can use the *Jarque-Bera test* which is defined as $JB = \frac{N}{6} \cdot (S^2 + \frac{1}{4}(K-3)^2)$, where N is the number of observations, S is the *skewness*, and K is the *kurtosis* [7].

The null hypothesis of this test assumes that there is no difference between our data's *skewness* and *kurtosis* and the normal distribution's *skewness* and *kurtosis*. Table 1.3 shows the *skewness* and *kurtosis* of the bitcoin simple return series, as well as the results of the *Jarque-Bera test*.

Skewness	Kurtosis	Jarque-Bera
0.9967583	12.42849	14503

Table 1.3: Results of the Jarque-Bera test.

This very large value for the *Jarque-Bera test* has a p value of about zero which indicates we can reject the null hypothesis of normality. To state this another way, we could sample about 1×10^{16} values from the normal distribution and never create the histogram which represents the bitcoin simple returns. Osterrieder (2016) analyze the basic statistical elements of bitcoin in greater detail and also fit many different non-normal curves to the *log* return series. They also conclude that many cryptocurrencies, including bitcoin, are non-normal, fat-tailed distributions, and they recommend fitting the *log* returns with a student's t-distribution. For more details regarding distributions which fit the simple returns of bitcoin see [9]. We will not care about the underlying distribution because, as we will see, characterizing the losses of the simple returns does not depend on the underlying distribution.

Chapter 2

Extreme Value Theory

2.1 Introduction

The concept of averaging in statistical theory is well known, and its use is widespread. A classic example of averaging comes from looking at IQ test scores. We might assume that the IQ scores of a population are normally distributed, and then use a sample average to estimate the mean and standard deviation of the population so that, in the future, we could calculate the probability of a given person's IQ score being in a certain range. By the *Central Limit Theorem*, if the sample is large enough, then the sample mean and standard deviation are precisely the parameters which give the most likely population normal distribution.

Extreme value theory (EVT), on the other hand, wants to understand events which are unlikely or extreme. This is important when we are trying to mitigate risk. For example, a hydrologist building a dam might want to know the probability, that on any given day of the year, there will be a very large amount of rain, so that they know how large to build a dam. The large amount of rain may be very unlikely, but from this probability we can estimate the average amount of time we might expect to see before this very large amount of rain, and thus we can build the dam to accommodate this event and given that it is a long time away, possibly rebuild in the future if need be. Another interesting example would be an environmental scientist studying air pollution. If biological evidence suggests that a certain concentration of ozone in the air is harmful to human health, they might use that concentration as a threshold, and then use extreme value theory to calculate the probability of the concentration of ozone being a certain number of times more than that threshold. Again, while it may be very unlikely on any given day of the year to see high ozone levels, it is important to estimate how likely this event may be. A good starting point for further examples of EVTs use can be found in Reiss and Thomas (2007).

2.2 Core Concepts

We begin with a data set $X = \{X_1, ..., X_n\}$. There is a subset of this data set $A \subseteq X$ which are considered the extreme values. There are two standard methods of determining these values: The first method, called peaks over threshold (POT), begins by choosing a threshold $0 < \mu \in \mathbb{R}$ and then taking the subset of extreme values to be $A = \{X_i \in X : 1 \leq i \leq n | X_i > \mu\}$. For the POT method, in the case where our extreme value data set of interest may have negative values, we take the negative of the data set, and then once we have chosen the threshold, take the negative of that value as our threshold for the original data set. We may also take the negative of our data set and let the threshold be a positive number, as long as the context clearly shows that it represents loss, or negative returns.

The second method, called the block maximum (BM), begins by partitioning the data set into equal parts, (or as equal as possible), and then creating the extreme value data set by taking the maximum value in each partition. If we split our data set into regular intervals say, $X_{1,k}, ..., X_{k,2k}, ..., X_{m-1k,mk}$ and then take the maximum n values from each interval, we can create our extreme value data set in this manner as well. For example, we might consider the price series of bitcoin from August 2013 to August 2016 and then partition the prices up into monthly blocks. We would then select the largest price from each month to create the extreme value data set.

The method used to extract the extreme value data from the original data set, whether it be by POT or BM, determines the type of distribution class which will be used to model the extreme data. The generalized extreme value distribution is used with the BM approach while the generalized Pareto distribution (GPD) is used with the POT approach [12]. In this project, we only consider the POT method because we want to compare different threshold choices for bitcoin with the choice made by Osterrieder (2017); see [10] for more details. Thus we are concerned with understanding the GPD. In the next section, we explore the GPD and its connection with EVT. We also develop some theory regarding the GPD.

2.3 Generalized Pareto Distribution

Reiss and Thomas begin by listing three sub-models for the GPD [12]. We have

$$W_0(x) = 1 - exp(-x), \quad x \geq 0 \tag{2.1}$$

$$W_{1,\alpha}(x) = 1 - x^{-\alpha}, \quad x \geq 1 \tag{2.2}$$

$$W_{2,\alpha}(x) = 1 - (-x)^{-\alpha}, \quad -1 \leq x \leq 0 \tag{2.3}$$

In Equations (2.2) and (2.3), α is the shape parameter of the distribution. We can redefine this shape parameter as $\xi := \frac{1}{\alpha}$ which in the literature is referred to as the *extreme value index*. From Drees et. al. (2000) we have that the Hill estimator is a popular way to estimate this parameter. We discuss this more in Sections 2.4 and 3.1. Note that $W_{1,\xi} \to W_0$ as and $W_{2,\xi} \to W_0$ as $\xi \to 0$. This leads us to the following combined model for the GPD.

Definition 2.1 [Generalized Pareto Distribution]

$$F_{(\xi,\mu,\beta)}(x) = \begin{cases} 1 - (1 + \frac{\xi(x-\mu)}{\beta})^{-\frac{1}{\xi}} & \text{for } \xi \neq 0 \\ 1 - exp(-\frac{x-\mu}{\beta}) & \text{for } \xi = 0 \end{cases}$$

where $\beta > 0, \xi, \mu \in \mathbb{R}$.

The scale parameter for this distribution is β, the shape parameter ξ, and the location parameter is μ which in extreme value theory is the threshold for the POT method. This cumulative distribution has the usual properties required, i.e.,

$$0 \leq F_{(\xi,\mu,\beta)}(x) \leq 1, \forall x \in dom(F_{(\xi,\mu,\beta)}(x)),$$

$$\lim_{x \to -\infty} F_{(\xi,\mu,\beta)}(x) = 0,$$

$$\lim_{x \to \infty} F_{(\xi,\mu,\beta)}(x) = 1,$$

$$F_{(\xi,\mu,\beta)} \text{is non-decreasing.}$$

In the first case, when $\xi \neq 0$, we have that

$$F_{(\xi,\mu,\beta)}(x) = \begin{cases} 1 - (1 + \frac{\xi(x-\mu)}{\beta})^{-\frac{1}{\xi}} & \text{for } \xi > 0 \implies x \geq \mu \\ 1 - (1 + \frac{\xi(x-\mu)}{\beta})^{-\frac{1}{\xi}} & \text{for } \xi < 0 \implies \mu \leq x \leq \mu - \frac{\beta}{\xi} \end{cases}$$

We point out these two different cases of $\xi \neq 0$ to emphasize that when $\xi > 0$, we have an infinite support. And when $\xi < 0$, our support is finite. The GPD is a parametric probability density function (pdf) used in the POT estimation of the tail of a distribution. The pdf is defined as follows.

Definition 2.2

$$f_{(\xi,\mu,\beta)}(x) = \begin{cases} (1 + \frac{\xi(x-\mu)}{\beta})^{-(1+1/\xi)} & \xi \neq 0 \\ exp(-\frac{x-\mu}{\beta}) & \xi = 0 \end{cases}$$

We note here the analytic relationship $f_{(\xi,\mu,\beta)}(x) = F'_{(\xi,\mu,\beta)}(x)$ between the cumulative distribution and its density. We show the three different cases for the shape parameter in Figure 2.1.

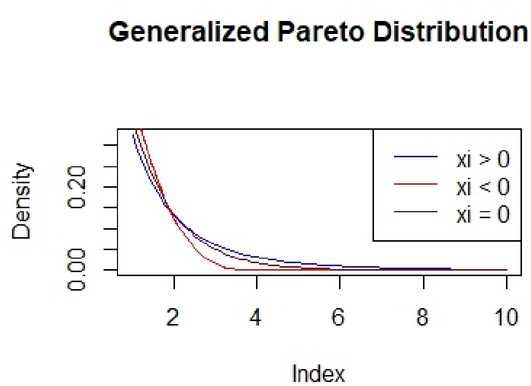

Figure 2.1: Three different cases of the shape parameter for the GPD.

In general, the conditional excess distribution is given by.

Definition 2.3 [Conditional Excess Distribution] $F_\mu(x) = P(X \leq x | X > \mu)$.

If F is the cumulative distribution function then this definition can be rewritten as follows.

Definition 2.4 $F_\mu(x) = \frac{F(x)-F(\mu)}{1-F(u)}$, $x > \mu$.

Reiss and Thomas (2007) note that GPDs are the only continuous distribution functions which have what is called *POT-stability*. This property is specified by the following.

Definition 2.5 [POT-Stability] For a certain choice of constants b_μ and a_μ,

$$F_\mu(b_\mu + a_\mu x) = G(x) \tag{2.4}$$

where G is a GPD.

What this means is that, for the underlying distribution F from which the data is sampled, if F_μ is shifted and scaled by specific constant factors, then F_μ becomes exactly a GPD.

The justification for using the GPD to model the excesses comes from the Pickands-Balkema-de Haan theorem [10].

Theorem 2.6 (Pickands-Balkema-de Haan) *Let $X_1, ..., X_n$ be a sequence of independent and identically distributed random variables with F_μ as their conditional excess distribution function. For a large class of underlying distribution functions F, if the threshold μ is large, then we have that F_μ is well approximated by the GPD.*

This theorem allows us to use the GPD to model the tail of bitcoin returns. We need a large enough threshold in order to justify applying the Pickands-Balkema-de Haan theorem, but we also need a small enough threshold in order to keep the variance of the model low [10]. As we will see, this idea comes into play when using different methods to select the threshold.

2.4 Maximum Likelihood Estimation

Fitting a distribution to a set of data depends on the type of the distribution. There are models which depend on a finite or countable amount of parameters. These models are called parametric. The GPD is a parametric model, and thus it is appropriate to estimate the best parameters based on maximum likelihood estimation (MLE). We are assuming our data to be independent and identically distributed (iid), meaning that each data point is independent from all other data points, i.e., $P(X_i|X_j) = P(X_i), \forall i \neq j, 1 \leq i, j \leq n$ and identically distributed meaning that each data point is assumed to be a sample from the same distribution. Then we can consider the likelihood function $L(\theta)$ defined as follows.

$$L(\theta) = \prod f(X_i|\theta)$$

where θ is the vector of parameters for the probability distribution, f. Since our data is given to us, we want to choose the value of parameters $\hat{\theta}$ that maximizes the likelihood function. We can do this using calculus by taking the first derivative of the likelihood function, setting it equal to 0 and solving for θ, i.e., solving the following equation for θ.

$$\frac{d}{d\theta}[L(\theta)] = 0.$$

Since we have a function of two variables (recall that the threshold of the GPD is fixed prior to parameter estimation), we can use the partial derivatives to estimate the parameter vector [14]. Since we are dealing with a product, we may want to create a summation instead by taking the *logarithm* of the likelihood function. The *logarithm* is a strictly increasing function which means the *log* likelihood reaches its maximum at the same location that the likelihood function does. And thus we solve.

$$\frac{d}{d\theta}[log(L(\theta))] = 0$$

to find our parameter vector θ. Drees et. al. (2000) state that the Hill estimator described in Definition 3.2, with $H(n-1)$, is the maximum likelihood estimator of the extreme value index for the case where the index is positive.

Chapter 3

Estimating the Threshold

In many cases, precedent or some standard, will set the threshold of the extreme value data set. For example, if we know that a certain amount of particulate in the air is harmful for human health, we would want to model the particulate amount above this threshold. In the case of bitcoin, we want to know what extreme loss will be bad for our monetary health. Bader and Yan (2016) suggest some various "rules of thumb" have been applied to find the threshold. These are "select the top 10% of the data" or "select the top square root of the sample size" [2]. For us, since we are considering $1,077$ negative bitcoin returns, this would amount to either the largest 108 negative returns, or the largest 33 negative returns respectively. This is quite a disparity. Traditionally, the mean residual life (MRL) plot or mean-excess plot was used to subjectively estimate the threshold. Far and Wahab (2016) defines this plot as follows.

$$\big(\mu, \frac{1}{n_\mu}\sum_{i=1}^{n_\mu}(X_i - \mu)\big) : \mu < X_{max}$$

, where $X_i > \mu, i = 1, ..., n_{mu}$. We then plot this set for a range of threshold choices μ. We can think of the plot as showing the excess mean over the corresponding threshold.

Osterrieder and Lorenz (2017) use the negative of the bitcoin/USD returns to create the MRL plot. Figure 3.1 shows this plot, and from this plot they choose a threshold of $\mu = -0.004$ [10]. This choice corresponds to an extreme value data set with the 150 largest negative returns.

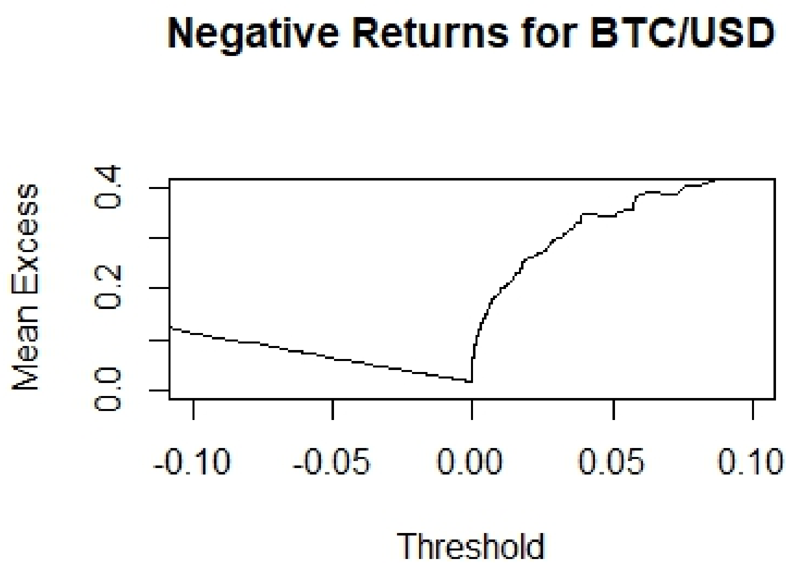

Figure 3.1: MRL plot of the *log* normal bitcoin to dollar returns.

This choice is justified since the increase in the MRL plot after this point indicates a fat-tail and "a positive gradient above some threshold is a sign of Pareto behavior in the tail" [10]. Using the MRL plot to determine the threshold allows for some subjectivity since we are visually choosing a value beyond which the graph is approximately linear [10].

3.1 Hill Plot

The Hill plot has been used to help subjectively determine the threshold of an extreme value distribution. Drees et. al. (2000) describe the Hill estimator as a way to estimate the *extreme value index*, which if we recall from Section 2.3, is defined as $\xi := \alpha^{-1}$. This value is made explicit in the following equation. They begin by considering a heavy tailed distribution F which they define as a distribution which satisfies the following condition.

$$1 - F(x) \sim x^{-\alpha} L(x), x \to \infty, \alpha > 0 \tag{3.1}$$

"where L is a slowly varying function satisfying $\lim_{t\to\infty} \frac{L(tx)}{L(t)} = 1$ for all $x > 0$" [4]. The value $1 - F(x)$ represents the tail of the distribution F. This equation says that the tail has a distribution which is slowly decaying at a rate which is based on the value of α.

To begin, we first need to order our data $X_{(1)} \leq ... \leq X_{(n)}$. Note here that we require our data to be non-zero and positive. Then, the Hill estimator is defined as follows.

Definition 3.1 [Hill Estimator] $H(k) = \frac{1}{k} \sum_{j=1}^{k} \ln(X_{(n+1-j)}) - \ln(X_{(n-k)})$

where $k = 1, ..., n-1$. From Zhou et. al. (2007), the variance of the Hill estimator is stated as follows.

$$Var(H(k)) \propto \frac{1}{k}. \tag{3.2}$$

As $n \to \infty$, for a sequence $k_n \to \infty$ such that $\frac{k_n}{n} \to 0$, we have that $H(k_n) \xrightarrow{P} \xi$ [4]. As was discussed in Section 2.4, $H(n-1)$ is the maximum likelihood estimator for the positive shape parameter of the GPD.

To create a Hill plot, we take the reciprocals of $H(k)$ from Definition (3.1) and plot them against k. Figure 3.1 shows the Hill plot of the negative bitcoin returns.

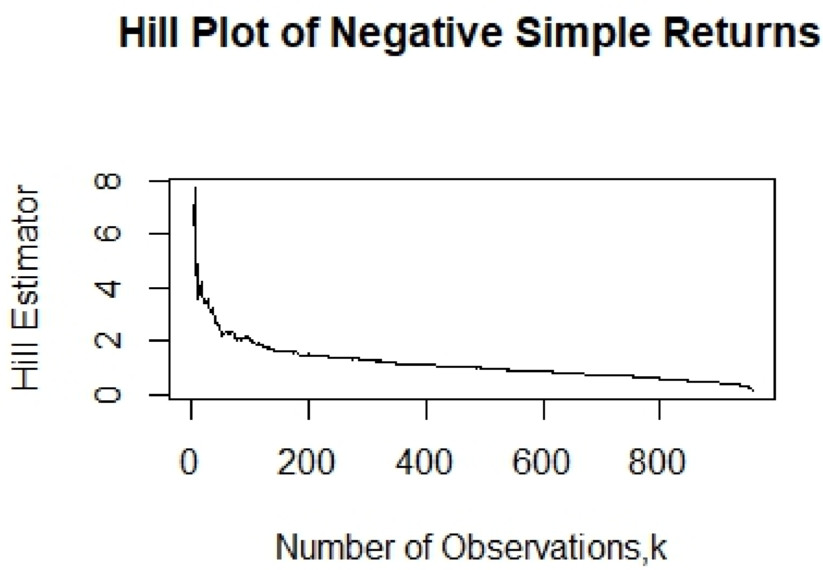

Figure 3.2: Hill plot for negative bitcoin returns.

When there are extreme values in our data set, for small values of k, the Hill plot will vary greatly. Once k increases beyond a certain value, the Hill plot will have small variance and can be fitted with a straight line. We use this point, called the turning point, to choose our threshold. Bader and Yan (2016) describe this value as follows: “The threshold is the kth smallest order statistic beyond which the parameter estimates are deemed to be stable." Previously, this turning point was calculated subjectively by observation [15]. In Figure 3.2 we notice that from observations 0 to about 50 on the x-axis, the Hill estimator varies quite a lot. After this, the Hill estimator is just about linear. Then subjectively, we might choose the 50 largest negative returns as our extreme value data set. Zhou et. al. (2007) use the Hill plot as part of their algorithm to calculate the threshold for the extreme values of copper futures.

3.2 Quantifying the Threshold

Reiss and Thomas (2007) give an ad hoc procedure as an automatic choice for the number of extremes. They begin by defining the following function.

$$A(k) := \frac{1}{k} \sum_{i \leq k} i^{\beta} |\xi_i - med(\xi_1, ..., \xi_k)| \tag{3.3}$$

where ξ_k is the calculated shape parameter based on k upper extremes and β is a smoothing parameter. They choose the threshold as the k^* value which minimizes this function. Reiss and Thomas (2007) note the following when calculating the GPD parameters for k upper extremes: "if k is small, there is a strong fluctuation of the shape parameter for varying k. For an intermediate k of extremes, the values of the estimates for the shape parameter stabilize around the true value. If k is large, the model assumption may be strongly violated and one observes a deviation of the shape parameter." This fluctuation may also be observed in the Hill plot of the bitcoin return series. Using this procedure, we calculated $k^* = 67$ which corresponds to a threshold of $\mu = 0.04117647$ which is about a 4% price decrease.

Zhou et. al. give a new method which is based on the Hill plot which calculates the *turning point* and thus the threshold for the POT model [15]. This *turning point* can be thought of as the first value where the Hill estimator stabilizes, as described in Reiss and Thomas (2007). They first characterize the stationary part of the Hill plot by fitting a straight line to a subset of this plot, in the range of $[n/10] \leq k \leq [n/2]$, where $[\cdot]$ represents the nearest integer function. The procedure is as follows: let $\kappa_1 = [n/10]$ and $\kappa_2 = [n/2]$. First, we want to estimate the parameters β_0 and β_1 for the following.

$$H(k) = \beta_0 + \beta_1 k + \epsilon(k) \tag{3.4}$$

where $\epsilon(k)$ is the error term with $E(\epsilon(k)) = 0$. Here we note that the error term is assumed non-constant and depends on our index k. We make this assumption since we know that the variance of the Hill plot decreases as k increases. Then the parameter estimates are given by the following.

$$\hat{\beta} = (Z'W^2Z)^{-1}Z'W^2H \tag{3.5}$$

where

$$Z = \begin{bmatrix} 1 & \kappa_1 \\ 1 & \kappa_1 + 1 \\ \vdots & \vdots \\ 1 & \kappa_2 \end{bmatrix}$$

$$H(k) = (H(\kappa_1), H(\kappa_1 + 1), ..., H(\kappa_2))'$$

$$W = \begin{bmatrix} \sqrt{\kappa_1} & & & \\ & \sqrt{\kappa_1 + 1} & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & & & \sqrt{\kappa_2} \end{bmatrix} \tag{3.6}$$

The fitted line is then used to calculate the turning point. We first want to consider the stationary series of residuals

$$e(k) = H(k) - (\hat{\beta}_0 + \hat{\beta}_1 k) \tag{3.7}$$

where $1 \le k \le \kappa_2$.

Then we calculate the standard deviation of this series for the interval $\kappa_1 \le k \le \kappa_2$.

$$\hat{s} = \sqrt{\frac{1}{\kappa_2 - \kappa_1} \sum_{k=\kappa_1}^{\kappa_2} e(k)^2} \tag{3.8}$$

Then we have

$$k^* = max\{k : |e(k)| \ge \lambda \hat{s}, 1 \le k \le \kappa_1\} \tag{3.9}$$

and finally, the threshold is $\mu = X_{(n-k^*)}$.

As k increases, we are getting closer to the plateau of our Hill plot. We want the largest k value because we want the closest k value to our plateau which cannot be modeled within error of our line. We noted earlier that the error term (ϵ) of Equation (3.4) is non-constant, which violates the assumption of a constant variance, error term, for ordinary least squares regression. Equation (3.5) is the *weighted least squares* estimation of the parameters [11]. The matrix W in Equation (3.6) is a matrix of weights where each weight is the reciprocal of the co-variance of the data point associated with the *kappa* value. In the matrix W, the values along the diagonal increase in value, reflecting a decrease in variance and an increase in relative information [11]. In Equation (3.9), λ is a "proper positive factor" [15]. It's important here to note that the "proper positive factor" is a subjective choice. Zhou et. al. (2007) conclude that the threshold must not exceed 10% of the data. Therefore we only consider thresholds in the range of $1 \leq k \leq \kappa_1$. Zhou et. al. (2007) describe the turning point (threshold) as having "a relatively large deviation from the fitted stationary straight line." The residuals $e(k)$ give us deviations from the fitted stationary line. Thus, the turning point will be part of the residual series. We are choosing the largest k value corresponding to a residual which is beyond the range of the error term of the stationary line. To make this description more concrete, if $k^* = 40$ then this means that the value $H(40)$ was not well estimated by the stationary line and that the stationary line represents the Hill estimator within the error range beyond this k^* value.

Chapter 4

Results

We decided to calculate the k^* values using the method in Zhou et. al. (2007) for a range of λ values from 0 to 3 by 0.01. There were a total of 4 values calculated. The results of the k^* values and their corresponding thresholds can be seen in Table 4.1.

k^*	Threshold
3	\$161.61
48	\$29.52
93	\$18.09
96	\$17.83

Table 4.1: Thresholds for losses calculated by the algorithm from Zhou et. al.

We also plotted these values on our Hill plot, which can be seen in Figure 4.1.

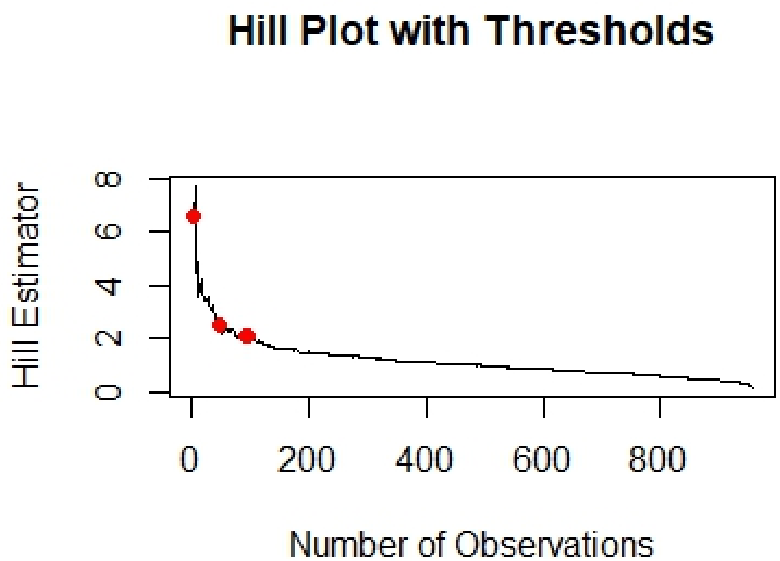

Figure 4.1: The red dots are where $k^* = 3, 48, 93, 96$.

We can reject the threshold associated with $k^* = 3$ based on the Hill plot because this value is in the region of the Hill plot where visually we can see the high fluctuation. Thus, we consider the other 3 thresholds as viable candidates. In addition to these three thresholds, we will include the threshold that was chosen by Osterrieder and Lorenz (2017), which is the negative return value of \$10.91 associated with 150 extreme value data points in the tail. We will also include the values calculated from the rules of thumb in the beginning of chapter 3 and the value calculated from the ad hoc procedure in Reiss and Thomas (2007). GPD models were fit for each of these thresholds, and the resulting parameters are shown in Table 4.2.

Method	Model Number	k^*	Threshold	Normalized Threshold	ξ	β
Quantification	1	48	\$29.52	0.1219	-0.1685	0.0792
Quantification	2	93	\$18.09	0.0849	0.0267	0.0596
Quantification	3	96	\$17.83	0.0833	0.0343	0.0588
MRL plot	4	150	\$10.91	0.0560	0.0212	0.0595
Ad Hoc	5	67	\$23.30	0.1023	-0.0719	0.0695
Rule of Thumb	6	108	\$15.43	0.0745	-0.0063	0.0628
Rule of Thumb	7	34	\$37.25	0.1501	-0.1601	0.0738

Table 4.2: All of the calculated thresholds with their corresponding parameters for their GPDs.

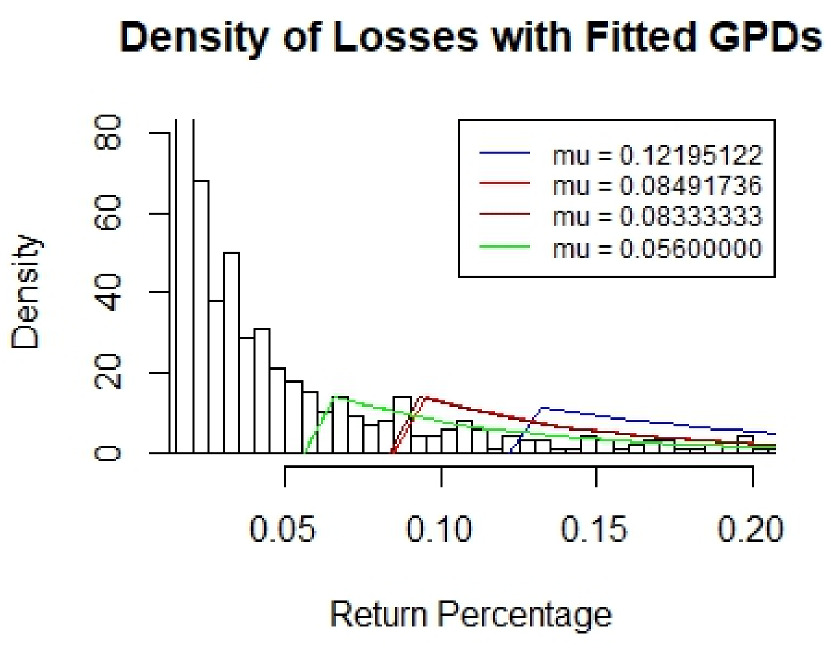

Figure 4.2: The calculated parameters along with their thresholds.

Figure 4.2 shows the densities of the GPD models plotted over the negative returns. Notice that the negative shape parameter corresponds to the first threshold choice of models $1, 5, 6, 7$. This is important because the GPD model associated with a negative shape parameter has a finite support. In the case of model 1 for example, its support is from $\mu \leq x \leq \mu - \frac{\beta}{\xi}$ which is $0.1219 \leq x \leq 0.1219 - \frac{0.0792}{-0.1685}$. In practice, we want to use a probability distribution which does not assume a finite support since theoretically there is no upper limit to the gains or losses of a financial device, since there is no upper limit to its price. The case in point is the maximum price of bitcoin in August 2016 was \$1,147.00 while the maximum price of bitcoin in the year 2017 was \$19,247.37. It has since decreased in value but still remains well above the maximum price of 2016.

Far and Wahab (2016) gives a way to calculate the mth observation return level, based on the threshold choice, and fitted scale and shape parameters of the GPD. This is, in their words, "the value which is exceeded on average once every m observations." They calculate this value with the following formula:

$$x_m = \mu + \frac{\beta}{\xi}[(m \cdot \frac{N}{N_t})^{\xi} - 1] \tag{4.1}$$

where μ is the threshold, β and ξ are the scale and shape parameter respectively, N is the total number of data points beyond the threshold, and N_t is the total number of data points. Table 4.3 shows the calculated mth observation values for approximately one trading week, one trading month, and one trading year respectively.

Model Number	m = 5	m = 20	m = 240
1	-0.0905	0.0517	0.2366
2	-0.0056	0.0751	0.2276
3	-0.0037	0.0755	0.2274
4	-0.0071	0.0747	0.2278
5	-0.0374	0.0675	0.2313
6	-0.0138	0.0736	0.2285
7	-0.0864	0.0524	0.2359

Table 4.3: The values we expect once every m observations.

The negative values for one trading week suggest that every five trading days an investor might expect one day where they lose something small beyond zero or even have no return. This is consistent with the large amount of zero returns for bitcoin. For a smaller trading period of about a month, an investor could expect one day where they lose about 5% or more for models 1 and 7, and about 6.7% to 7.5% or more with respect to the other models. And for a trading period of about a year, an investor could expect about one day where they lose 22.70% to 23.60% or more. In the short term, the negative values of all of the models indicate an expected loss of about zero while the smaller monthly return loss and slightly larger yearly return loss may indicate that models 1 and 7 are slightly more optimistic over shorter time intervals and less so in the long run.

Another way of deciding which model to use is to see how it performs with new data or future data. The models are fit with data from August 2013 to August 2016. We can now look at the data from August 2013 to August 2018 and compare the empirical probability against the probability given by the model. The calculated values for the models can be seen in Table 4.4 and the empirical calculated values can be see in Table 4.5.

Model Number	$P(X \leq 0.20 \mid X > \mu)$	$P(X \leq 0.30 \mid X > \mu)$
1	0.6592	0.9405
2	0.8477	0.9680
3	0.8532	0.9688
4	0.9052	0.9802
5	0.7726	0.9584
6	0.8658	0.9734
7	0.5103	0.9140

Table 4.4: Probability values calculated by the models fit on data from August 2013 to August 2016.

Corresponding Model	Threshold	$P(X \leq 0.20 \mid X > \mu)$	$P(X \leq 0.30 \mid X > \mu)$
1	0.1219	0.7547	0.9622
2	0.0849	0.8879	0.9827
3	0.0833	0.8916	0.9833
4	0.0560	0.9389	0.9906
5	0.1023	0.8414	0.9756
6	0.0745	0.9057	0.9855
7	0.1501	0.6388	0.9444

Table 4.5: Empirical probabilities based on bitcoin data from August 2013 to August 2018.

We then calculate the mean squared error (MSE) for each model (k) relative to its predicted value and the empirical calculation. This is defined as:

$$MSE_k := \frac{1}{2}\sum_{i=1}^{2}(predicted_i - empirical_i)^2 \tag{4.2}$$

Where i represents the index corresponding to the calculated value at 0.2 and 0.3. The results are displayed in Table 4.6.

Model Number	MSE
1	0.0047896323
2	0.0009159099
3	0.0008433904
4	0.0006212563
5	0.0025131992
6	0.0008713189
7	0.0087186384

Table 4.6: MSE between the calculated and empirical probabilities.

We can see that the model with the smallest MSE is model 4 corresponding to the model from Osterrieder and Lorenz (2017). Not far behind are models $2, 3, 6$ which correspond to two models calculated by the algorithm from Zhou et. al. (2007) and the ad hoc procedure in Reiss and Thomas (2007). Those models have a very small difference in their MSE, yet using them will give one different probability predictions. Thus, we can recommend that it is up to the user to decide their risk tolerance and their outlook with respect to the future, and consider these factors when choosing a threshold and model.

Chapter 5

Conclusion

Bitcoin continues to be a subject of controversy. Advocates believe bitcoin should be left alone to develop and mature. Skeptics believe that it has a positive future as long as it is has the right restrictions and regulations, while opponents believe it will be a passing fad. Within the financial community, potential investors see bitcoin as very unstable and full of promise while institutional investors see bitcoin as unstable and full of risk. Large risk, large reward, and large losses seem to be the daily headlines. For example, on November $20, 2018$ a Forbes article [1] entitled, "Bitcoin: What is Behind the Recent Plunge & How Bad Can it Get?" talks about why bitcoin's price has dropped by 78% since its all-time peak of nearly $\$20,000$ in 2017.

As we have shown in this report, bitcoin is highly non-normal with many extreme return values. The prevalence of these dramatic price swings should alarm anyone who takes seriously the notion that it will replace our current cash system in the near future. We have also shown that estimating the cutoff value of extreme returns can be either subjective or quantifiable and that these two methods lead to different models and different conclusions about the risk associated with the extreme values. The subjective approach to decide the cutoff value of extreme returns can be just as valid as the quantifiable approach given the right expertise of the analyst. However, different thresholds will give different models for our data. For example, comparing the calculations from choosing model 4 from Osterrider and Lorenz (2017), and model 2 calculated by the algorithm from Zhou et. al. (2007), we have a big difference in their corresponding thresholds - namely \$18.09 and \$10.91. The expected values exceeded on average once every m observations are not very different between these two models. However, if we look at the comparison between our predicted probabilities and the probabilities calculated by future data, it suggests that model 2 is less accurate than model 4 but not by much. Model 4 was the most accurate, according to MSE, in predicting the future data. It's important to note that the models with the largest MSE were also those estimated to have a negative *shape* parameter and thus a finite support. This must be taken into consideration by the practitioner.

For future work, it is important to note that there are more quantification algorithms in the literature; see [13] and [2] for more details. It would be important going forward, to compare the performance of the models created by even more methods of threshold selection. In EVT, we can use the GPD to estimate the entire cumulative distribution for our data. It would be interesting to see the performance of these so-called mixture models relative to their underlying GPDs. There are also other ways of using the GPDs which would allow for comparison, including value-at-risk and expected shortfall calculations; see [10] and [12]. Despite the suggestions made by Hudson and Gregoriou (2015), one still might perform the extreme value analysis on the logarithmic return series since this procedure is common in the financial literature; again see [10] and [15]. In this report, the BM method was not explored because comparing different thresholds for the POT method was the primary focus. Thus, it might be worth analyzing this method. However, there is an analytic relationship, with some restrictions, between GPDs and the distributions associated with the BM approach according to Reiss and Thomas (2007). Therefore, it may not yield a significant difference.

Bibliography

[1] Aslam, N. (2018, November 20). *Bitcoin: What is Behind the Recent Plunge & How Bad Can it Get?*. Retrieved from https://www.forbes.com/sites/naeemaslam/2018/11/20/bitcoin-what-is-behind-the-recent-plunge-how-bad-can-it-get/

[2] Bader, B. & Yan, J. & Zhang, X. (2016). Automated Threshold Selection for Extreme Value Analysis via Goodness-of-Fit Tests with Application to Batched Return Level Mapping. *arXiv:1604.02024*.

[3] Ccn. (2018, May 12). *Big Banks are Investing Heavily in Blockchain and Crypto: $364 Billion Investment Firm.* Retrieved from https://www.ccn.com/big-banks-are-investing-heavily-in-blockchain-and-crypto-364-billion-investment-firm/.

[4] Drees, H. & de Haan, L. & Resnick, S. (2000). How to Make a Hill plot. *The Annals of Statistics, Vol. 28, No. 1*, pp. 254-274.

[5] Far, S.S. & Wahab, A.K.Abd. (2016). Evaluation of Peaks-Over-Threshold Method. *Journal of Ocean Science.*

[6] Hudson, R. & Gregoriou, A. (2015). Calculating and Comparing Security Returns is harder than you think: A Comparison between Logarithmic and Simple Returns. *International Review of Financial Analysis Volume 38*, pp. 151-162.

[7] Jarque C.M. (2011). Jarque-Bera Test. In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg.

[8] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.

[9] Osterreider, J. (2016). The Statistics of Bitcoin and Cryptocurrencies. *Zurich University of Applied Sciences*.

[10] Osterrieder, J. & Lorenz, J. (2017). A Statistical Risk Assessment of Bitcoin and Its Extreme Tail Behavior. *Annals of Financial Economics, Vol. 12, No.1.*

[11] Pennsylvania State University (2018). *Weighted Least Squares*. Retrieved from https://onlinecourses.science.psu.edu/stat501/node/352/.

[12] Reiss, R.-D. & Thomas, M. (2007) *Statistical Analysis of Extreme Values 3rd Edition*. Boston, MA: Birkhauser.

[13] Scarrott, C. & MacDonald, A. (2012). A Review of Extreme Value Threshold Estimation and Uncertainty Quantification. *REVSTAT - Statistical Journal, Vol. 10, No. 1*. pp. 33-60.

[14] Wasserman, L. (2004) *All of Statistics*. New York, NY: Springer.

[15] Zhou, C. et al., (2007). A New Method to Choose the Threshold in POT Model. *Financial Engineering Research Center for Shanghai Jiaotong University*.