An Analysis of S&P 500 Index Returns: An Extreme Value Theory Approach

Andrew Benson

December 15, 2021

1 Introduction

In the years following the 2008 financial crisis, banks stepped up efforts to adequately capture risk exposure in the businesses they support. In financial analytics, the use of the Gaussian distribution is widespread to model asset returns such as in the Capital Asset Pricing Model (CAPM), Modern Portfolio Theory (Markowitz-Theory), the or Black-Sholes options pricing model to model Brownian motion. In reality, most assets manifest more rare events than the Gaussian distribution would suggest. Although the Gaussian distribution does well to model central tendency, the distribution underestimates the amount of risk often seen in the real world. The tails the distribution are too thin to allow for rate events. Although many studies have attempted to find a way to deal with the amount of kurtosis displayed in empirical asset returns, researchers and financial professionals continue use the Gaussian distribution despite its shortcomings due to its theoretical properties such as being a stable distribution which works well for portfolio analytics [3]. If a measure of central tenancy is not needed directly in analysis, extreme value theory lends itself well when modeling the extremes and can provide a more accurate picture of tail risk. In this paper, extreme value theory will be applied to S&P 500 index data to see if the theory provides a good fit for the data. A few applications will also be discussed.

2 Theory

In extreme value theory there are two approaches to modeling extreme values, the block maximum approach and and peaks over threshold approach (POT). Although this paper will focus on the block maxima approach, the POT approach will be briefly introduced.

2.1 Block Maxima Approach

If a block maxima approach is used, the data collected are typically measurements over time. Time is then subdivided into chunks, for example one year, but can be smaller or larger depending on how often the data were able to be sampled. The annual maxima series is then generated from the overall sample. In other words, the maximum for each year is found from the sample of data collected. The annual maxima series is then fitted to one of three distributions outlined in the following theorem.

Theorem 1[a]: Suppose $M_n = \max\{X_1, ..., X_n\}$ where $X_1, ..., X_n$ is a sequence of independent random variables having common distribution F. If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n > 0\}$ such that

$$\Pr\{(M_n - b_n)/a_n\} \leq z\} \to G(z) \text{ as } n \to \infty$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\begin{aligned}
\mathrm{I}: G(z) &= \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, -\infty < z < \infty \\
\mathrm{II}: G(z) &= \begin{cases} 0, & z \leq b \\ \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, & z > b \end{cases} \\
\mathrm{III}: G(z) &= \begin{cases} \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b \\ 1, & z \geq b \end{cases}
\end{aligned} \tag{1}$$

[a]For a reference on Theorem 1, please see reference [1].

Explained in other words, Theorem 1 states that $M_n* = (M_n - b_n)/a_n$

converges in distribution to one the three families (I, II, or III) which are known as the Gumbel, Fréchet, and Weibull families respectively. A weakness of Theorem 1 is that it can be difficult to know which one of the three distributions to choose. Indeed this is a nontrivial task. Thankfully, it is possible to rewrite the Gumbel, Fréchet, and Weibull families such that they are combined into a single family of models known as the Generalized Extreme Value distribution function or (GEV) distribution which can be written as

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}. \tag{2}$$

Maximum likelihood estimation (MLE) is then typically used find the best fit parameters.

2.2 Peak Over Threshold Approach

The peak over threshold approach, although not applied to the data in this paper, will be briefly mentioned here. This is another approach that seeks to generate a sample of maximums or minimums. Instead of creating blocks of time such as in the block maxima approach, a line is drawn so to speak, where only observations that exceed the threshold are considered. This subset of the data containing only extreme observations can be fitted to a distribution through a different theorem than what was stated in Theorem 1. Asymptotic results and theorems, not given in this paper, claim that the data can be fit to a generalized Pareto distribution [1].

3 Modeling

Data for daily adjusted closing prices on the S&P 500 from were collected from 1960 through 2020. A plot of the data over time is shown in Figure 1. Although theory allows that distributions of a the minimum or maximum may be constructed, for this analysis, only the minimum will be considered. As introduced above, a block maximum approach will be used for modeling. The time series data was divided into partitions of one year. The minimum return was then found for each year creating an annual minima series. A plot of the annual minima series is shown in Figure 2. Before jumping straight into making a model, some of the assumptions of Theorem 1 will be examined to make sure they are satisfied as best as possible.

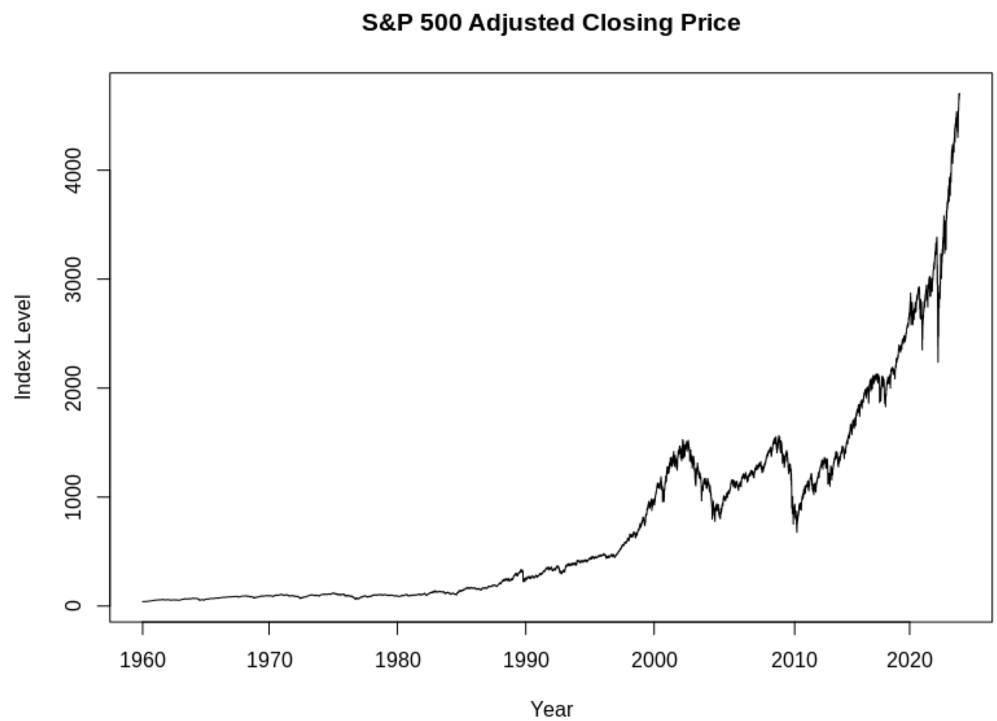

Figure 1

3.1 Model Assumptions

For this paper the data are assumed to be independent. This seems reasonable given research on claims asset prices can be modeled as a martingale [4]. Further, when working with time series data, such as the stock data here, a common concern is lack of stationarity. A lack of stationarity would violate the assumption in Theorem 1 that, $X_1, X_2, ...X_n$ have a common distribution, as well as cause misleading results when attempting to perform extreme value analysis. Clearly the S&P 500 index level measurements are not identically distributed as the mean continues to rise over time. As a suggested solution, Coles claims that empirical studies have shown an approximation to stationarity can be achieved by taking the logarithm of ratios of successive observations [1]. This operation results in what is also known as log-daily

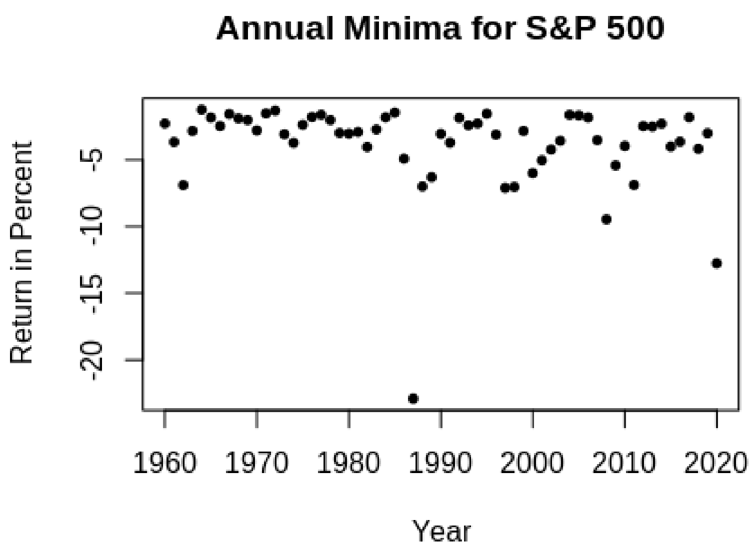

Figure 2: A plot of the minimum return for each year using the block maxima approach.

returns. Log-daily returns can be written mathematically as

$$r_t = \ln(p_t/p_{t-1}) \tag{3}$$

where r_t represents the return in time period t and p_t represents the S&P adjust closing price in time period t. A plot of the log daily returns is shown in Figure 3. The mean is now centered around zero and the variance remains somewhat constant through time. The log-daily returns were used for model fitting, graphics, and all analysis for this point forward.

3.2 Model Fitting

For modeling, R statistical software was used in conjunction with the 'evd' package [5]. The evd package provides a maximum likelihood estimation function to calculate the parameters of the GEV distribution. Note that before performing the maximum likelihood estimation, the block minima data were changed from negative values to positive values since all the data were negative. The MLE estimates are shown in Table 1.

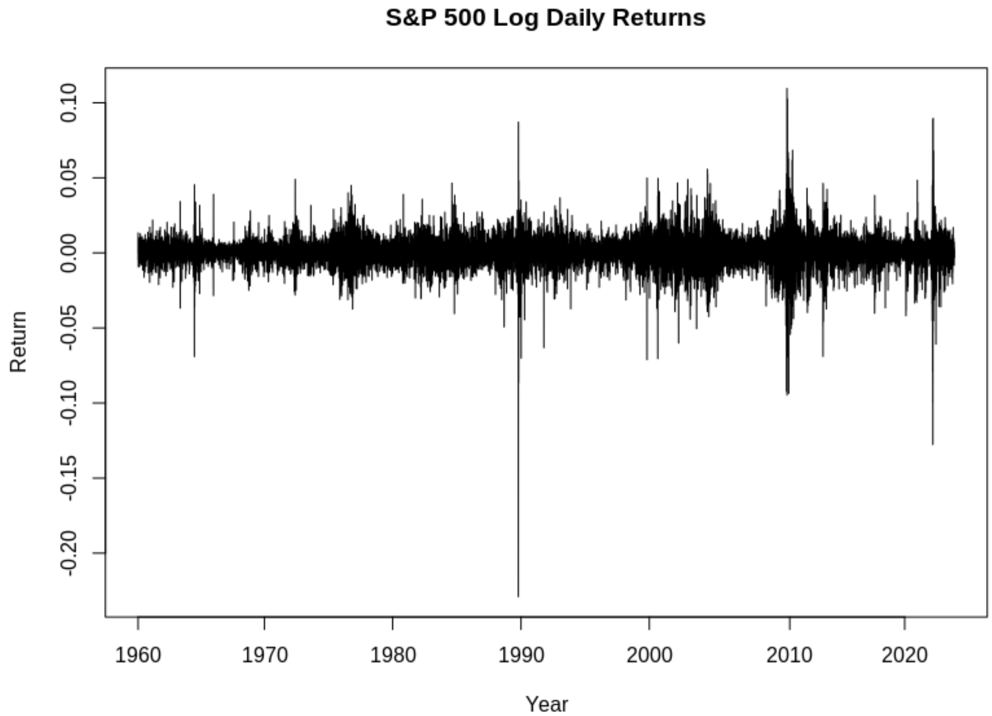

Figure 3: An approximation to stationarity can be generated by log daily returns which is shown here.

	Lower Bound	Point Estimate	Upper Bound
$\hat{\mu}$ (location)	0.0206	0.0237	0.0268
$\hat{\sigma}$ (scale)	0.0079	0.0106	0.0134
$\hat{\xi}$ (shape)	0.2202	0.5181	0.8159

Table 1: Maximum likelihood estimates for the GEV distribution along with the upper and lower bounds found using a 95% confidence interval.

Note that the shape parameter, $\hat{\xi}$, from Table 1 is greater than zero and has a small confidence interval $\hat{\xi}$. Since this parameter is greater than zero, even when considering lower bound of the confidence interval, theory suggests the data best fit a corresponding distribution be of Type II, or the Fréchet distribution. A plot of the Fréchet probability density function with the maximum likelihood estimates is shown in Figure 4.

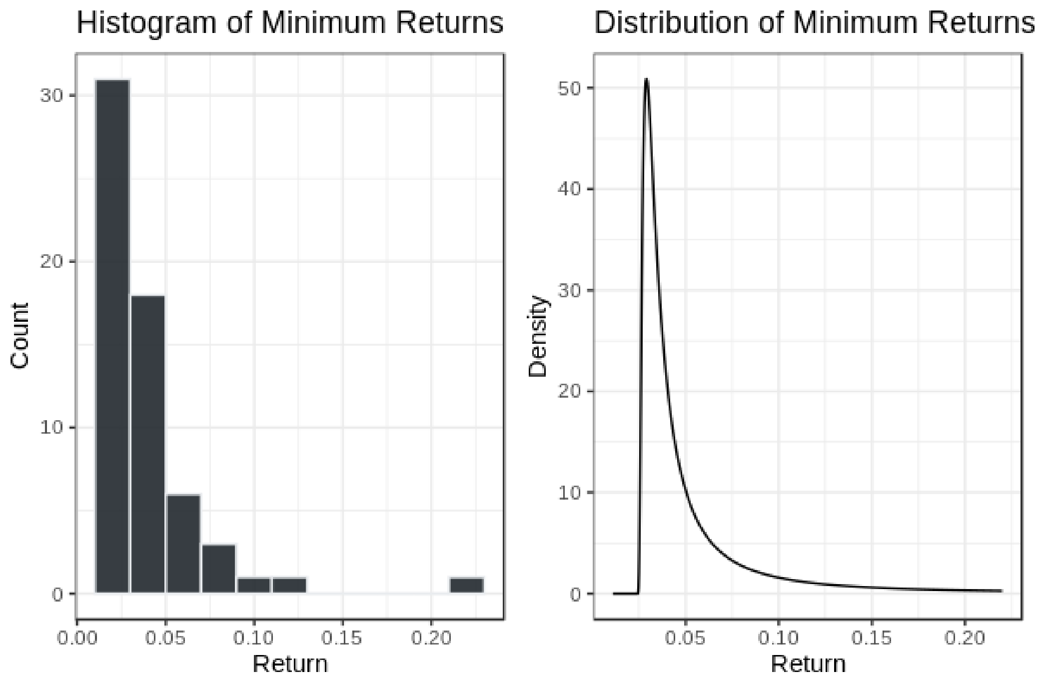

Figure 4: On the left is a histogram of the block minima series log-daily return data. On the right is the Fréchet distribution fitted with the MLE location and scale point estimate parameters from Table 1.

A probability plot was created to check the fit of the data. The empirical distribution function (ECDF) was checked against the fitted Fréchet distribution and the results are shown in Figure 5. In the left panel of Figure 5, the probability plot highlights an important issue with the fit of the data to the Fréchet distribution. The issue is the fitted distribution does poorly at estimating smaller losses; around zero to two percent. When the losses are around zero to two percent, the fitted distribution predicts probabilities of zero where the ECDF shows a higher probability of losses in that range. Losses between zero and two percent were then truncated for plotting purposes to show that once the losses are larger than two percent, the data match the fitted distribution better. This change is shown in the right panel of Figure 5. This is somewhat intuitive. A loss of only one or two percent isn't really an extreme event and could likely be captured adequately by a Gaussian distribution.

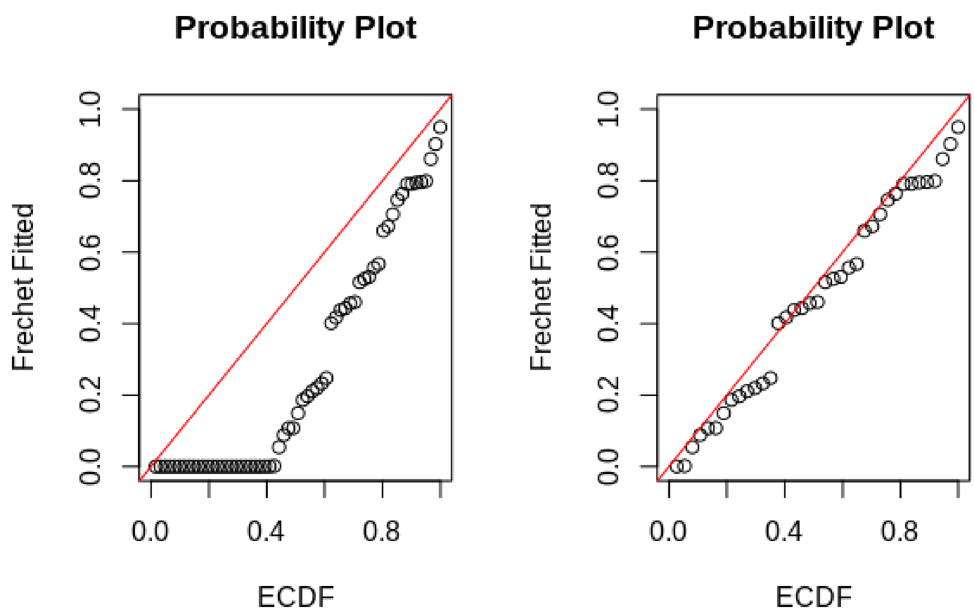

Figure 5: Probability plots of the ECDF vs the Fréchet distribution fitted with the MLE point estimates. The plot on the right is the same as the left except only losses greater than two percent are plotted. This shows the a fitted distribution is a good fit under extreme losses, but perhaps not under small losses around zero to two percent.

In addition to the probability plots, a Kolmogrov-Smirnov test was run. The results are shown in Table 2. The null hypothesis is that the block maxima log-daily return data come from a Fréchet distribution. The alternative is that the sample comes from a different distribution. The test shows a very small p-value indicating that the null hypothesis should be rejected. This adds some evidence to suggest the Fréchet distribution may not be an acceptable fit for the data.

What this modeling and analysis shows is that the model may be a good fit for extreme values where the losses in one day are greater than approximately two percent. The analysis also reveals the model is not a good fit for smaller losses. Although this model is not perfect, it does give

Test Statistic	P value	Alternative Hypothesis
0.5738	3.798e-09***	two-sided

Table 2: Results for the Kolmogrov-Smirnov two sample test. The first sample (empirical data) was compared to the fitted Fréchet distribution.

a hint of what what may have been a better approach. It appears that perhaps a POT approach would have been better route. Only observations greater than two or three percent could have been chosen and fitted to the generalized Pareto family. This could be something done in the future with more time.

3.3 Applications

Although this model cannot be accepted as accurate in light of the evidence presented above, a simple application will be demonstrated to put some of the data in context and show an application of extreme value distributions. For example what is the maximum loss that may be observed during a period of 100 years? Gilli and Këllezi provide such a method they call return levels in their study that flows directly from using the GEV distribution [2]. From MLE using $\hat{u}, \hat{\sigma}$, and $\hat{\xi}$, Gilli and Këllezi define the computation for return levels, $\hat{R}_k$, as

$$\hat{R}_k = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left(1 - \left(-\log(1-\frac{1}{k})\right)^{-\hat{\xi}}\right) & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left(-\log(1-\frac{1}{k})\right) & \hat{\xi} = 0. \end{cases} \tag{4}$$

Using the estimates from Table 1, it follows that from top line in Equation (4) that R_{100} was estimated to be 0.346. This can be interpreted to mean that a once in a 100 year loss will be around 34.6%. This is assuming that the model is accurate which was questioned to some extent in the previous section. To put this estimate into context, the largest ever one day loss on the S&P 500 was on October 19, 1987 where the S&P 500 was down 22.9%.

Although the extreme value distributions can't be applied where central tendency of data is needed, it does to well to estimate tail risk. Value at risk (VAR), a risk management tool used by banks, would be an application where this could be used and a possible future study could go into this more in detail using what was learned here (i.e. using the POT approach and then calculating VAR).

4 Conclusion

In this study, data from the S&P 500 were fitted to an extreme value distribution using a block maxima approach. The data were fitted to the generalized extreme value distribution using MLE. The fitting process and extreme value theory indicated that a Fréchet distribution would be a good model for the data. The modeling and analysis showed mixed results when checking the fit of the data. The results show that the data perform poorly when the data were not very extreme; from zero to two percent losses. However, the model performs well under more extreme losses such as two or more percent. A POT approach fitted to a generalized Pareto distribution would have likely have been a better fit since losses only greater two percent could have been considered for study. In conclusion, it appears the block maxima approach may not be the best method for this data.

References

[1] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

[2] Manfred Gilli et al. An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2):207–228, 2006.

[3] Robert L Hagerman. More evidence on the distribution of security returns. *The Journal of Finance*, 33(4):1213–1221, 1978.

[4] Paul A Samuelson. Proof that properly anticipated prices fluctuate randomly. In *The world scientific handbook of futures markets*, pages 25–38. World Scientific, 2016.

[5] Alec Stephenson and Chris Ferro. *evd: Functions for Extreme Value Distributions*, 2018. R package version 2.3.3.