Applications of Fat Tail Models in Financial Markets

Sudhalahari Bommareddy
Saad El Beleidy
Sujitreddy Narapareddy
Numan Yoner

Sponsor: Dr. Kuo Chu Chang

Executive Summary

Normal distributions are commonly used in many modeling techniques to describe reality yet in many cases, reality is closer to a fat-tailed distribution than a normal one. This critical assumption of normality results in potentially inaccurate models, especially when modeling risk and options prices. The aim of this project is to develop a fat-tailed distribution model that fits the S&P500's daily percent returns to be applied in standard financial calculations.

The project uses whole data models of t-distribution and mixed normal distributions as well as extreme scenario models of Generalized Extreme Value distribution (GEV) and Generalized Pareto Distribution (GPD) to calculate Value at Risk (VaR) and options prices. The results show that the models chosen exhibit fatter tails than the data and by a larger difference than that between the data and the normal distribution. Of all the models, the closest to the data is the mixed normal model. This leads to the recommendation of the project being to use the mixed normal model as the distribution in estimating VaR and options prices so that investment decisions include the expectation of more extreme scenarios. For VaR, this can also be achieved by using the normal model with a base value for VaR or by using a multiplier.

Acknowledgements

We would like to acknowledge, with thanks, the assistance and advice of our sponsor Dr. Kuo Chu Chang and our academic advisor Dr. Kathryn Blackmond Laskey.

Table of Contents

Context

Financial engineering is a multidisciplinary field that involves the use of mathematical techniques and analyses to solve financial problems. It uses a variety of tools such as statistics, concepts of economics, computer science, and applied mathematics to help address the financial market, present and future. Financial engineering is primarily used as an analysis technique for financial corporations such as corporate banks, hedge funds, and investment banks.

Risk is an important concept in Financial Engineering. It can be defined as the amount of exposure to loss. Investments are generally made in anticipation of a positive return but the value of this return is uncertain, unpredictable and potentially unfavorable. The practice of financial engineering uses tools to determine a set of decisions that minimize future risk.

Problem: Normal Assumption

The normal distribution is a probability distribution in which all of the corresponding values are plotted in a symmetrical fashion, and most of the results are grouped tightly around the mean value of the distribution. The values are equally likely to be plotted on either side of the mean. Values are typically grouped close to the mean and then tail off symmetrically away from the mean with the spread defined by standard deviation.

The normal distribution is a commonly occurring statistical pattern which has been observed in distinct phenomena such as product manufacturing, and biological variables including height, blood pressure etc. The distribution enables us to specify that majority of the possible values of any unknown distribution are likely to be limited between two real limits.

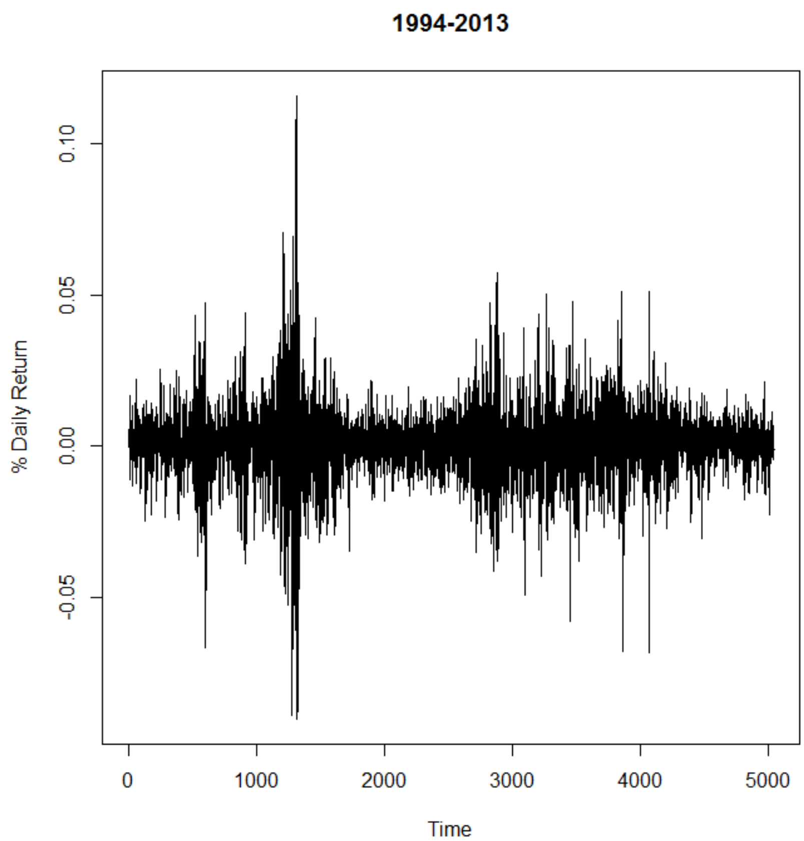
Normality is assumed in many problems initially as the distribution is understood comprehensively through many prior studies and the foundational central limit theorem infers that for a sufficiently large data-set the mean is distributed normally. This enables us to draw effective conclusions from a data-set and explain the observations using well-defined principles.

S&P500’s Normality

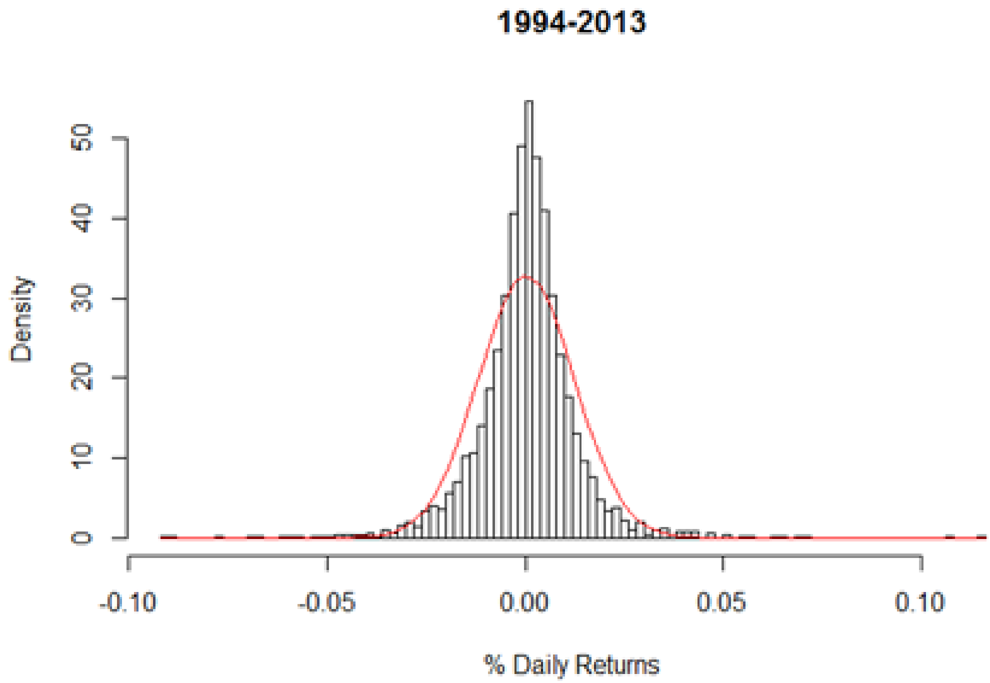
The normal distribution assumption is so prevalent in many domains including financial asset pricing models of S&P 500. The normal distribution is often utilized because of its universal occurrence in many phenomena as well as its relative simplicity. A key feature of its simplicity is the three-sigma rule specifying that virtually all of the values are constrained within three standard deviations of the distribution’s mean.

S&P 500 data is assumed to be normal because analysis and plotting of the sufficiently large data set actually shows that most of the price movements are described adequately by the normal distribution (bell-shaped curve). However, the key problem with applying this assumption with S&P 500 data is that the exceptional price movements in the distribution are not as extremely infrequent as assumed with a data-set explained by normal distribution.

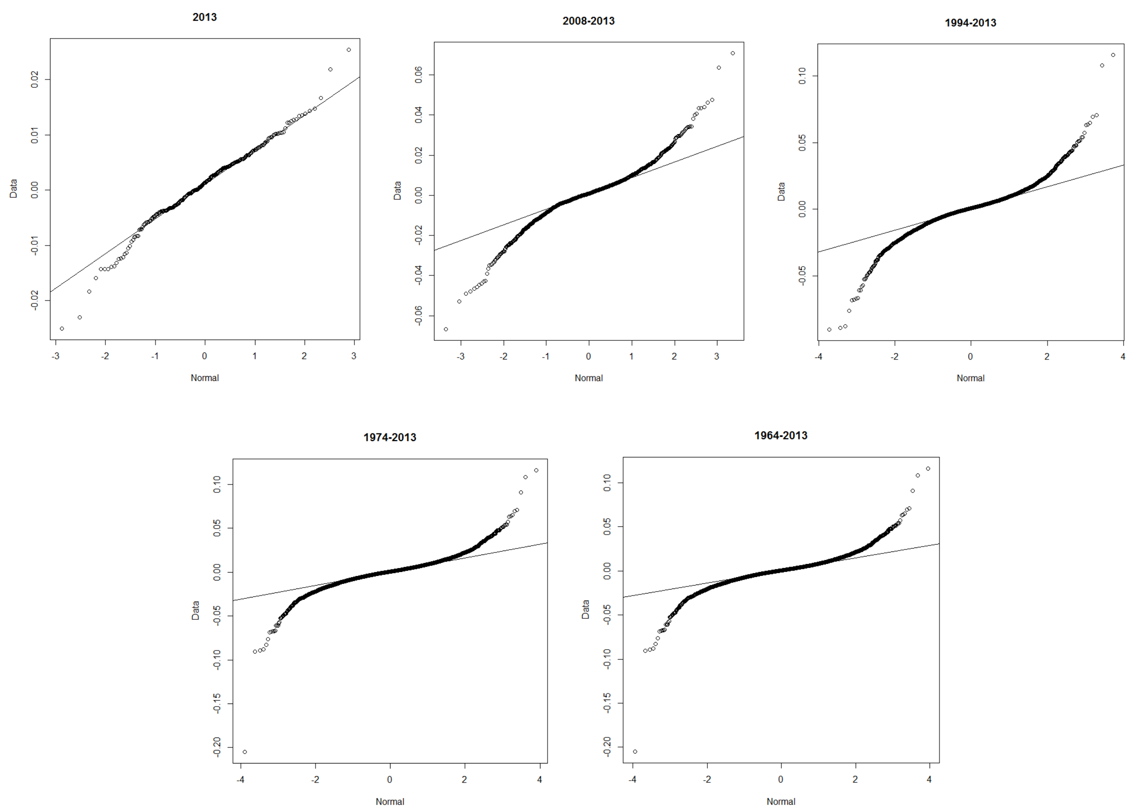
The following graph represents daily percent returns of the S&P 500 Index for the past 20 years. The graph shows how the daily percent returns (y axis) change over time (x axis) in days since January 1st 1994. We can clearly see that the percentage change is not always the same , there are also many extreme results which the normal distribution would not take into consideration.



The following histogram represents the daily returns of S&P 500 Index with a normal fit shown in red. We can see that normal distribution does not account for the fat tailed nature of the data and therefore, is not a good representation of the data for the purposes of this project..



The following are the Q-Q plots of the 1, 5, 20, 40 and 50 years data sets of S&P 500 Index daily returns. From the plots we can notice that the 1 year is very much close to the normal but as duration of the is increasing the data is going away from the normal, this is due to the occurrence of more and more extreme outcomes over the period of time.



Objective

The quest for reliable financial modeling techniques has increased in response to the highly volatile and seemingly unpredictable nature of the financial markets. Large losses and returns occur more frequently than predicted under the assumption of normality. This was seen in the figures above describing S&P500 daily percent returns data which has lead to the following objective:
The aim of this project is to develop a fat-tailed distribution model that fits the S&P500's daily percent returns to be applied in standard financial calculations.

Financial Applications

There are several standard financial calculations being done by all stakeholders interested in financial modeling. Of those calculations, we chose to focus on two of the most popular; Value at Risk (VaR) and Options pricing.

Value at Risk (VaR)

Value at Risk has been established as a standard tool among financial institutions to depict the downside risk of a market portfolio. It measures the maximum loss of the portfolio value that will occur over some period at a specific confidence level due to risky market factors. Value at risk is used by risk managers in order to measure and control the level of risk which the firm undertakes. VaR is measured in two variables; the confidence level in percent, and the time frame in days. Technical details about VaR will be covered in the model applications section below.

Importance of VaR

VaR is an important metric used by the firms to determine the capital amount that serves as a reserve for unexpected losses.The most important application of VaR is in Basel III. Basel III is a comprehensive set of reform measures, developed by the Basel Committee on Banking Supervision, to strengthen the regulation, supervision and risk management of the banking sector. These measures aim to:

- improve the banking sector's ability to absorb shocks arising from financial and economic stress, whatever the source
- improve risk management and governance
- strengthen banks' transparency and disclosures.

The VaR-based market risk capital requirement for a portfolio under the Basel III requires daily calculation of a basic VaR metric for the portfolio. This metric is usually calculated using the normal distribution but is multiplied by a factor to take into consideration the fatter-tailed nature of reality. More information about Basel III can be found at http://www.bis.org/bcbs/basel3.htm.

Options Pricing

European Options provide the holder with the right to buy or sell a specified quantity of an underlying asset at a fixed price (called a strike price or an exercise price) at or before the expiration date of the options. Since it is a right and not an obligation, the holder can choose not to exercise the right and allow the option to expire. There are two types of options - call options (right to buy) and put options (right to sell). Throughout this paper, the term option refers to a European Option.

A call option gives the buyer of the option the right to buy the underlying asset at a fixed price (strike price or K) at any time prior to the expiration date of the option. The buyer pays a price for this right.

- By expiration, if the value of the underlying asset (S) > Strike Price (K)
  Buyer makes the difference: S - K when they exercise the option
- If the value of the underlying asset (S) < Strike Price (K)
  Buyer does not exercise.

More generally, the value of a call increases as the value of the underlying asset increases and vice versa.

A put option gives the buyer of the option the right to sell the underlying asset at a fixed price at any time prior to the expiration date of the option. The buyer pays a price for this right.

- By expiration, if the value of the underlying asset (S) < Strike Price (K)
  Buyer makes the difference: K - S when they exercise the option
- If the value of the underlying asset (S) > Strike Price (K)
  Buyer does not exercise

More generally, the value of a put decreases as the value of the underlying asset increases and vice versa.

Black Scholes Model

As expected, options pricing is a very important field and so many financial models attempt to accurately value them. Of these, Black-Scholes model is the most popular. It takes into account the option of investing in an asset earning the risk-free interest rate. It also acknowledges that the option price is purely a function of the volatility of the stock's price (the higher the volatility the higher the premium on the option). Volatility of a stock is the standard deviation of the daily percent returns. Black-Scholes treats a call option as a forward contract to deliver stock at a contractual price, which is, of course, the strike price.

The value of a call option in the Black-Scholes model can be written as a function of the following variables:

$$S = \text{Current value of the underlying asset}$$
$$K = \text{Strike price of the option}$$
$$t = \text{Life to expiration of the option}$$
$$r = \text{Riskless interest rate corresponding to the life of the option}$$
$$\sigma^2 = \text{Variance of the underlying asset}$$
$$d_1 = \frac{ln(\frac{S}{K})+(r+\sigma^2)t}{\sigma\sqrt{t}}, \quad d_2 = d_1 - \sigma\sqrt{t}$$

$$\text{The value of a call option is}:$$
$$C = SN(d_1) - Ke^{-rt}N(d_2)$$

Implied Volatility

The implied volatility of an option contract is the value of the volatility that, when input in an option pricing model, will return a theoretical value equal to the current market price of the option. In principle, the implied volatility can be inferred from computed options prices by inverting the Black-Scholes formula.

VIX is a trademarked ticker symbol for the Chicago Board Options Exchange Market Volatility Index, a popular measure of the implied volatility of S&P 500 index options. It shows the market's expectation of 30-day volatility and is a widely used measure of market risk and is often referred to as the "investor fear gauge”.

Scope

The scope of the project is developing models to fit daily returns of S&P 500 for each of 1 year, 5 years, 20 years, 40 years and 50 years. The developed models will be compared against each other as well as the normal distribution. All developed models will then be used to calculate Value at Risk and options prices.

The final selected model must meet the following requirements:

1. The model must obtain a mean value within a 99% confidence interval of the respective dataset's mean
2. The model must obtain a standard deviation value within a 99% confidence interval of the respective dataset's standard deviation
3. The model must obtain a kurtosis value within a 95% confidence interval of the respective dataset's kurtosis
4. A Kolmogorov-Smirnov (K-S) statistical test shall not reject the hypothesis that the data arise from the fitted model. (see the model verification section for more details)

Models will then be used to calculate VaR as well as options prices to measure their effectiveness.

Modeling Approach

Standard statistical analysis is conducted to better understand the data and develop parameters that can be used in generalizing the data's distribution. After that, the parameters and further analysis of the data is used in two approaches in finding a generalized distribution for the data. The first approach involves inspecting the entire data set and learning a general fit for the data. The second approach involves inspecting the extreme values of the data to develop a model that describes them alone.

The major assumption being made in this project is that the daily percent returns are independent of each other. That is to say that if the S&P500 goes down 2% today, that does not affect in any way how the S&P500 will behave tomorrow.

Whole Data Modeling

After obtaining the key parameters of the data, we generate a distribution to fit the data. We will be looking at two types of distributions; the t-distribution and a mixture of normal distributions. These distributions are commonly used to describe fatter than normal tailed data sets.

t Distribution

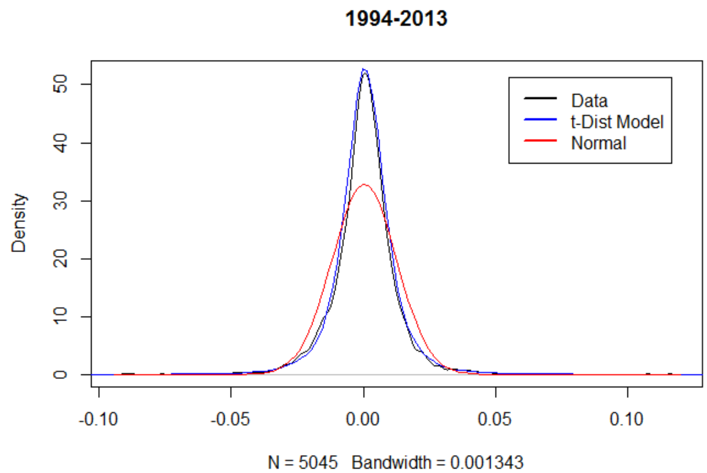
The t distribution, also commonly known as the Student's t- distribution, has a density function of:

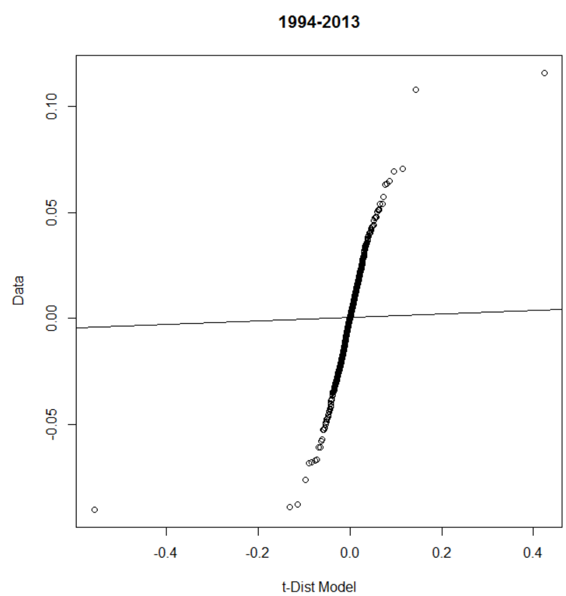
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

Where v is the number of degrees of freedom.

Since the t-distribution has fatter tails than the normal distribution it should be a more realistic approximation of the distribution for the data set. The std() and stdfit() functions found in the fGarch package in R are used to develop this model. Further information about this package can be found at http://cran.r-project.org/web/packages/fGarch/index.html

The following are sample graphs of the fit of the t-distribution model generated for the 20 year data set. First we have a graph of the density of the model compared to the data and the normal fit of the data followed by a QQ plot of the distribution and the data.





We can see that even though the t distribution seems to have a very good fit around the mean of the data, there are much more extreme scenarios found in the t distribution than the data.

Mixed Normal Distribution

The second distribution used to fit the data will be obtained by mixing two normal distributions of varying standard deviations. This aggregation of normal distributions with varying standard deviations creates fatter tails than a normal distribution and can hopefully more accurately represent the data's distribution.

The procedure used to generate a mixed normal distribution is as follows:

1. Generate a Boolean random variable (Y) whose values are 1 with probability p and 0 with probability 1-p
2. Generate two Gaussian random variable X_1 and X_2 to be mixed with standard deviations α and β such that the variance of the mixed model is equal to 1. This method results in only one of the standard deviations being a parameter to the distribution while the other one is calculated as follows

$$\beta = \sqrt{\frac{1-p\alpha^2}{1-p}}$$

3. Generate a standard normal random variable Z
4. Loop through all values of the generated Boolean random variable (Y) and for each value
    a. If it is equal to one, equate the output model value to α times the value from the standard normal random variable Z,
    b. otherwise equate the output model value to β times the value from the standard normal random variable Z

In mathematical terms:

$$Y = Bool(p)$$
$$Z = Norm(0,1)$$
$$\beta = \sqrt{\frac{1-p\alpha^2}{1-p}}$$

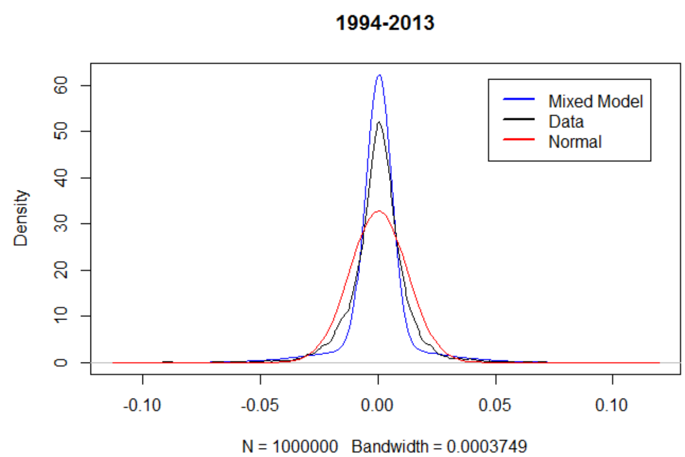
$$For\ each\ Y$$
$$if\ Y = 1,\ X = \alpha Z$$
$$if\ Y = 0,\ X = \beta Z$$

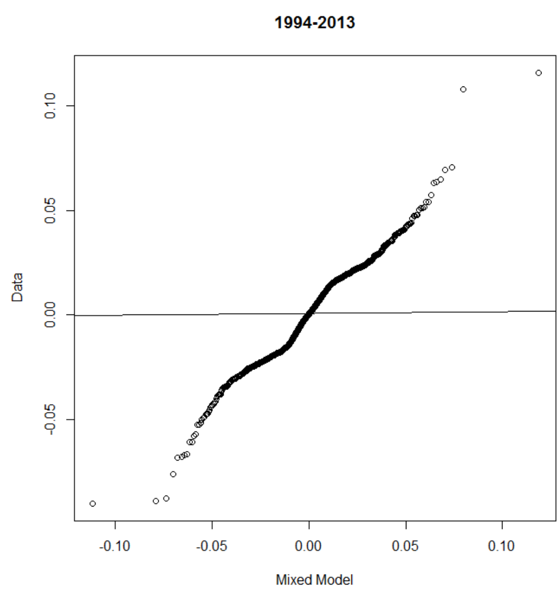
This generated distribution has a mean of 0, standard deviation of 1 and a kurtosis value equal to

$$\gamma = 3(p\alpha^4 + (1-p)\beta^4) - 3$$

In order to find a mixed normal distribution that fits the dataset accurately, MS Excel Solver is used to obtain values for p and α based on the kurtosis value of the data. From there, the model is generated and plotted on a graph alongside the data for visual comparison.

The following are sample graphs of the fit of the mixed normal model generated for the 20 year data set. First we have a graph of the density of the model compared to the data and the normal fit of the data followed by a QQ plot of the distribution and the data.





We can see that the mixed model doesn't do as well as the t-distribution for the values close to the mean but it does better in terms of extreme scenarios.

Extreme Scenario Modeling

As briefly mentioned above, when modeling the extremes (max/min) of random variable, Extreme Value Theory (EVT) plays the same fundamental role as the Central Limit Theorem (CLT) when modeling aggregation of random variables. In both cases, theory suggests what the limiting distributions are.

EVT has two significant results. First, the asymptotic series of maxima (minima) is modeled and under certain conditions the distribution of standardized maxima (minima) is shown to converge to a Generalized Extreme Value (GEV) distribution.

The second significant result concerns the distribution of excess values over a given threshold, where one is interested in modeling the behavior of excess loss once a high threshold is reached. This type of distribution is called the Generalized Pareto Distribution (GPD).

For the purpose of the project, values greater than 2% in magnitude were considered extreme scenarios. These are days in which the daily percent gain or loss was greater than 2%.

Generalized Extreme Value (GEV) Distribution

The Generalized Extreme Value distribution can be developed as follows:
Let X_n be a series of iid random variables and M_n be the maxima of values of X within certain blocks of size m such that $M_n = Max\,(X_1, X_2, \ldots X_n)$. Then M_n follows the GEV distribution and H(x):

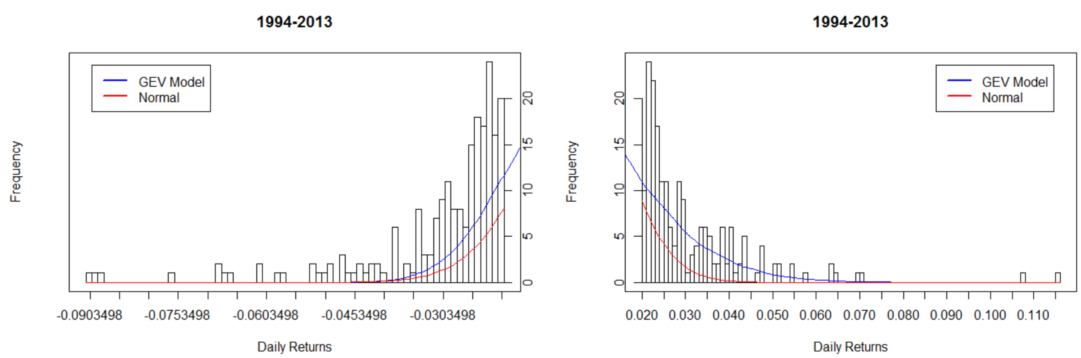
$$H_{(\xi,\mu,\sigma)} = \begin{cases} \exp\left(-[1+\xi(x-\mu)/\sigma]^{-1/\xi}\right) & if\ \xi \neq 0 \\ \exp(-e^{-(x-\mu)/\sigma}) & \xi = 0 \end{cases}$$

$$while \qquad 1 + \zeta(x-\mu)/\sigma > 0$$

The parameters μ, σ, ξ correspond, respectively, to location, scale and shape (tail index) parameters.

To find and generate accurate models for the data sets, the gev() and gevFit() functions of the fExtremes R package were used. Further information about these functions can be found at http://cran.r-project.org/web/packages/fExtremes/index.html

Sample outputs for the GEV models for the 20 year data set are as follows:



We can see that GEV has fatter tails than the normal and seems closer to the data.

Generalized Pareto Distribution (GPD)

The Generalized Pareto Distribution can be developed as follows:

Let X be a random variable with distribution F and a threshold given x_f, for U fixes < x_f, F_u is the distribution of excesses of X over the threshold U.

$$F_u(x) = P(X - u <= xI\ X > u),\ x >= 0$$

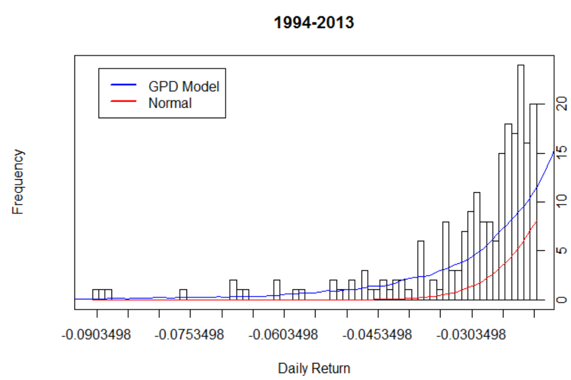
Once the threshold, u, is determined by an estimation procedure, the conditional distribution of F is approximated by a GPD.

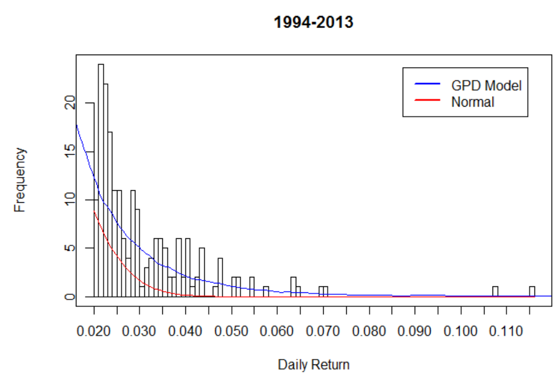
$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-1/\xi} & \text{if} \quad \xi \neq 0 \\ 1 - e^{-y/\sigma} & \text{if} \quad \xi = 0 \end{cases}$$

The parameters σ, ξ correspond, respectively, to scale and shape (tail index) parameters. The interesting property of tail index parameter is that there is a relationship between this parameter and the t distribution's degrees of freedom.

To find and generate accurate models for the data sets, the gpd() and gpdFit() functions of the fExtremes R package were used. Further information about these functions can be found at http://cran.r-project.org/web/packages/fExtremes/index.html

Sample outputs for the GPD model for the 20 year data set are as follows:





We can see that GPD has fatter tails than normal and seems to better fit the data as well. It also has fatter tails than GEV. GPD's tails also maintain their thickness further away from the mean.

Model Verification

The requirements developed for the models during the proposal stage had focused on the whole data models. The extreme scenario models will be considered in their VaR results.

For the verification of the developed whole-data models with our requirements, we get the following results:

1. The model must obtain a mean value within a 99% confidence interval of the respective dataset's mean

Data Set	Model	Mean	Mean Diff	Mean Pass?
1 year	Mixed Normal	0.001050733	0.002809	Pass
5 years	Mixed Normal	0.000674637	0.00609	Pass
20 years	Mixed Normal	0.000340924	0.017166	Fail
40 years	Mixed Normal	0.000340405	0.030595	Fail
50 years	Mixed Normal	0.000313388	0.019198	Fail
1 year	T - Dist	0.001275583	0.210583	Fail
5 years	T - Dist	0.001015543	0.514486	Fail
20 years	T - Dist	0.000680093	0.96061	Fail
40 years	T - Dist	0.00043666	0.24352	Fail
50 years	T - Dist	0.000405344	0.318256	Fail

The mixed model is the only model that passes this requirement and this only occurs for the 1 and 5 year data sets.

2. The model must obtain a standard deviation value within a 99% confidence interval of the respective dataset's standard deviation

Data Set	Model	Standard Dev.	SD Diff	SD Pass?
1 year	Mixed Normal	0.006962838	0.0014	Pass
5 years	Mixed Normal	0.01227532	0.0002	Pass
20 years	Mixed Normal	0.01216536	0.0011	Pass
40 years	Mixed Normal	0.01100889	0.0003	Pass
50 years	Mixed Normal	0.01023801	0.0057	Pass
1 year	T - Dist	0.006985151	0.0018	Pass
5 years	T - Dist	0.01404796	0.1446	Fail
20 years	T - Dist	0.01312552	0.0801	Fail
40 years	T - Dist	0.01137252	0.0333	Fail
50 years	T - Dist	0.01039655	0.0097	Pass

All mixed models pass this requirement as well as the t-models for 1 and 50 years.

3. The model must obtain a kurtosis value within a 95% confidence interval of the respective dataset's kurtosis

Data Set	Model	Kurtosis	Kurt. Diff	Kurt. Pass?
1 year	Mixed Normal	1.316811	0.011529	Pass
5 years	Mixed Normal	4.069714	0.001355	Pass
20 years	Mixed Normal	8.352593	0.00371	Pass
40 years	Mixed Normal	19.531	0.003376	Pass
50 years	Mixed Normal	20.91449	0.002748	Pass
1 year	T - Dist	2.224212	0.669617	Fail
5 years	T - Dist	46.59059	10.46364	Fail
20 years	T - Dist	77.11566	8.266795	Fail
40 years	T - Dist	181.9548	8.284757	Fail
50 years	T - Dist	0.010397	0.102399	Fail

All mixed models pass this requirement while all t models fail it.

4. A Kolmogorov-Smirnov (K-S) statistical test shall not reject the hypothesis that the data arise from the fitted model. (see the model verification section for more details)

Data Set	**Model**	**K-S**
1 year	Mixed Normal	D = 0.0521 p-value = 0.5006
5 years	Mixed Normal	D = 0.0272 p-value = 0.3076
20 years	Mixed Normal	D = 0.0554 p-value = 7.916e-14
40 years	Mixed Normal	D = 0.0994 p-value < 2.2e-16
50 years	Mixed Normal	D = 0.0972 p-value < 2.2e-16
1 year	T - Dist	D = 0.0088 p-value = 0.3473
5 years	T - Dist	D = 0.0212 p-value = 0.6325
20 years	T - Dist	D = 0.0181 p-value = 0.08711
40 years	T - Dist	D = 0.0089 p-value = 0.4554
50 years	T - Dist	D = 0.0088 p-value = 0.3473

In agreement with the sponsor, this requirement is removed to focus on results of the applications of the developed models.

Model Verification Conclusions

Based on the results of the model verification process we can see that the preliminary requirements may have been too conservative for the models. Results from the VaR and options pricing applications of the models will be used to come to general conclusions about applications of the chosen fat tailed models.

Model Applications

Two applications of the developed fat tail model will be pursued. Of the most common fields of modeling in financial markets, risk mitigation and options pricing are extremely interesting. For risk mitigation, the application chosen will be calculating value at risk (VaR) and for options pricing, a Monte Carlo simulation for options pricing will be developed.

Value at Risk (VaR)

Value at Risk, or VaR, is a very important metric for taking risk into consideration when making important financial decisions. It can be calculated as the x-value of the density of the N-day returns at which the area under the curve is equal to one hundred minus a set confidence level.

Calculation Method

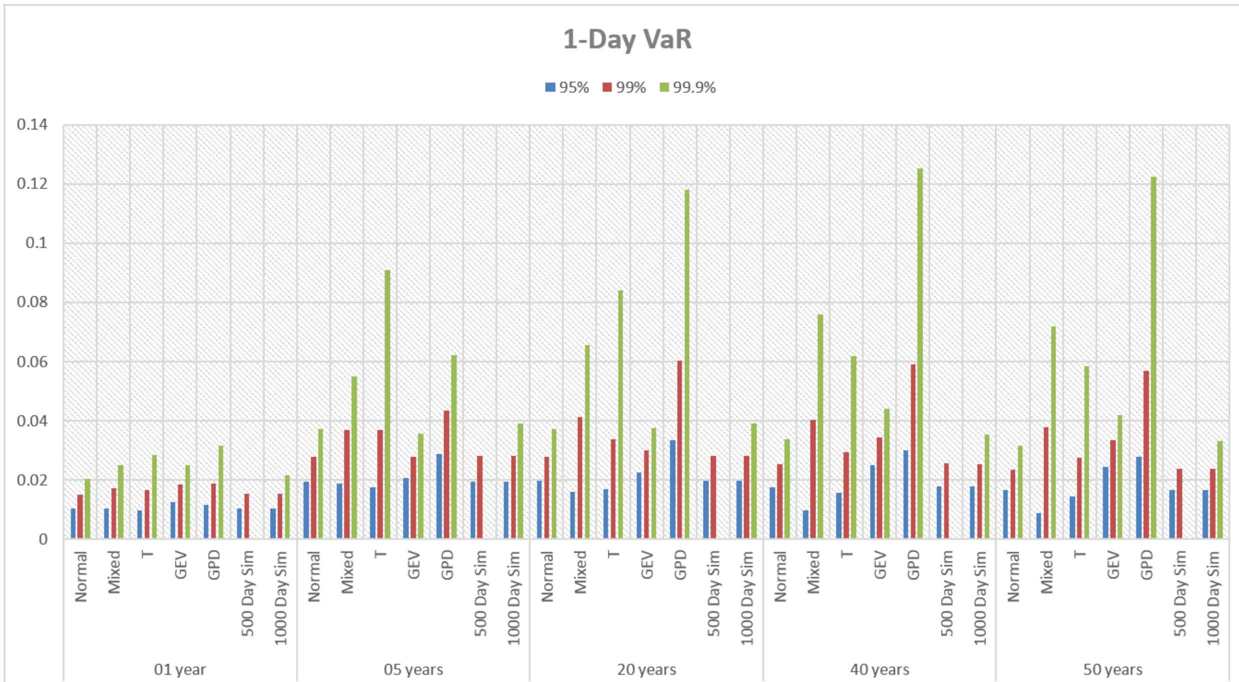
The method used to calculate this value for the generated models is quite simple. The model of 1 million points is provided and the code then sorts the entire data set in ascending order. Based on the confidence level, the number of points that are needed to generate one hundred minus the confidence level is calculated. Using that value, we find that point in the ordered data set to find the VaR. For example, for a 95% confidence level, 5% of 1,000,000 is the number of points. We then find the 50,000th point in the ordered data set and that gives us the VaR.
This method is applied to all models using daily % returns to calculate the 1-day 95%, 99% and 99.9% Values at Risk.

Benchmark Calculation

After the model values are calculated they must be compared to a benchmark value for verification of the results. The benchmark used is calculated by running a simulation of the data set. Using a function in R called sample(), which produces a random sample of n data points from a dataset, we generate 500 and 1000 days of daily % returns. From these values, we can calculate the VaR using the method described previously. We then repeat this run for 1,000,000 replications and take the average VaR from all the simulations. This provides two benchmark values for each data set.

Results

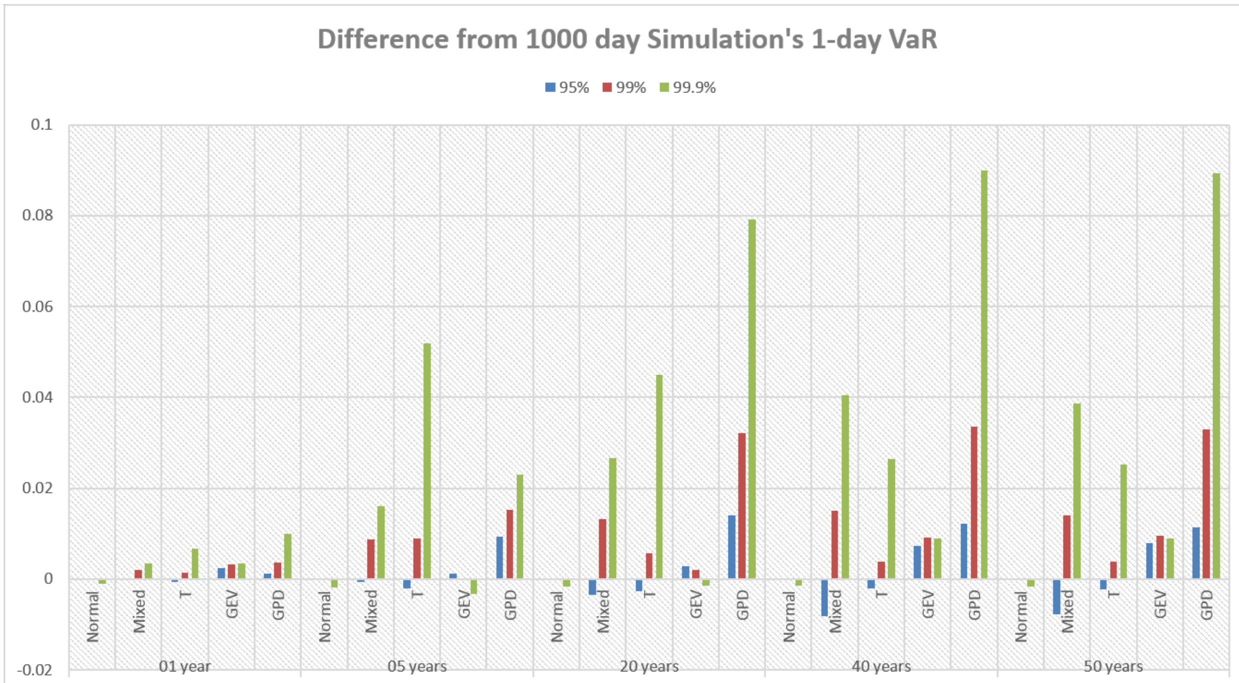
Below are the Values at Risk results shown on a chart The chart is organized by data set in order of smallest to largest followed by each model in the order of Normal, Mixed, T, GEV, GPD and the two benchmark outputs from the simulation. The bars are then displayed for each model, 95% then 99% and finally 99.9%. The Values at Risk are shown as a percentage of the assets' value.



The following are the important observations from the chart:

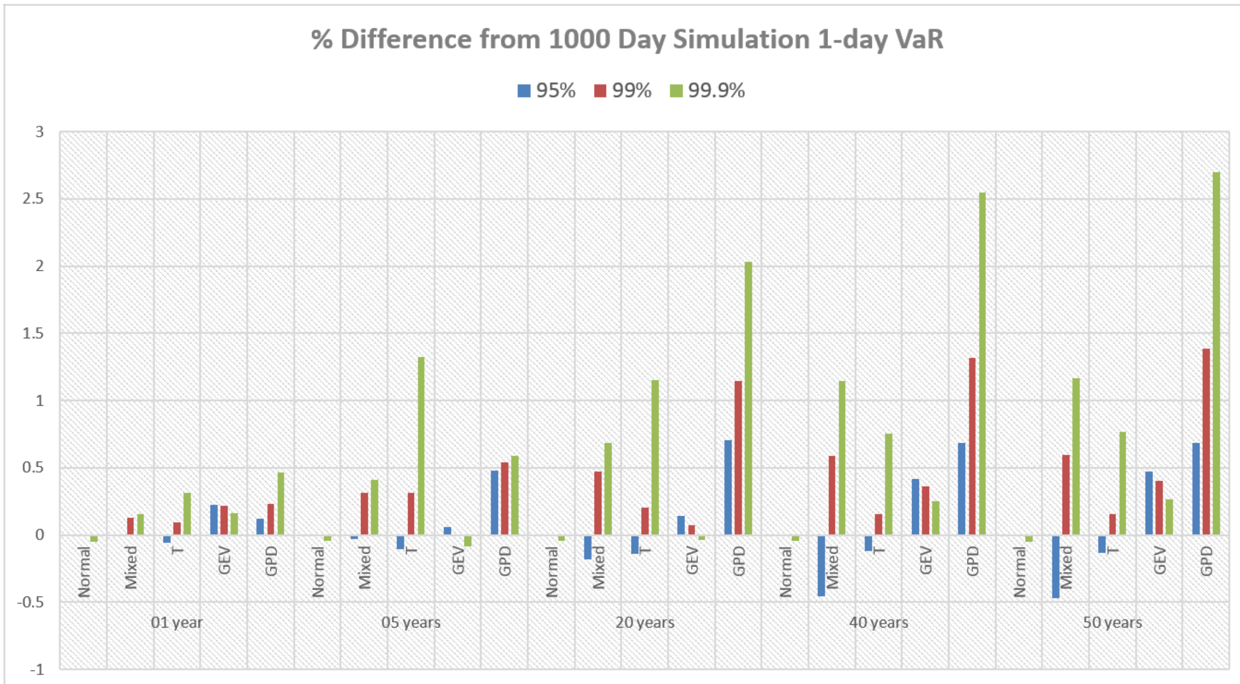
1. As the confidence level increases the Values at Risk increase.
2. As the data set increases the Values at risk also increase. This is due to more extreme results being included in the data set.
3. The extreme value models also show larger Values at Risk due to fatter tails in the models. However, with less data they are closer to the whole data models due to less availability of extreme results.
4. The simulation results appear to be closer to the whole data models, specifically the normal model, than the extreme scenario models.

To better compare the results to the simulation outputs, we analyze the difference between the 1000 day simulation and the models. First, we compare the differences and then the % difference as shown in the graphs below. The graphs are organized the same way with the data sets followed by the models (excluding the simulations) and the bars showing the same 95, 99 and 99.9% confidence levels.



We can see that the normal distribution is generally the closest to the simulation results but it shows results that are less in magnitude than the simulation results. The next best whole data model is the Mixed model as long as we don't take into consideration the 95% level. Since this confidence level is, in fact, rarely used in practice the Mixed model seems to be the best approximation after normal.

For more accurate comparison we look at the percent difference from the VaR simulation:



The normal model is the closest to the results however it provides values that are smaller than the actual results. The mixed model is the next best one and its results are larger than the simulation results. The mixed model should however, only be considered in cases where a 99% or larger VaR is considered. Since, in practice the 99% VaR is the most common, we conclude that for VaR, the mixed model is the next best model to the normal model.

Options Pricing

For analyzing the effects of fat-tailed models on options pricing we looked at options of S&P500 with varying expiration dates and strike prices. Namely options expiring 30,60 and 90 days with strike prices above and below the start price by 1, 2 and 5%. Option prices for each model as well as the historical data were calculated using a Monte Carlo simulation.

Simulation

The simulation takes the input as the model of S&P 500 daily % returns and then runs for the duration of the options, simulating the underlying asset. The difference from the strike price is then calculated to obtain the value of the options. This value is then discounted using the interest rate input to the simulation. This is then repeated for 100,000 replications and the average value for the call and put options are calculated to provide the final result for each model. The entire procedure follows the following equations:

$$V(n,m) = S * \prod_{i=1}^{n} (1 + d + m_i)$$

$$C(n,K,m) = \frac{\sum_{i=1}^{R} Max(V(n,m) - K, 0) * e^{\frac{-rn}{252}}}{R} \quad , \quad P(n,K,m) = \frac{\sum_{i=1}^{R} Max(V(n,m) - K, 0) * e^{\frac{-rn}{252}}}{R}$$

$V(n,m)$ = *value of the asset after n days using model m*

S = *Start price of the asset*

n = *number of days to model the asset*

d = *drift rate* (*used as* 0.1)

m_i = *the value generated from model m for day i*

$C(n,K,m)$ = *the price for a call options with expiration n trading days and strike price K using model m*

$P(n,K,m)$ = *the price for a put options with expiration n trading days and strike price K using model m*

r = *the risk − free interest rate* (*used as* 0.05)

R = *the number of replications for the simulation* (*used as* 100,000)

Benchmark Calculation

For comparing the simulation to actual results, historical options prices were not freely available so we used an approximation for them. Using Black-Scholes with the implied volatility value from VIX on December 31st 2013 we were able to create a metric for comparison to the simulation results.

The % difference from the calculated values of Black Scholes is used to measure each model's effectiveness. This is calculated as follows:

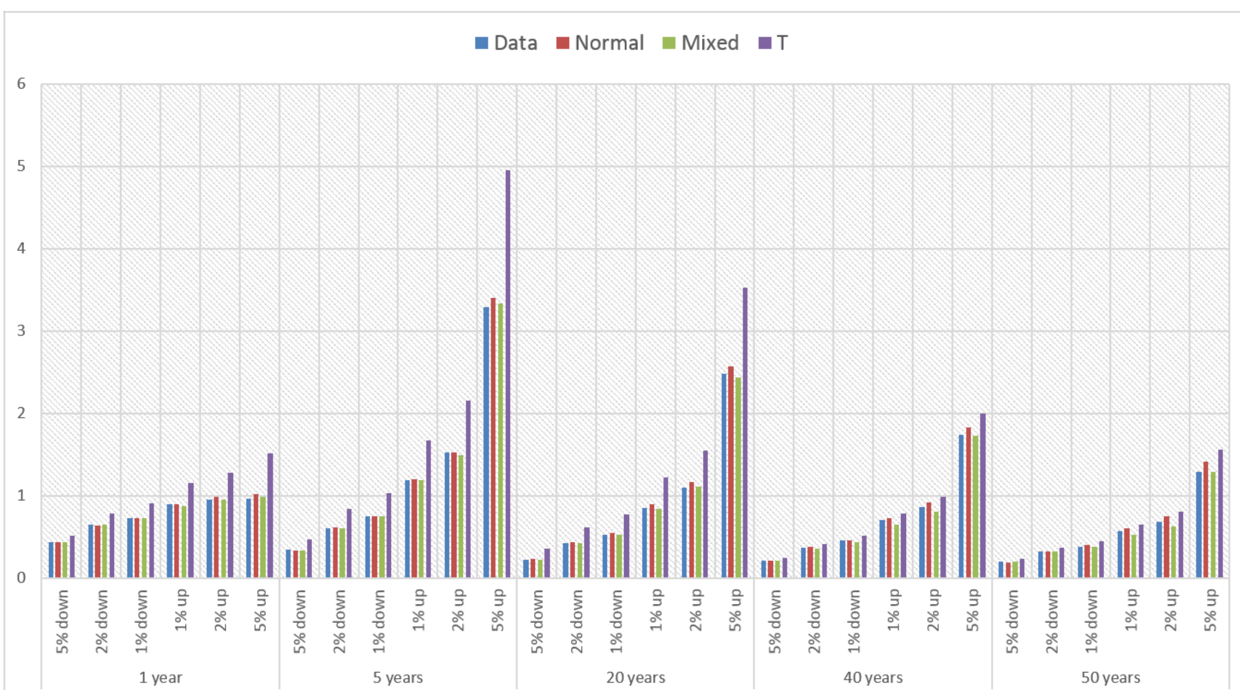
$$Benchmark = \frac{BlackScholes_{Price} - Simulation_{Price}}{BlackScholes_{Price}}$$

From now on, this value shall be referred to as the benchmark.

Results

Below are the results of the benchmark values for call options. The graph is grouped again by the dataset but now followed by the strike prices in ascending order and the bars show each model's w benchmark value.

ith the Data followed by Normal, Mixed and T. The extreme value models were not used for options pricing due to them not modeling the rest of the scenarios.



The following are the important observations from the chart:

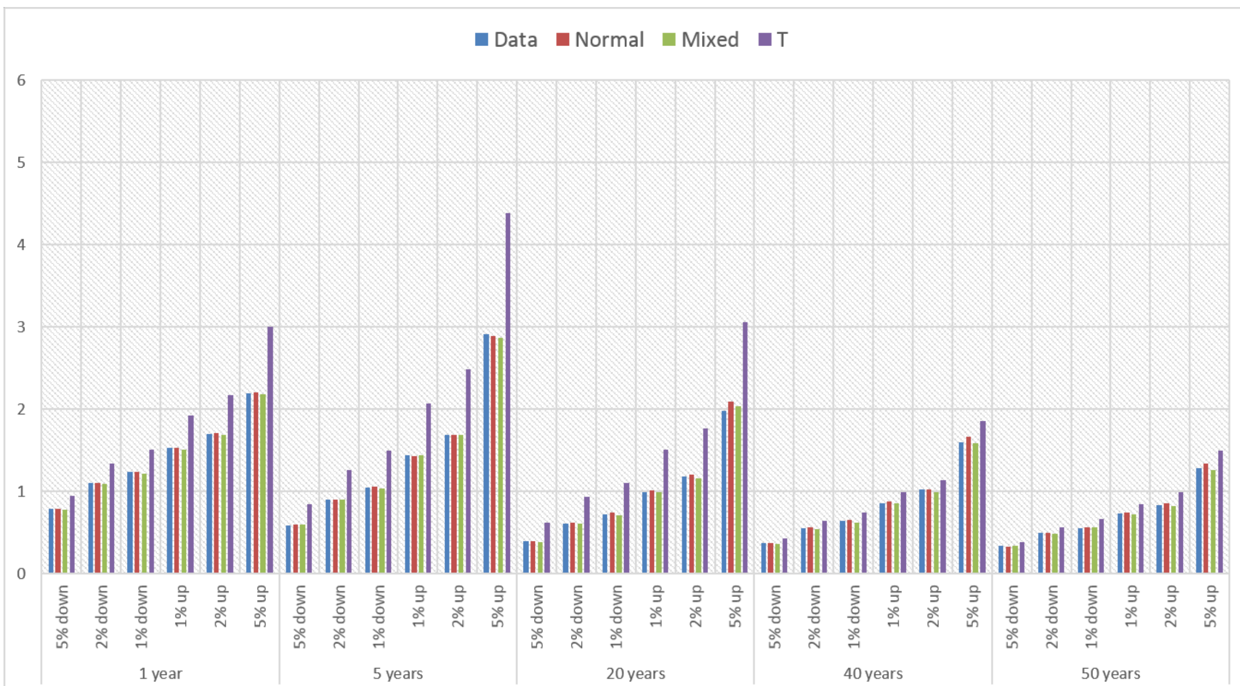
1. As the strike price increases the benchmark increases.

This is due to the benchmark being a percent value and the values of lower strike call options having a much higher magnitude. For example, a $5 increase on a $20 option will result in a much smaller benchmark than a $2 increase on a $1 option.

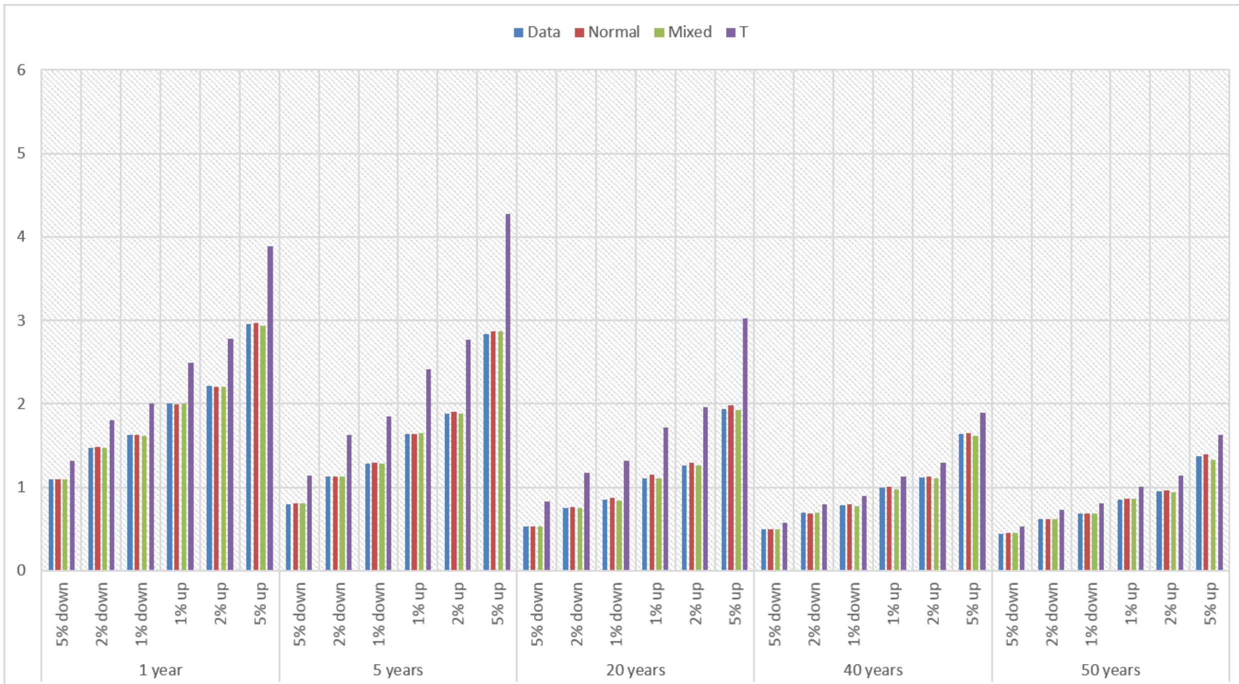
2. The benchmark values for the data, normal model and mixed model are closer to each other whereas the T model is significantly higher.

This is due to the extreme scenarios that are more frequent in the T model which lead to higher price estimates for the options.

3. Using larger datasets stabilizes the effect of larger differences of T from the other models due to more extreme scenarios being available in the data.
4. The mixed model is generally the closest to the data followed by the normal model.



As the duration of the option prices increases, the benchmark value increases as well. This is due to more extreme scenarios occurring and may also be due to the fact that VIX is an accurate estimate for 30 days but may prove less accurate for a longer duration.



The trend noticed in the option duration of 60 days is again noticed in 90 days. We can also see that the benchmark for the larger strike prices decreases. This is due to the price of the 5% greater strike price rising and the the benchmark (percent difference) decreasing since the difference from Black Scholes is smaller than the difference due to increase in duration.

Conclusions

Based on the model verification, Value at Risk calculations and Options pricing application, we've reached the following conclusions:

1. The data does indeed have fatter tails than the normal distribution, but the developed models have a fatter tail than the data and by a larger magnitude.

$$Normal_{tail} < Data_{tail} << MixedModel_{tail}$$

2. The normal model is the most accurate general model for financial applications but it is not good enough.

The normal model produces the closest VaR's but they are smaller than the actual data. A better way of using the normal model would be to introduce a multiplier of the normal model or a base value that is added to the normal for VaR calculations.
For options pricing, the normal and mixed models are quite close in comparison with no significant difference.

3. The mixed model is the next best model.

The mixed model can be a valid estimator for VaR as long as a 99% or higher VaR is considered. It will still provide values that are larger than accurate which means that the investor would be on the safe side.
For options pricing, the normal and mixed models are quite close in comparison with no significant difference.

Recommendation

The normal distribution is not a good enough estimator in financial applications that deal with risk and extreme scenarios such as Value at Risk calculation . We recommend the use of the normal and mixed normal models in making investment decisions. This allows the investor to have a lower and upper bound of risk to their investment:

- The normal model underestimates the actual risk
- The mixed model overestimates the actual risk

Future Work

For next steps or future plans of the project, we recommend investigating the best approach of calculating VaR from the recommended methods as well as testing other scenarios of options pricing to validate our expectation that the mixed model would be a more reliable model for calculating options prices.

Another approach going forward with this project is to model the standard deviation of the daily % returns as opposed to the daily % returns themselves. This could investigate the dependence of daily returns on each other and relax that assumption that is made in this project.

References

Value At Risk: A Quantile-Based Distribution Approach For Incorporating Skewness And Fat-Tailedness, Doowoo Nam
Alternative statistical distributions for estimating value-at-risk: theory and evidence, Cheng-Few Lee and Jung-Bin Su

Implementable tail risk management: An empirical analysis of CVaR-optimized carry trade portfolios, Hakan Kaya, Wai Lee, Bobby Pornrojnangkool

An Empirical Examination of Extreme Value Theory Methods in VaR Estimation, Zhenhong Faa

Capturing fat tails, Zari Rachev,Boryana Racheva-Iotova and Stoyan Stoyanov

Estimating Flexible, Fat-Tailed Asset Return Distributions, Craig Friedman, Yangyong Zhang, and Jinggang Huang

Estimation of portfolio return and value at risk using a class of gaussian mixture distributions, Kangrong Tan, Meifen Chu
A Comparative Software Review for Extreme Value Analysis, Eric Cilleland, Mathieu Ribatet, Alec G Stephenson

VaR-x: Fat tails in financial risk management, Ronals Huisman, Kees G Koedijk and Rachel A J Pownall

Fat-tailed models for risk estimation, Stoyan V. Stoyanov, Svetlozar T. Rachev, Boryana Racheva-Iotova, Frank J. Fabozzi

Fat tails via utility based entropy, Craig Freedman, Yangyong Zhang, Wenba Cao

An Application of Extreme Value Theory for Measuring Financial Risk, Manfred Gilli and Evis KÄellezi

Option Pricing and Implied Tail Indices under the Generalised Extreme Value (GEV) Distribution, Amadeo Alentorn

Intrinsic bubbles and fat tails in stock prices: a note, Prasad V. Bidarkota and Brice V. Dupoyet

Risk Forecasting with GARCH, Skewed t Distributions, and Multiple Timescales, Alec N. Kercheval and Yang Liu

Portfolio optimization with serially correlated, skewed and fat tailed index returns, M. Glawischnig and I. Seidl

Return Distributions and Applications, Young Do Kim

Computational aspects of risk estimation in volatile markets: A survey, Stoyan V. Stoyanov, Svetlozar T. Rachev, Frank J. Fabozzi

Backtesting Value-at-Risk based on Tail Losses, Woon K. Wong

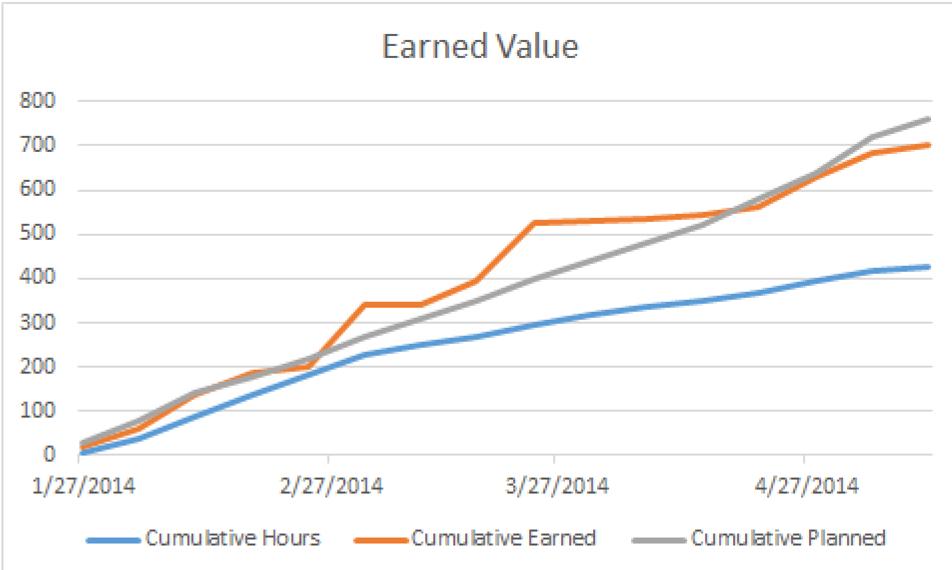
Statistical modeling of finance data and extensions of options pricing framework, Xin Zhong
Steps in Applying Extreme Value Theory to Finance: A Review, Younes Bensalah

The Tail Fatness and Value-at-Risk Analysis of TAIFEX and SGX-DT Taiwan Stock Index Futures, Yu Chuan Huang, Chu-Hsiung Lin, Chang-Cheng Chang Chien and Bor-Jing Lin

The tale of the tail: extreme-value patterns of financial returns, Angelo Corelli

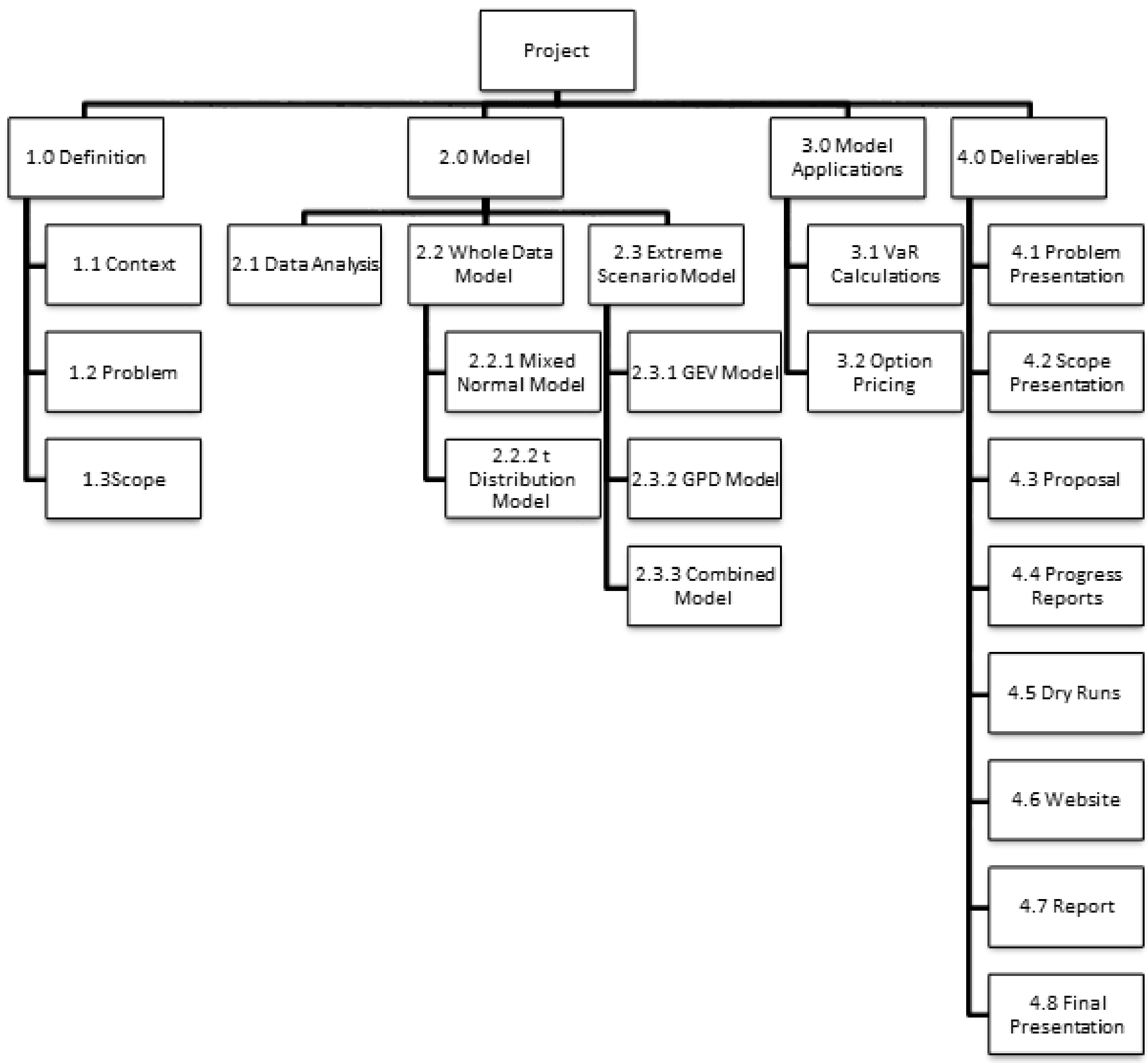
Appendix - EVM and Project Management

The earned value chart for this project is as follows:



The project required a lot of effort up front regarding researching the topic and better understanding the context. After understanding the context we began coding the analyses and the model which was somewhat time consuming since we attempted to manually code all necessary functionality or rely on built-in functions in R. After coding extensively in R, we found packages that could assist in the analysis of the model. This required some additional coding but resulted in the speeding up of our progress afterwards. We also spent more time documenting our results than expected as well as less time coding than expected which resulted in somewhat of a balance. The final earned value is less than planned due to not pursuing the initially proposed combined model approach and postponing it to future work.

The following is our WBS followed by our Gantt Chart with the critical path marked in red:



	WBS		Name	Duration	Start	Finish	Predecessors
1	**1**		**⊟Definition**	**10d**	**01/22/2014**	**02/04/2014**	
2	1.1		Context	1w	01/22/2014	01/28/2014	
3	1.2		Problem	1w	01/22/2014	01/28/2014	
4	1.3		Scope	1w	01/29/2014	02/04/2014	3
5	**2**		**⊟Model**	**33d**	**02/03/2014**	**03/19/2014**	**4SS 3d**
6	2.1		Data Analysis	1w	02/03/2014	02/07/2014	
7	**2.2**		**⊟Whole Data Model**	**10d**	**02/06/2014**	**02/19/2014**	**6SS 3d**
8	2.2.1		Mixed Normal Model	1w	02/06/2014	02/12/2014	
9	2.2.2		t Distribution Model	1w	02/13/2014	02/19/2014	
10	**2.3**		**⊟Extreme Scenario Model**	**15d**	**02/20/2014**	**03/12/2014**	**6SS 3d**
11	2.3.1		GEV Model	1w	02/20/2014	02/26/2014	
12	2.3.2		GPD Model	1w	02/27/2014	03/05/2014	
13	2.3.3		Combined EVT, Whole Data Model	1w	03/06/2014	03/12/2014	
14	2.4		Model Verification	1w	03/13/2014	03/19/2014	
15	**3**		**⊟Model Applications**	**20d**	**03/20/2014**	**04/16/2014**	**5SS 4w**
16	3.1		VaR Calculations	2w	03/20/2014	04/02/2014	
17	3.2		Option Pricing Calculations	2w	04/03/2014	04/16/2014	
18	**4**		**⊟Deliverables**	**75d?**	**01/27/2014**	**05/09/2014**	
19	4.1		Problem Presentation	1d?	01/27/2014	01/27/2014	
20	4.2		Scope Presentation	1d?	02/04/2014	02/04/2014	
21	4.3		Proposal	1d?	02/11/2014	02/11/2014	
22	**4.4**		**⊟Progress Reports**	**16d?**	**03/04/2014**	**03/25/2014**	
23	4.4.1		Progress Report	1d?	03/04/2014	03/04/2014	
24	4.4.2		Progress Report	1d?	03/25/2014	03/25/2014	
25	**4.5**		**⊟Dry Runs**	**6d?**	**04/22/2014**	**04/29/2014**	
26	4.5.1		Presentation Dry Run	1d?	04/22/2014	04/22/2014	
27	4.5.2		Presentation Dry Run	1d?	04/29/2014	04/29/2014	
28	4.6		Website	1d?	05/05/2014	05/05/2014	
29	4.7		Report	1d?	05/05/2014	05/05/2014	5,15
30	4.8		Final Presentation	1d?	05/09/2014	05/09/2014	5,15