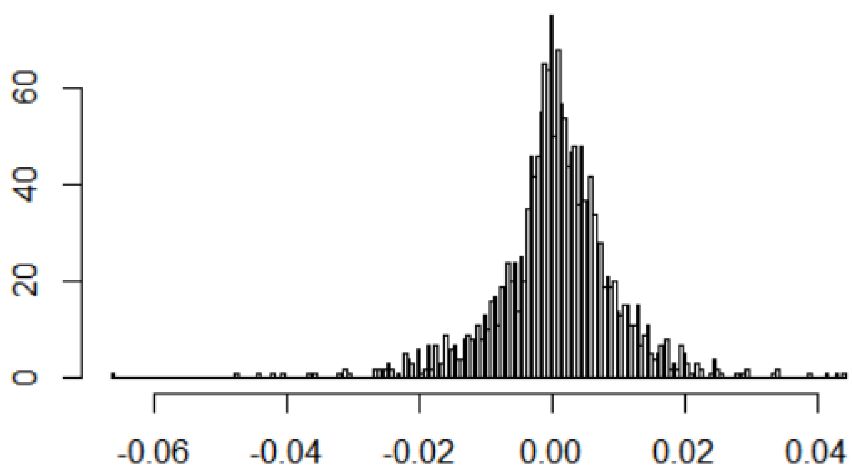
Modeling Stock Returns and Pricing Options

Overview

The distribution of a stock's returns can be just as important as its expected return. Accurately modeling the distribution helps us measure the risk of the investment. It is also critical in pricing derivatives such as options. This paper will examine how different distributions can be fit to a stock's empirical returns. A T-distribution is shown to fit this data set best. Splitting the distribution and modeling the positive returns separately from the negative returns can also add value. We will then examine how these distributions can be used to price options, but conclusions about the profitability of trading on such a method will require further investigation.

Modeling Stock Returns

The VOO fund was examined since it tracks the S&P 500 (and since Yahoo stock data had that annoying-ass error the other week). 1-day log returns for this fund have the distribution shown:

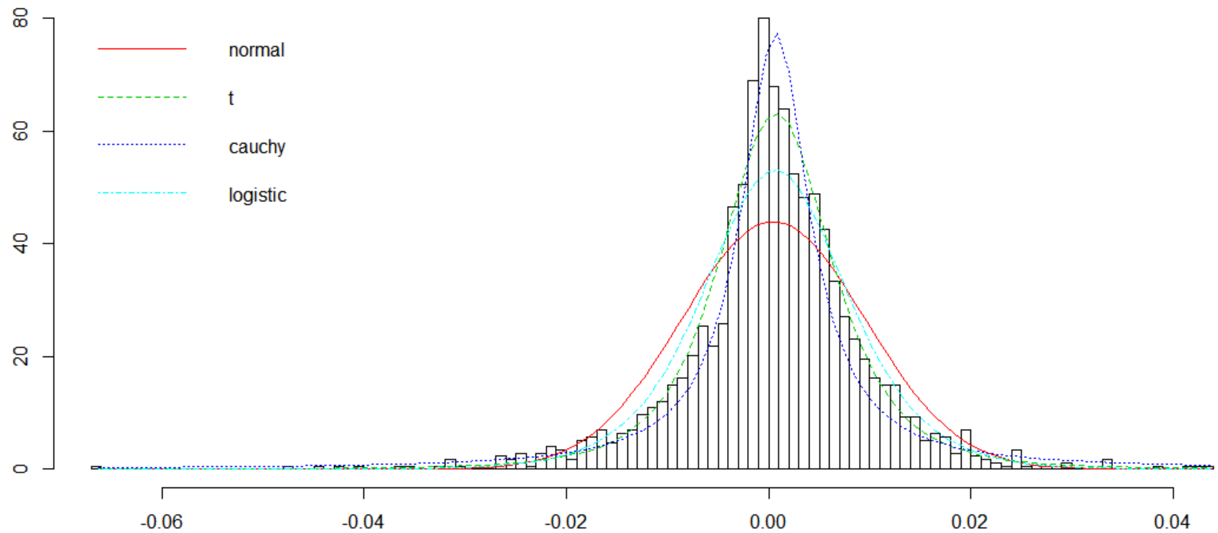



```
Mean = 0.0004647934
Sd   = 0.00907076
```

Several distributions were fit to these returns. Log-likelihood results are below:

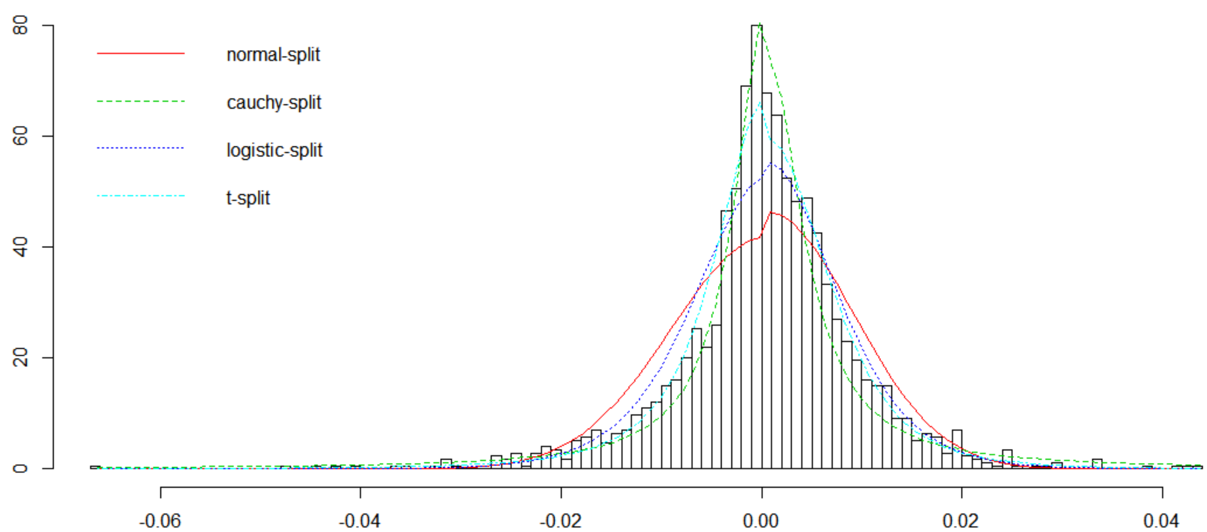
	name	loglik	notes
1	t	5861.753	location=0.0007369255, scale=0.005814103, df=2.999417
2	logistic	5824.292	location=0.0006537631, scale=0.004706983
3	cauchy	5739.547	location=0.0006721006, scale=0.004105272
4	normal	5707.676	mean =0.0004647934, sd =0.0907076

As shown, the T-distribution is the best fit for these returns. The normal distribution – which is widely used in academic circles – is actually the worst of the four options. The PDF for each distribution is plotted over the empirical distribution in the chart below. Notice how the normal distribution fails to capture both the central mass of the distribution as well as its long tails. The t-distribution captures this shape much better, pulling close to gather the middle while still going wide in the tails.



Even visually, we can see how the negative returns have a much longer tail than the positive side. For this reason, it may make sense to split the distribution in half and build separate distributions for the upper and lower returns. We do this for each of the four returns and get:

```
           name   loglik
        t-split 5865.367
 logistic-split 5825.133
   cauchy-split 5740.614
   normal-split 5712.015
```



Once again, the t-distribution is the best fit. Splitting the distribution in half improved the score slightly. The parameters for each half of the t-distribution are:

Modeling Stock Returns and Pricing Options

Tail	Location	Scale	DF
lower	0.0005545295	0.005358364	2.375896
upper	0.0005545295	0.006277177	3.837722

As might be expected, the degrees of freedom parameter is lower for the negative tail. This will extend the lower tail and make it "fatter" than the upper tail. The upper tail does have a larger scale variable though.

If we take into account the extra number of parameters we've created, splitting the distribution may not be warranted. The AIC is shown for each fit below. This takes into account the number of parameters in each model and penalizes both the t and t-split distribution for their extra parameters. When this is taken into account, the simple, non-split t-distribution appears best.

```
          name   loglik params    akaike
             t 5861.753      3 -5855.753 <- best
      logistic 5824.292      2 -5820.292
        cauchy 5739.547      2 -5735.547
        normal 5707.676      2 -5703.676

          name   loglik params    akaike
       t-split 5865.367      6 -5853.367 <- 2nd best
logistic-split 5825.133      4 -5817.133
  cauchy-split 5740.614      4 -5732.614
  normal-split 5712.015      4 -5704.015
```

The practitioner will need to determine if the split is warranted. In the next section we will look at the same stock returns over 30 days. In this case, the split distribution does appear best, as shown in the scores below:

```
          name   loglik params    akaike
             t 3176.146      3 -3170.146 <- 2nd best
      logistic 3168.327      2 -3164.327
        normal 3110.153      2 -3106.153
        cauchy 2989.924      2 -2985.924

          name   loglik params    akaike
       t-split 3212.097      6 -3200.097 <- best
logistic-split 3192.817      4 -3184.817
  normal-split 3162.528      4 -3154.528
  cauchy-split 2986.870      4 -2978.870
```

Pricing Options

So how do these different distributions affect how we might price options? The tables below show the prices of options on the VOO fund on August 4, 2017. The current price of the fund on this date was $227.26.

Modeling Stock Returns and Pricing Options

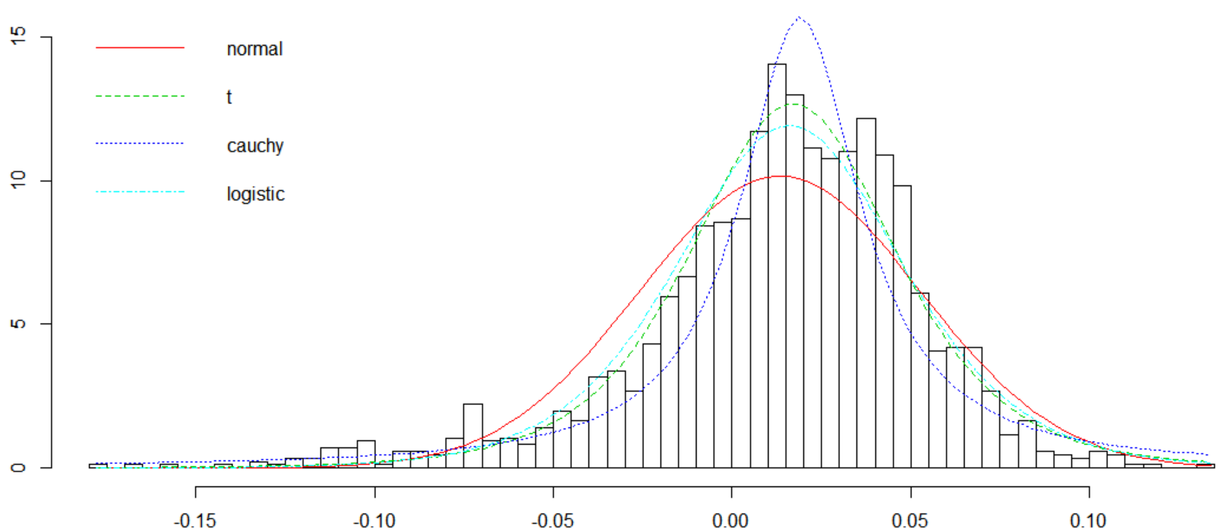
Calls for September 15, 2017

Contract Name	Last Trade Date	Strike	Last Price	Bid	Ask	Change	% Change	Volume	Open Interest	Implied Volatility
VOO170915C00230000	2017-08-04 9:48AM EDT	230.00	1.20	0.95	1.15	0.30	33.33%	3	24	7.30%
VOO170915C00235000	2017-07-28 11:46PM EDT	235.00	0.55	0.00	0.25	0.00	-	3	3	7.44%

Puts for September 15, 2017

Contract Name	Last Trade Date	Strike	Last Price	Bid	Ask	Change	% Change	Volume	Open Interest	Implied Volatility
VOO170915P00199000	2017-07-28 11:47PM EDT	199.00	0.10	0.10	0.35	0.00	-	1	1	23.68%
VOO170915P00200000	2017-07-28 11:47PM EDT	200.00	0.10	0.10	0.40	0.00	-	1	1	23.61%
VOO170915P00205000	2017-07-28 11:47PM EDT	205.00	0.15	0.25	0.50	0.00	-	1	1	20.92%
VOO170915P00210000	2017-07-28 11:47PM EDT	210.00	0.40	0.40	0.60	-0.05	-25.00%	1	2	17.88%
VOO170915P00215000	2017-07-28 11:47PM EDT	215.00	0.60	0.65	0.90	0.00	-	9	10	15.67%
VOO170915P00220000	2017-08-02 9:30AM EDT	220.00	1.05	0.90	1.20	0.00	-	4	8	12.46%
VOO170915P00225000	2017-07-31 3:55PM EDT	225.00	2.01	1.70	2.25	-0.04	-1.95%	5	15	10.65%
VOO170915P00230000	2017-08-04 3:55PM EDT	230.00	3.60	3.30	3.80	-0.30	-7.69%	2	13	6.98%

There are 6 weeks – or 30 trading days – until these options expire. Stock returns over 30 days were calculated and distributions fit to them. Results are below. Once again, the t-distribution provided the best fit:




```
      name   loglik
1        t 3176.146
2 logistic 3168.327
3   normal 3110.153
4   cauchy 2989.924
```

Modeling Stock Returns and Pricing Options

As mentioned earlier, the split t-distribution provides an even better fit, even after accounting for the additional parameters. When we use any of these distributions to price options, we end up with very different prices than those observed in the market:

Calls:

		Implied Volatility		Price		
Strike	Ask	Yahoo!	GBS	Normal	T	Split-T
230	1.15	7.30%	7.16%	3.85	4.08	4.12
235	0.25	7.44%	7.28%	1.83	1.92	1.98

Puts:

		Implied Volatility		Price		
Strike	Ask	Yahoo!	GBS	Normal	T	Split-T
199	0.35	23.68%	23.19%	0.0002	0.03	0.06
200	0.40	23.61%	23.11%	0.0004	0.03	0.06
205	0.50	20.92%	20.49%	0.004	0.05	0.09
210	0.60	17.88%	17.51%	0.02	0.10	0.14
215	0.90	15.67%	15.34%	0.13	0.21	0.24
220	1.20	12.46%	12.20%	0.51	0.48	0.49
225	2.25	10.65%	10.42%	1.47	1.16	1.17
230	3.80	6.98%	6.83%	3.35	2.71	2.75

The strike price, ask price, and implied volatility come from Yahoo! Finance. The GBS implied volatility is calculated using the "fOptions" package in R. As an example:

```
> GBSVolatility(1.15, "c", 227.26, 230, 30/250, 0.00, 0.00)
[1] 0.07155463
```

The formula above uses a risk-free rate of 0% since this more closely matched the results given by Yahoo. The 1-month treasury yield is currently 1%. If we use this in the model instead of a 0% risk-free return, we get results that differ even more from Yahoo's results:

```
> GBSVolatility(1.15, "c", 227.26, 230, 30/250, 0.01, 0.01)
[1] 0.06848211
```

The calculated volatilities match those from Yahoo rather closely. The actual mean and standard deviation of these returns is 1.3% and 3.9%. Annualized, these are 11.13% and 11.32%. If we examine returns over 250 days, we see a mean of 10.11% and a standard deviation of 8.01%. This latter value seems closer to the implied volatilities near the money, but this seems low compared to historical averages.

The table of prices (above) were calculated by taking 1,000,000 random samples from each distribution. The following R code then calculates the expected value of an option with a current price of S and exercise price X:

```
S <- 227.26
X <- 235
TypeFlag <- "c"  # “c” = call, “p” = put
r.sample <- result$fits[['t-split']]$sample(1000000)
S.T.sample <- S * exp(r.sample)
V.T.sample <- switch(TypeFlag,
  c=pmax(0, S.T.sample-X),
  p=pmax(0, X-S.T.sample)
  )
mean(V.T.sample)
```

The results above imply that the call options are under-valued quite significantly. The 230 call option is valued at $3.85 - $4.12 even though the market price is $1.15. The 235 call option is valued at $1.83 - $1.98 while the market is at $0.25. This makes sense when we notice that the average 30-day return is 1.3%. The expected value of the stock is then $230.32. There’s a 50% chance that the option will expire in the money, and the upside averages out to more than $1.15. Conversely, it appears that the put options are over-valued. The market is placing a much higher probability than we would that the stock will drop by more than 10% to 200.

It is well-known that empirical distributions under-estimate the probability of large losses – those “Black Swan” events that may still occur even though nothing like them is in our 10 years of sample data. This would make us hesitate to purchase any of these put options. However, should we consider buying call options? For whatever reason, the market option prices imply that this fund has much less upward potential than history would suggest. This is likely due to the high valuations out in the market today. Once again, the trader is left to determine whether the historical model is a good predictor of the future, or is the market taking into account other factors and producing prices that are better than those based on historical returns.

Conclusion

Further investigation is required to determine which is a better method for pricing options. If we fit probability distributions to historical returns and use these to estimate the future value of an option, can we identify mis-pricing opportunities? Or does the market know something that we do not know? In either case, it will still be important to be able to model the distributions of stock returns. Even if we are just forecasting stock prices, our model of the error distribution will help us quantify risk and uncertainty in our forecasts. For this purpose, a T-distribution seems best. Splitting the data and modeling the upside returns separately from the downside returns was also shown to add value in some cases.