Logarithmic Stock Price Models
09/10/2025

After many years of trial and error, it was Matthew Osborne of the Naval Research Laboratory (NRL) who demonstrated in 1959 that the stock price followed a lognormal distribution, not a normal distribution. This is further formalized in the stock price (S) stochastic differential equation:

$$d(log(S)) = \mu dt + \sigma dw$$

where μ is the drift coefficient, σ volatility, and w a Wiener process following $N(0, 1)$. Therefore, the logarithmic stock price follows a normal distribution, $N(\mu, \sigma)$.

Normal Distribution

The stock price data of SPY (SP500, from 2005 to 2025) are collected and tabulated in their logarithmic values. The script below (Figure 1) performs the Maximum Likelihood Estimation of the model parameters. The stock() function reads the price data from a file ("CSPY.prn") and the function retn() computes the "simple daily returns" from the prices. Since the prices are listed in their logarithmic values, exp() function is used to covert them to the real prices. The "logarithmic return" is defined as the logarithmic value of the "simple daily return" plus 1. The logarithmic return also equals $d(log(S))$, therefore follows $N(\mu, \sigma)$. The function cdf() computes the "empirical cumulative distribution function" of the logarithmic returns.

```
require(stats4)
tkr="SPY"
stock=function(fn){dat=read.table(file=fn, header=F);return(dat)}
retn=function(x){len=length(x);retn=vector(length=len); for (i in
2:len) retn[i]=exp(x[i]-x[i-1])-1; return(retn)}
cdf=function(x){len=length(x); cdf=x; for (i in 1:len)
cdf[i]=i/len; return(cdf)}
dec=function(y){return(sprintf("%.2f",y))}

b0=stock(paste("C",tkr,".prn",sep="")); c0=b0[,3];
r1=log(1+retn(c0));  #r1=retn(c0);
cdf1=cdf(r1); r0=sort(r1)

## base model
NegLogLik2=function(mu,sigma){-
sum(log(dnorm(r1,mean=mu,sd=sigma)))}
MLE.fit2=mle(minuslogl=NegLogLik2,start=list(mu=mean(r1),sigma=sd(r
1)),method="Nelder-Mead")
```

Figure 1. Lognormal Distribution Model

Figure 2 compares the parametric model CDF using the estimated parameters against the empirical CDF (or EDF). The curve agrees with the data well in general except within a small center portion of the logarithmic returns. However, the critics have questioned the validity of using $N(\mu, \sigma)$ for the distribution, not for the small disagreement as shown in Figure 2, but a phenomenon known as the "fat tails" in the EDF that do not actually shown in Figure 2. They often cite the "fat tails" to explain the "black swan events" of sudden, deep drawdowns in the stock prices. The critics advocate the use of the "stable distribution" (also known as the Cauchy distribution) as the Wiener process:

$$pdf(x) = \frac{1}{\pi(a + x^2)}$$

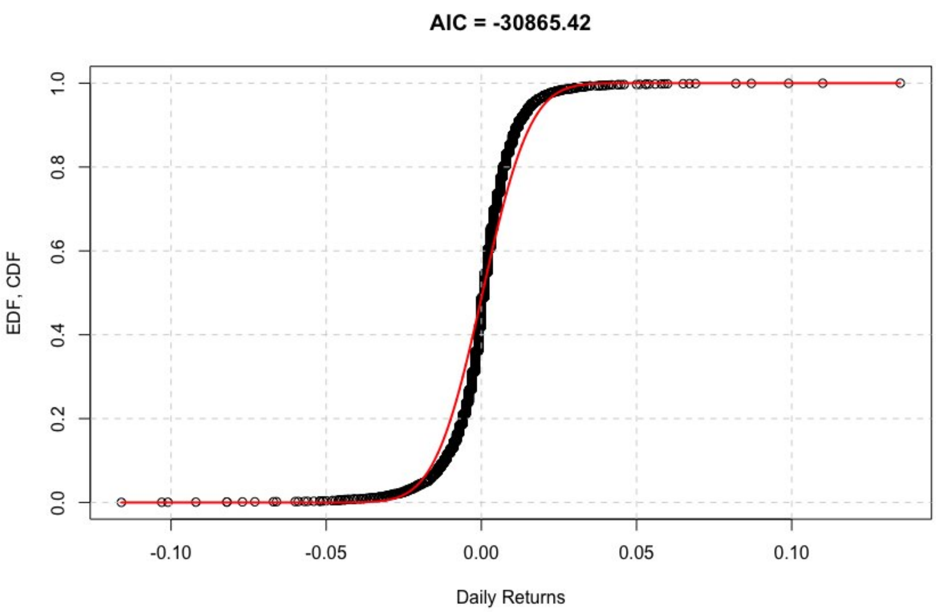

Figure 2. CDF (red line) of the Normal Distribution and EDF (circles) of SPY Logarithmic Returns.

Stable Distribution

A two-parameter stable distribution is defined as below:

$$cdf(x,\mu,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+((x-\mu)/\sigma)^2}$$

where μ and σ are also called location and scale parameters, respectively. Figure 3 lists the script for estimating these parameters. Figure 4 compares the model CDF against the EDF of the price data.

```
## stable distribution
NegLogLik2=function(loc,scl){sum(-dcauchy(r1,location=loc,
scale=scl, log=TRUE))}
MLE.fit2=mle(minuslogl=NegLogLik2, start=list(loc=median(r1),
scl=IQR(r1)/2),method="Nelder-Mead")
```

Figure 3. The Stable Distribution Model.

To the credit of the stable distribution proponents, the center portion of the data now match the model quite well. At the same time, however, to the great dismay of the proponents, the “fat tails” shown as part of the model do not match the actual data.

In his seminal book “*Analysis of Financial Time Series*”, Ruey Tsay of University of Chicago proposed the use of a mixture of normal distributions as the Wiener process:

$$pdf(\mu,\sigma_1,\sigma_2) = (1-X)N(\mu,\sigma_1) + XN(\mu,\sigma_2)$$

where X is a Bernoulli random variable such that $P(X=1)=\alpha$ and $P(X=1)=1-\alpha$. We can assume X as an additional parameter of the combined distribution.

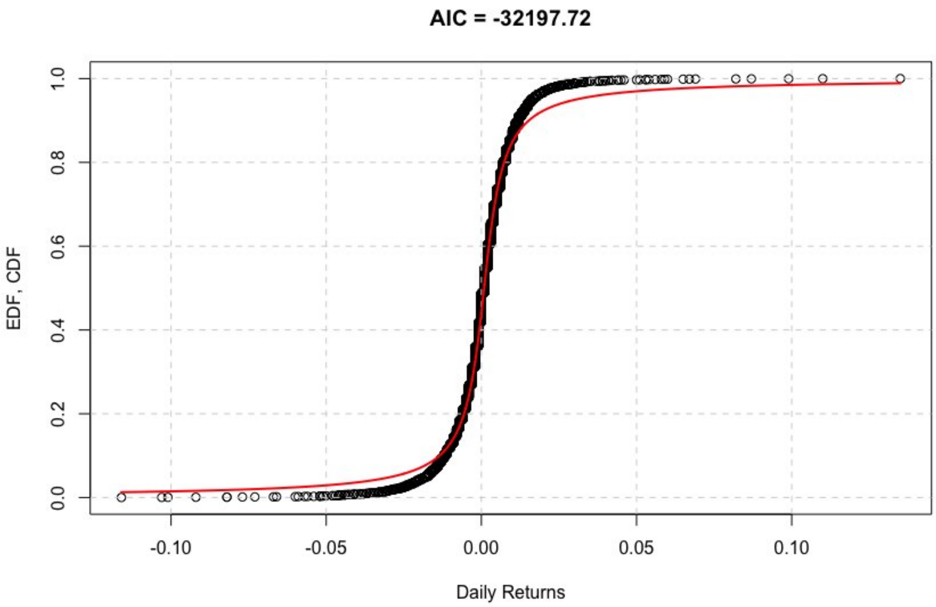

Figure 4. CDF (red line) of the Stable Distribution and the EDF (circles) of SPY Logarithmic Returns.

Mixture of Normal Distributions, Identical Means

This mixture model may be viewed as two random processes all follow the normal distribution with the same mean. The second process has a higher volatility and fewer numbers (fraction of occurrences) than the first. This often exhibits in the stock returns as the so-called "volatility clustering". The maximum likelihood estimation procedure for this model is listed in Figure 5 where p is the fraction of the second process in the entire random events and the initial guess of the fraction is set to 0.5.

```
## two sigma
NegLogLik2=function(mu,sigma1,sigma2,p){-sum(log((1-p)*dnorm(r1,
mean=mu,sd=sigma1)+p*dnorm(r1,mean=mu,sd=sigma2)))}

MLE.fit2=mle(minuslogl=NegLogLik2, start=list(mu=mean(r1),sigma1
=0.5*sd(r1),sigma2=2*sd(r1),p=0.5),method="Nelder-Mead")
```

Figure 5. The Mixture of Normal Distributions Model with Identical Means.

The model CDF is plotted along with the EDF of the price data in Figure 6. It is apparent that the model now matches the data quite nicely, much better than both the simple (two parameters) normal distribution and the stable distribution. Also, the Akaike's Information Criterion (AIC) of the mixture normal distribution model is smaller than those of the simple normal distribution and the stable distribution.

To make a more general model, the identical mean of the mixture of normal distributions can be assumed to be different and the maximum likelihood estimation script is modified in Figure 7. This model technically follows a bimodal distribution with each mode being of normal distribution.

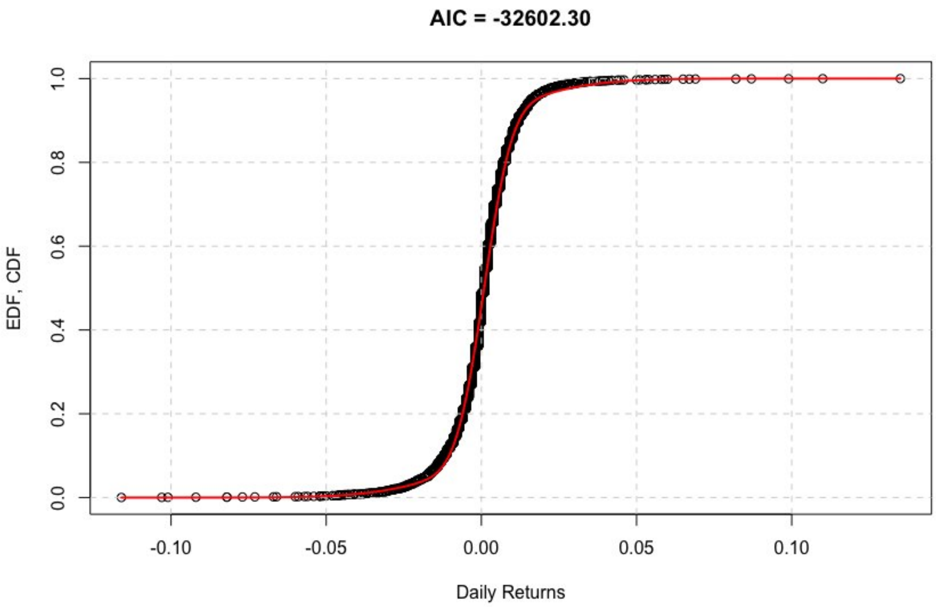

Figure 6. CDF (red line) of the Mixture of Normal Distributions and the EDF (circles) of SPY Logarithmic Returns.

Mixture of Normal Distributions, Different Means

The MLE procedure for mixture of normal distributions with different means is only a slight change from that in Figure 5 with an addition of the mu2 parameter.

```
## bi-modal
NegLogLik2=function(mu1,mu2,sigma1,sigma2,p){-sum(log((1-
p)*dnorm(r1,mean=mu1,sd=sigma1)+p*dnorm(r1,mean=mu2,sd=sigma2)))}

MLE.fit2=mle(minuslogl=NegLogLik2,start=list(mu1=mean(r1),mu2=
mean(r1),sigma1=0.5*sd(r1),sigma2=2*sd(r1),p=0.5),method="Nelder-
Mead")
```

Figure 7. The Mixture of Normal Distributions Model with Different Means.

Figure 8 shows the calculated model CDF and the actual price EDF. The match between the two is excellent throughout the range of the logarithmic returns.

NOTE: AIC and Confidence Interval

The Akaike's Information Criterion (AIC) is defined as:

$$AIC = 2k - 2log(\mathcal{L})$$

where k is number of parameters and $log(\mathcal{L})$ is the value of logarithmic maximum likelihood. The confidence interval (CI) for the sample size n, standard error S, at the level of $\alpha \times 100\%$ is

$$CI = x \pm t_{n-k,\alpha/2}\frac{S}{\sqrt{n}}$$

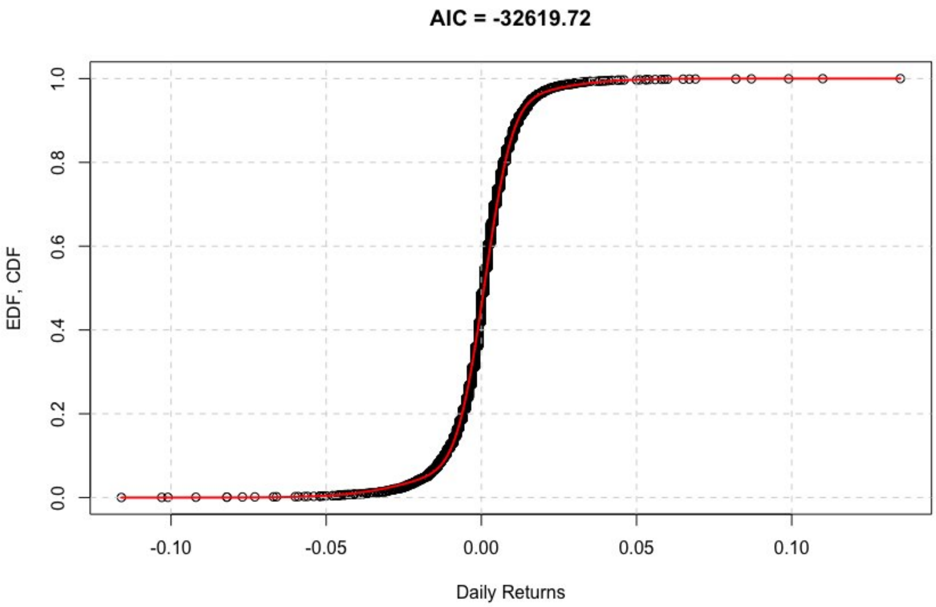

Figure 8. CDF (red line) of Mixture of Normal Distributions with Different Means and the EDF (circles) of SPY Logarithmic Returns.

Concluding Remarks

The table below summarizes the results of several models for the SPY price data. The mixture normal distribution bi-modal model has the lowest AIC, suggesting that this model more accurately describes the price data. The high volatility returns comprise approximately 18% of the price data. The high volatility returns also have a negative mean (-0.0028). If one would avoid these high volatility events, the annualized return could be substantially enhanced (0.001*252*0.82 versus 0.00037*252).

Model	Parameters	AIC

Normal Distribution	`Estimate  Std. Error`
	`mu      0.0003743938 0.0001687948`
	`sigma   0.0121191810 0.0001157358`
		-30865

Stable Distribution	`loc     0.0008793409 9.785848e-05` `scl     0.0046777808 8.131899e-05`	-32198

Mixture, Identical Means	`mu      0.000894451  0.0001203584`
	`sigma1  0.006958160  0.0001578241`
	`sigma2  0.024679472  0.0009173811`
	`p       0.176078991  0.0147939153`
		-32602

Mixture, Different Means	`mu1     0.001082988  0.0001279923`
	`mu2    -0.002831794  0.0008799129`
	`sigma1  0.006910551  0.0001567754`
	`sigma2  0.024196653  0.0008748467`
	`p       0.179974190  0.0148528845`
		-32620