An Introduction to Modeling Stock Price Returns With a View Towards Option Pricing

Kyle Chauvin

August 21, 2006

This work is the product of a summer research project at the University of Kansas, conducted under the guidance of Dr. Bozenna Pasik-Duncan. It presents a rudimentary description of the procedures and applications of stock price modeling, and furthermore seeks to replicate the study of the Hyperbolic distribution carried out by Eberlein and Keller, 1995.

1 The Need for Modeling

Financial options, namely stock options, are ways in which investors can manage the risk level of their portfolios and control the timing of various cash flows. Because, in most basic terms, this class of derivatives consists of agreements to buy or sell financial assets (here, shares of stock) at a prescribed time in the future, determining their fair market value requires a prediction of the future price of the asset to be bought or sold. Mathematical models are therefore employed to most accurately predict the future behavior of stock prices.

The owner of a European-style\(^1\) stock option has the "right, but not the obligation" to either buy\(^2\) or sell\(^3\) a share of stock \(S\) at time \(T\) for price \(K\). In financial terms, \(S\) is the underlying asset, \(T\) the time of expiration, and \(K\) the strike price. The fair trading value of the option is derived as following:

\(^{1}\)There are several varieties of options. While European options require that the option owner wait until expiry to exercise the option, American-style options allow the owner to exercise it at any point before expiry. This paper exclusively discusses European-style options, though many of the conclusions can be applied to Americans, as it has been shown that the optimal time of exercising an American option is also at expiry.

\(^{2}\)Making the option a "call option"

\(^{3}\)In this case, a "put option"
Let the price of the underlying stock be defined by the stochastic process

\[ \{ S_t \}, \ t = 0, 1, 2 \ldots \]  \hspace{1cm} (1)

where \( t \) denotes successive trading days. Assuming we are dealing with a call option, the profit made by exercising the option is

\[ S_T - K \]

Note that for option pricing, the value of the stock at expiration can only be defined by its expected value, furthermore, the owner of the option will only exercise the option when there is a profit to be made. Therefore the value of the option is better defined by

\[ \mathbb{E}[\max(S_T - K, 0)] \]

Finally, the option value must be discounted for the risk free interest rate, \( r \). Considering that whatever profit is made from exercising the option at the end of the time period could theoretically have been invested in a risk-neutral bond at the beginning, the actual profit is discounted by a factor of \( e^{-rT} \). Therefore, the expected payoff of a stock option, and thus the fair market price of the option at time \( t = 0 \) is

\[ e^{-rt} \mathbb{E}[\max(S_T - K, 0)] \]  \hspace{1cm} (2)

The need to accurately predict \( \mathbb{E}(S_T) \), which is the only piece of (2) unknown at time \( t = 0 \), requires modeling of \( \{ S_t \} \). In particular, we are interested in the behavior of the stock’s returns, its day to day change in price.

2 Stock Returns and their Distributions

Defining Returns

Given \( \{ S_t \}, t = 0, 1, 2 \ldots \) the price of stock \( S \) at say the close of each trading day,\(^4\) there are two ways to define a daily ”return” of the stock. The simplest, though less convenient method is to define return \( Y_t \) by

\[ Y_t = \frac{S_t - S_{t-1}}{S_{t-1}} \]  \hspace{1cm} (3)

\(^4\)However, \( S_t \) could also be said to represent the price of a stock at the end of every \( j \) trading days, or every month or year. Such definitions have produced different results with respect to distribution fitting.
which is just the ratio of the stock price’s daily movement to its price the day before. However, in order to express \( S_T \) through \( S_0 \) and \( \{ Y_t \} \) we need the more complex expression

\[
S_T = S_0 \prod_{i=1}^{n} (Y_i + 1)
\]  

(4)

If we instead define the return by

\[
X_t = \ln \frac{S_t}{S_{t-1}} = \ln S_t - \ln S_{t-1}
\]

(5)

then we can then express \( S_T \) by

\[
S_T = S_0 \exp\left\{ \sum_{i=1}^{n} X_i \right\}
\]

(6)

conveniently containing a sum and not a product of successive returns. For this reason \( X_t \) is used as the standard definition.

Assuming that the \( X_i \) are independent and identically distributed, the distribution of \( X_i \) provides all the needed information to calculate \( S_T \). While the Normal distribution is the most obvious and most widely discussed choice for modeling \( S_T \), it pales in comparison to the Hyperbolic distribution, whose family of distributions also includes the Student’s T (which converges to the standard normal). Other distributions that can be used include the Binomial model (which also converges to the normal), and the \( \alpha \)-stable class of distributions.

**The Normal Distribution**

For reference, the Normal distribution is easily defined by its probability density function

\[
f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

(7)

If the returns of a stock are said to be normally distributed, then the overall process \( \{ S_t \} \) is a case of Geometric Brownian Motion. GBM is a simple deviation from Standard Brownian Motion in that if \( B(t) \) is a stochastic process governed by Standard Brownian Motion:

\[
(B_t - B_s) \sim N(\mu, (t-s)), \; t > s > 0
\]

(8)

then \( G_t \), governed by Geometric Brownian Motion, is:

\[
(G_t - G_s) = G_0 e^{(t-s)\mu + \sigma B(t)}
\]

(9)
Thus, the advantage of assuming Normal distribution of returns lies in its connection to Brownian Motion, a topic widely studied and well within the academic “mainstream”.

The Student’s T Distribution

A well known substitute for the Normal, the Student’s T distribution has a probability density function of

\[
f_X(x|v) = \frac{\Gamma\left[\frac{1}{2}(v + 1)\right]}{\sqrt{\pi v} \Gamma\left(\frac{1}{2} + \frac{x^2}{v}\right)(v+1)/2}
\]  

Advantages of the T distribution include that it is uniquely defined by one parameter, \(v\), the number of degrees of freedom, and that it can be more accurate in dealing with small sample sizes where the variance is not previously known. However, the T distribution approaches the Standard Normal as \(v\) approaches infinity, so it is little better in fitting stock returns from a large sample.

The Hyperbolic Distribution

The Hyperbolic distribution, a member of the Generalized Hyperbolic distribution class, has p.d.f.

\[
f_X(x|\alpha, \delta, \beta, \mu) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha K_1(\delta \sqrt{\alpha^2 - \beta^2})} e^{-\alpha \sqrt{\delta^2 + (x-\mu)^2} + \beta (x-\mu)}
\]

or

\[
f_X(x|\chi, \psi, \beta, \mu) = \frac{\sqrt{\psi}}{2\sqrt{\chi} \psi + \beta^2 K_1(\sqrt{\chi}\psi)} e^{-\sqrt{\psi^2 + \beta^2(\chi + (x-\mu)^2)} + \beta (x-\mu)}
\]

where

\[
K_\nu(x) = \frac{1}{2} \int_0^\infty x^{\nu-1} e^{-\frac{1}{2}x(z+\frac{1}{z})} dz
\]

This distribution was originally developed to model sand distributions in South Africa, though by now its primary application is in finance. The advantage of using a Hyperbolic-based model for stock prices is high accuracy, which well exceeds that of the Normal and T distributions. However, the complexity of its definition introduces problems not seen with its alternatives. Firstly, the parameters of the Hyperbolic distribution, while not com-
pletely arbitrary, are not indicators of real-life conditions.\textsuperscript{5} Secondly, parameter estimation is a far more sophisticated process, best achieved through statistical software.

3 Numeric Results

The purpose of this section is to verify the claims of Eberlein and Keller (1995), who stated that the Hyperbolic distribution is far more powerful in modeling actual stock returns than the more widely accepted Normal distribution. In light of the ease with which American stock data could be obtained, and in the interest of applying the Hyperbolic distribution beyond the German stocks studied by Eberlein and Keller, six stocks from the NYSE (AT&T, Bank of America, Coca-Cola, Disney, Ford, and Pfizer) were selected for inspection. All of these stocks were chosen for their high volume of trading. Their daily closing prices from January 10, 2002 to December 30, 2004, adjusted for splits and dividends, were obtained from Yahoo!.com Finance, and verified by GlobalFinancialData.com. In all, the data for each stock consisted of prices for 1002 trading days, or 1001 daily returns. Histograms of the return data for each stock appear in Figure 1.

Fitting the Data

The statistical software R was used in all calculations. For each hypothesized distribution, the Normal, Student T, and Hyperbolic, parameters were estimated through R, sample distributions were simulated in order to generate QQ-Plots, and the Chi-Squared goodness of fit test was applied to measure the goodness of fit. Parameter estimation for the normal distribution consisted of the usual sample mean and sample standard deviation:

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_n \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]  

(14)

Because the Student’s T distribution is uniquely defined by the number of degrees of freedom, \(v\), the data itself had to be adjusted in order to determine the goodness of fit. For each return \(X_t\), an adjusted value was calculated by:

\[
X'_t = \frac{X_t - \bar{X}_n}{S_n/\sqrt{n}}
\]  

(15)

\textsuperscript{5}With the Normal distribution, the mean and variance of the model, properly adjusted, respectively represent the drift and volatility of a given stock.
Figure 1: Return Data from the NYSE Stocks
Parameter estimation with the Hyperbolic distribution requires powerful statistical software. In this study, the data was fitted to the Hyperbolic distribution through the "HyperbolicDist" package for R, created by David Scott. Tables 1 and 2 list the values of the parameters used.

### Goodness of Fit: the Chi-Squared Test

The return data from AT&T was fitted to the Normal, Student T, and Hyperbolic distributions and the results compared. To calculate goodness of fit, the $\chi^2$, or Pearson, test, was applied. 6 test. Both methods test the null hypothesis $H_0$ that data taking the "empirical distribution" $F_n$ actually comes from the known distribution $F_0$, against the alternative $H_1$ that $F_n \neq F_0$. What is more revealing, however, is the p-value for each fit: the probability of getting the empirical data set assuming $F_n = F_0$.

---

6Only because of time considerations was the Kolmogorov-Sminov test not also used. Adding this procedure to the study in order to verify the claims of the $\chi^2$ test would really be the first way to continue this research; a full discussion of improvements and extensions is given in section 5.
The \( \chi^2 \) test involves partitioning the sample space into mutually exclusive events. For each of these events \( A_i, i = 1, 2, \ldots, k \), we tabulate the number \( n_i \) of data points that fell within the event, and \( np_{io} \), the number of data points expected to have fallen within the event given that the data came from distribution \( F_0 \), where \( p_{io} = P_{F_0}(A_i) \) and \( n \) is the total number of data points \( (n_1 + n_2 + \ldots + n_k = n) \). These values are then used in calculating the \( D \) statistic, defined by:

\[
D^2 = \sum_{i=1}^{k} \frac{(n_i - np_{io})^2}{np_{io}}
\]  

(16)

A simpler version of this is:

\[
D^2 = \sum_{i=1}^{k} \frac{\text{Observed values} - \text{Expected values}}{\text{Expected values}}
\]  

(17)

In 1900 Karl Pearson proved that \( D^2 \) has a limiting distribution

\[
D^2 \longrightarrow D_0^2 \sim \chi^2(k - 1) \text{ as } n \rightarrow \infty
\]  

(18)

Thus a p-value for the test can be determined by \( p = P(\chi^2_{k-1} > d^2) \). The only caveat to this procedure is that the expected values \( np_{io} \) must be greater than or equal to 5 for each event.

In applying the \( \chi^2 \) test to each stock and distribution, the sample space was first partitioned into equal segments extending across the range of the data. Next, the events lying at the tails of the distributions, where the expected number of data points (even from a sample of 1000) was less than 5, were collapsed into each other.

**Results**

The findings of this study fully support Eberlein and Keller’s conclusion that the Hyperbolic distribution is a better fit for daily stock returns than the Normal. If the Quantile-Quantile plots do not make this clear enough, the results of the \( \chi^2 \) test are unmistakable. Table 3 gives a complete summary of the \( \chi^2 \) test results, and Figure 2 presents Q-Q plots for the AT&T return data, which is representative of the other stocks.

My sole concern in the results is that the p-values of fit for the Normal and Student’s T distributions are so small. Although I expected a significant difference between these and the p-values for the Hyperbolic distribution, such extreme figures are somewhat surprising. Perhaps the partitioning of
Table 3: P-values from the Chi-Squared Test

<table>
<thead>
<tr>
<th>Distribution</th>
<th>AT&amp;T</th>
<th>Bank of America</th>
<th>Coca-Cola</th>
<th>Disney</th>
<th>Ford</th>
<th>Pfizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>$2.67 \times 10^{-12}$</td>
<td>$1.11 \times 10^{-16}$</td>
<td>$3.67 \times 10^{-11}$</td>
<td>$3.87 \times 10^{-8}$</td>
<td>$3.66 \times 10^{-11}$</td>
</tr>
<tr>
<td>Student’s T</td>
<td>0</td>
<td>$5.10 \times 10^{-11}$</td>
<td>$5.81 \times 10^{-10}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hyperbolic</td>
<td>.166</td>
<td>.556</td>
<td>.486</td>
<td>.610</td>
<td>.916</td>
<td>.914</td>
</tr>
</tbody>
</table>

the sample space in the $\chi^2$ test was “disadvantageous” to the these two distributions, and a different sectioning would produce more expectable values. In any case, the admittedly extreme difference in p-values between the Normal / T group and the Hyperbolic goes far in highlighting the high accuracy of the latter.

4 Applications to Option Pricing

The most important application of stock price return modeling is the ability to give an accurate prediction of the stock’s future price. As discussed previously, this is a necessary factor in option pricing. If normality is assumed, than what is derived as a pricing formula is known as the Black-Scholes model.

An alternative to Black-Scholes is the binomial model, which instead of assuming a continuous distribution of returns, posits that each day, a stock either goes up or down by a fixed factor, with a certain probability. Carried out in time, the binomial model produces a "tree" of possible stock prices which can be "traced" backwards from the expiry time, when the value of the option is known, to the purchase time.

Eberlein and Keller, in their work with the Hyperbolic distribution, extend the Hyperbolic model of returns to an options pricing formula, similar to Black-Scholes, yet more accurate in that it draws from a more accurate model. Unfortunately, the mathematical sophistication needed to investigate and present this model, and others based on the Hyperbolic distribution, is beyond the scope of this author.

5 Ways to Continue and Build on this Research

There are several ways in which this research could potentially be extended. The first is investigating other distributions, and comparing fits. Namely, the Generalized and Normal Inverse Gaussian distributions, who like the
Figure 2: QQ Plots of AT&T Return Data fit to Various Distributions
Hyperbolic and Student’s T, are special cases of the Generalized Hyperbolic distribution. Stable distributions, of which the Normal, Cauchy, and Lévy are special cases, are another possibility. Finally, the Pareto distribution has been mentioned as an alternative.

A second path for continuing this study would be to further investigate the applications of return modeling, not just to stock options, but to futures and other derivatives.

Yet another option would be to compare the modeling here in finance to other branches of science using similar distributions and similar models.

Acknowledgments

The work I have done in completing this project was supported by a Research Experience for Undergraduates grant from the National Science Foundation, though my professor Dr. Bozenna Pasik-Duncan. I am incredibly appreciative of the opportunity to have worked on this study, for the knowledge, skills, and tools gained in the experience could not have been gained by other means. Finally I must thank Yeang-Yeow (Austin) Yuen, whose paper provided an excellent starting point for my own investigations and very useful reference.

References


