Time and Finance: Exploring Variance in the Black-Scholes Model

Edward Chase Skorupa
Bard College

Time and Finance: Exploring Variance in the Black-Scholes Model

A Senior Project submitted to
The Division of Science, Mathematics, and Computing
of
Bard College

by
Edward Chase Skorupa

Annandale-on-Hudson, New York
May, 2019

Abstract

In their 1973 paper, The Pricing of Options and Corporate Liabilities, Fischer Black and Myron Scholes published mathematical methods they had devised with the goal of accurately pricing European options. When using the model to predict future options prices, all input variables in the model can be empirically viewed, and calculated, at present time except for the future volatility of the underlying security. Retrospectively analyzing the volatility implied by the Black-Scholes model using price history shows that this implied volatility is an inaccurate estimate of actual future volatility. This project sought to explore the relationship between the implied future volatility of a stock and the Black-Scholes model, and if future implied volatility can be better predicted by incorporating an alternate measure of variance, Allan Variance. Allan Variance was first shown by physicist David Allan in Allan's Statistics of Atomic Frequency Standards to measure frequency stability in oscillations found in atomic clocks. While the preliminary work done in this paper does not suggest a simple way to incorporate Allan Variance into the Black-Scholes model, and measuring variance in financial securities in general, it is indeed worth further exploration.

Contents

Dedication

I dedicate this senior project to my parents, and my sister. Thank you for everything.

Acknowledgments

I would like to acknowledge the help I received from my adviser, Matthew Deady, as well as every math teacher I have had the pleasure of taking a course with while at Bard.

1
Introduction

Commonly referred to as the 2008 financial crisis, the financial shocks felt around the world in 2007 and 2008 began with the collapse of the U.S. Subprime-Mortgage lending market in the summer of 2007, an event that very few saw coming. Leading up to the summer of 2007, many of the warning signs were dismissed by financial experts and analysts as irrelevant and merely by-products of an increasingly globalized and technologically sophisticated economy. Instead of taking note of rapidly rising U.S. home values, financial institutions espoused that "the huge run-up in US housing prices was not at all a bubble, but rather justified by financial innovation(including sub-prime mortgages), as well as by the steady inflow of capital from Asia and petroleum exporters"[1] while "the huge run up in equity prices was similarly argued to be sustainable thanks to a surge in US productivity growth and a fall in risk that accompanied the 'Great Moderation' in macroeconomic volatility."[1] These arguments soon proved flawed as, seemingly inexplicably, default rates on subprime-mortgage lending skyrocketed thanks to a sudden and rather small drop in US house prices. The financial rot quickly spread throughout the global markets as the proliferation of complicated financial instruments, specifically collateralized debt obligations designed to essentially bundle large amounts of such individual debts, resulted in wide spread, and largely unforeseen risk.

While the question of what precisely led to the 2008 financial crisis and whether or not it could have been averted, let alone predicted, will continue to be discussed and analyzed in the context of previous financial crises, the 2008 crisis is unique compared to other crises such as the Great Depression, or the recession of the 1970s, because the proliferation of computer driven, quantitative financial modeling in finance is widely regarded to have played a role in the severity, if not the cause of the crisis. Trader-turned-writer Nassim Taleb touches on this point in his book *The Black Swan: The Impact of the Highly Improbable*, which explores the causes and implications of "Black Swan" events, defined as highly unlikely and highly consequential outlier events, in finance. While attempting to project and predict financial movements has always been inherent in the industry of finance, the computerization of the process has corrupted it to some extent. Before the advent of tools such as Microsoft Excel, "the activity of projecting, in short, was effortful, undesirable, and marred with self doubt"[2] while now any computer literate person is able to endlessly extrapolate any projection using such statistical tools. These unjustifiably extrapolated predictions often end up widely circulated in digital presentations where "the projection takes on a life of its own, losing its vagueness and abstraction and becoming what philosophers call reified, invested with concreteness."[2] While this trend has undoubtably contributed to difficulty in accurately understanding and predicting financial behavior, there still remain problems with the actual methods used in such predictions.

In Taleb's exploration of Black Swan events in finance he discusses the integral role human behavior plays in repeatedly facilitating such events, as he sees it. He posits that much of our inability to accurately predict real world events, such as the ones that govern financial movements, lies in our inability to fully understand the limitations of the methods we have for making predictions, and on a more fundamental level the limitations of our knowledge as we perceive it. In particular, Taleb discusses the work of 19th century philosopher and mathematician Henri Poincaré as an excellently laid out example of some of the fundamental limitations of mathematics as we currently understand it. Poincaré first formalized the concept of *nonlinearity*: a relationship in which a change in the output is not proportional to any change of the input.

As Taleb puts it, nonlinearities manifest in the real world as “small effects that can lead to severe consequences”[2] and are expressly “not an invitation to use mathematical techniques to make extended forecasts.”[2] Poincaré’s point essentially boils down to the fact that “as you project into the future you may need an increasing amount of precision about the dynamics of the process that you are modeling, since your error rate grows very rapidly.”[2] He uses the example of the *Three Body Problem* to provide a simple illustration of this point. Given two planet-like celestial objects moving in a closed system, it is not difficult to model their past movement and subsequently predict their future movement to a very accurate degree. However if even an extremely small object relative to the planets such as a comet is introduced to the system this changes. The small object may have no discernible impact on the movement of the planets initially, but over time the minuscule effect the comet has on the planet’s movements will come to have a large, and sometimes hazardous, effect on the large celestial bodies. In the world of financial predictions there is obviously much more in play in terms of small effects leading to large consequences, however the principle remains the same.

This project seeks to explore these nonlinearities that lead to Black Swan events in finance through attempting to better model and predict them. Developed to model pricing of European call options, the Black-Scholes model, named after economists Fischer Black and Myron Scholes, is widely used and accepted financial model that attempts to predict future financial based on sets of assumptions about current data. As such the Black-Scholes model will be examined in an effort to better understand and model the nonlinearities discussed earlier.

Named after atomic physicist David W. Allan, the Allan Variance, also referred to as a 2-sample M variance, was developed to measure frequency stability in oscillations, particularly those present in atomic clocks. In his paper *Statistics of Atomic Frequency Standards* Allan is able to show “a relationship between the expectation value of the standard deviation of the frequency fluctuations for any finite number of data samples and the infinite time average value of the standard deviation, which provides an invariant measure of an important quality factor of a frequency standard.”[7] The Allan Variance models frequency stability not due to systematic error, and is defined as one half of the time average of the squares of the differences between successive readings of frequency deviation sampled over the sampling period. Allan Variance is a special, 2 sample case of the more general M-Sample Variance given by:

$$\sigma_y^2(M,T,\tau) = \frac{1}{M-1}\left\{ \sum_{i=0}^{M-1}\left[\frac{x(iT+\tau)-x(iT)}{\tau}\right]^2 - \frac{1}{M}\left[\sum_{i=0}^{M-1}\frac{x(iT+\tau)-x(iT)}{\tau}\right]^2\right\} \tag{1.0.1}$$

where $x(t)$ is the phase angle, M is the number of frequency samples used in the variance, T is the time between each frequency sample, and τ is the time length of each frequency estimate. The Allan Variance case is then given by:

$$\sigma_y^2(\tau) = \mathbf{E}[\sigma_y^2(2,\tau,\tau)] \tag{1.0.2}$$

2
The Black-Scholes Model

2.1 What is the Black-Scholes Model?

While the average person may not be particularly knowledgeable, or involved with global financial markets, these markets have a large impact on many people's lives as it relates to their personal finances. Within these markets there is a never-ending process of people buying, selling, and conducting a number of other transactions involving different financial securities, commodities, and derivatives. This process, and the importance it holds, has lead numerous academics to attempt to explain these phenomena over the years, with varying results. One particular area of focus that is regarded as perhaps the most successfully explained are financial derivatives, specifically call options.

While there are many different models that are used in the pricing of call options, the most widely known is perhaps the Black-Scholes model. In 1973 Fischer Black and Myron Scholes published their paper *The Pricing of Options and Corporate Liabilities* in which they laid out what is now widely known as the Black-Scholes Formula. Black and Scholes would go on to win a Nobel Prize in Economics in 1997 for their work along with Robert Merton, who was able to show an alternate, and much more applicable derivation of the formula. Specifically, the formula gives the price of any European call option given the current stock price, strike price, time until expiration, risk-free interest rate, and the implied volatility of the stock. The formula is given by:

$$C = S * N(d_1) - K * e^{-rT} * N(d_2) \tag{2.1.1}$$

$$d_1 = \frac{ln(\frac{S}{K}) + (r + \frac{\nu^2}{2})T}{\nu\sqrt{T}}$$

$$d_1 = \frac{ln(\frac{S}{K}) + (r - \frac{\nu^2}{2})T}{\nu\sqrt{T}}$$

where S is the stock price, T is the time till expiration, K is the strike price, r is the risk-free interest rate, ν is the implied volatility, and $N(x)$ is the cumulative distribution function for a standard normal distribution. The implied volatility,ν , can also be viewed as the standard deviation of the stock's logarithmic returns, which means that ν^2 can be viewed as the variance of these logarithmic returns. While it is clear that $N(x)$ is a cumulative distribution function, it is much less clear what it calculates when substituting d_1 and d_2 for x. In his *Understanding $N(d_1)$ and $N(d_2)$: Risk-Adjusted Probabilities in the Black-Scholes Model*, Lars Nielsen describes $N(d_1)$ as "the factor by which the present value of contingent receipt of the stock exceeds the current stock price"[3] and $N(d_2)$ as "the risk-adjusted probability that the option will be exercised"[3]. In other words, they, in tandem, represent the likelihood the possessor of the call option will exercise it.

The Black-Scholes model itself is widely used in options pricing around the world due to its general accuracy in predicting options prices. In the following graph we see call option data for 480 stocks, whose companies have varying sizes, being analyzed over a 240 day trading period from January 1, 2014 to December 12, 2014. The full and dotted curved lines are both mean log returns for the forward call option prices calculated using the Black-Scholes model for 2 different values of risk-free interest rate, while the stochastic line is the mean actual price for call options for the stocks as observed in the market.

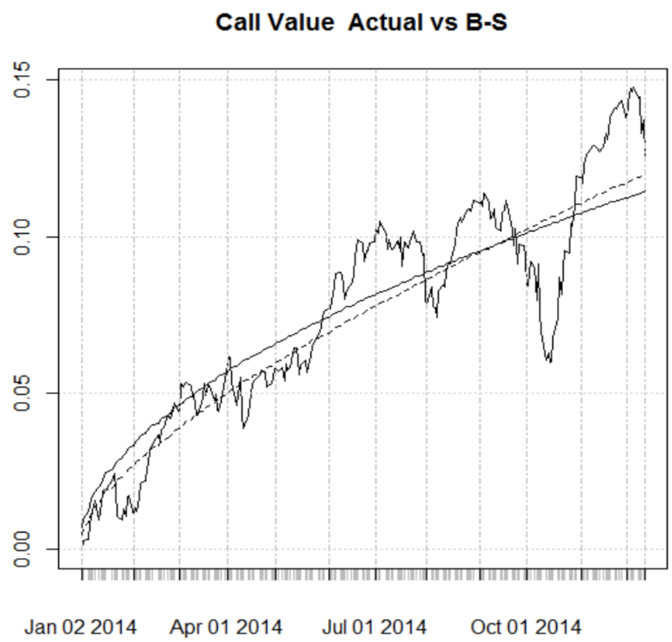


Figure 2.1.1.

As we can see, both calculations of forward options pricing using the Black-Scholes model are accurate to an extent, however fail to capture much of the noise of the inter-day movements of the actual call options prices, which reflect movement in the underlying securities.

In the following section we will show a derivation of the Black-Scholes formula that does not involve stochastic methods.

2.2 Derivation of the Black-Scholes Formula

We begin with some definitions.

Definition 2.2.1. A *Call Option* is defined as an agreement between a buyer and a seller in which the buyer has the right, but not the obligation, to purchase some financial instrument or asset at an agreed upon price during an agreed upon period. In an American call option the buyer has the option to purchase the asset at the agreed upon price at any time before the end of the agreed upon period, while in a European call option the buyer may only exercise his or her right to purchase at the end of the agreed upon period.

Definition 2.2.2. A *Put Option* is defined as an agreement between a buyer and seller in which the buyer has the right, but not the obligation, to sell some financial instrument or asset at an agreed upon price during an agreed upon period. A put options functions like the opposite of a call option.

Definition 2.2.3. We will define the *Forward Price of a Stock*, as the expected future returns of a given stock S from the starting price S_0 over the period of time t. We assume a risk-free interest rate r, and assume the stock does not pay dividends. The Forward Price, $F(t)$ is then given by:

$$F(t) = S_0 * e^{rt}$$

Definition 2.2.4. We classify a universe, or class, of stocks as a *Risk Neutral Universe* if for any stock A during time period t, the value of the stock at time $t = 0$, given by $C(A, 0)$, is equal to the expected value of the stock at time t discounted by the risk-free interest rate r to the value at $t = 0$:

$$C(A, 0) = e^{-rt}\mathbf{E}[C(A, t)]$$

Definition 2.2.5. The annual volatility of a stock, denoted ν, is defined as the standard deviation of the percent change of the stock over a year.

Definition 2.2.6. Given a continuous random variable X, the *Cumulative Distribution Function*(denoted C.D.F.), F is defined for all values x as:

$$F(a) = \mathbf{P}\{X \leq x\}$$

Definition 2.2.7. Given a continuous random variable X with C.D.F. $F(x)$ the *Probability Density Function*(denoted P.D.F.), $f(t)$, is given by:

$$F(x) = \mathbf{P}\{X \leq x\} = \int_{-\infty}^{x} f(t)dt$$

Definition 2.2.8. We define a continuous random variable X as a *Normal Random Variable* with parameters μ and $\sigma^2 > 0$ if the P.D.F. for X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where $-\infty < x < \infty$.

It then follows that the C.D.F. of a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$ would be

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

Definition 2.2.9. Given a continuous random variable X with P.D.F. $f(x)$, the *Expected Value* of X, $\mathbf{E}[X]$, is given by

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Definition 2.2.10. We define a continuous random variable X to be *Log-Normally Distributed* if given a normal random variable Y, $X = e^Y$ or that lnX is normally distributed.

Lemma 2.2.11. *Given a stock S with price S_t at time t (in years), and annual volatility ν, we assume S_t is a log normally distributed random variable such that $ln\frac{S_t}{S_0}$ is normally distributed with mean μ and variance σ^2. Then:*

$$\sigma^2(t) = \nu^2 t \tag{2.2.1}$$

$$\mu = (r - \frac{\nu^2}{2})t \tag{2.2.2}$$

Proof. We will prove equation (2.2.1) using induction.

At time $t = 1$ it is clear that $ln\frac{S_1}{S_0}$ has variance $\nu^2 = (\nu^2)1$.

Assuming that for $t-1$, $ln\frac{S_{t-1}}{S_0}$ has variance $(\nu^2)(t-1)$, for t we have:

$$\begin{aligned} ln\frac{S_t}{S_0} &= ln\frac{S_{t-1}S_t}{S_0 S_{t-1}} \\ &= ln\frac{S_{t-1}}{S_0} + ln\frac{S_t}{S_{t-1}} \end{aligned} \tag{2.2.3}$$

Which has variance $(\nu^2)(t-1)+\nu^2=\nu^2 t$.

Equation (2.2.2) is shown as follows:

$$\begin{aligned} F(b) &= \mathbf{P}\{S_t \leq b\} \\ &= \mathbf{P}\{S_0 e^x \leq b\} \\ &= \mathbf{P}\{x \leq ln\frac{b}{S_0}\} \\ &= \frac{1}{\sqrt{2\sigma\pi}}\int_{-\infty}^{ln\frac{b}{S_0}} e^{\frac{-(x_t-\mu)^2}{2\sigma}}dx \end{aligned} \tag{2.2.4}$$

To find the P.D.F. $f(x)$ for S_t we differentiate with respect to b:

$$f(x)=\frac{1}{\sqrt{2\sigma\pi x}}e^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\sigma}} \tag{2.2.5}$$

By our definition of the forward price of a stock in a risk-free universe we have $\mathbf{E}[S_t]=S_0e^{rt}$, so then

$$\begin{aligned} \mathbf{E}[S_t] &= \int_0^\infty \frac{1}{\sqrt{2\sigma\pi x}}xe^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\sigma}}dx \\ &= \frac{1}{\sqrt{2\sigma\pi}}\int_0^\infty e^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\sigma}}dx \end{aligned} \tag{2.2.6}$$

Now let $y=\frac{ln\frac{x}{S_0}-\mu}{\sqrt{\sigma}}$ where $dy=\frac{dx}{x\sqrt{\sigma}}$ and $x=S_0e^{y\sqrt{\sigma}+\mu}$ such that

$$\begin{aligned} \mathbf{E}[S_t] &= \frac{S_0}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{\frac{-y^2}{2}}e^{y\sqrt{\sigma}+\mu}dy \\ &= \frac{S_0}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{\frac{-y^2}{2}+y\sqrt{\sigma}+\mu}dy \\ &= \frac{S_0}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{\frac{-(y-\sqrt{\sigma})^2}{2}+\mu+\frac{\sigma}{2}}dy \\ &= \frac{S_0e^{\mu+\frac{\sigma}{2}}}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{\frac{-(y-\sqrt{\sigma})^2}{2}}dy \end{aligned} \tag{2.2.7}$$

Now let $x=y-\sqrt{\sigma}$. Then

$$\begin{aligned} \mathbf{E}[S_t] &= \frac{S_0e^{\mu+\frac{\sigma}{2}}}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{\frac{-x^2}{2}}dx \\ &= S_0e^{\mu+\frac{\sigma}{2}} \end{aligned} \tag{2.2.8}$$

Then since $\sigma=\nu^2t$ we have $S_0e^{\mu+\frac{\nu^2t}{2}}=S_0e^{rt}\Rightarrow\mu=(r-\frac{\nu^2}{2})t$. □

Theorem 2.2.12. *The Black-Scholes formula is as follows: Given a stock S in a risk-neutral universe with initial price S_0 and log-normally distributed price S_t at time t, the price of a European call option C with strike price K, expiration time $t = T$, and risk-free interest rate r at time $t = 0$ is given by*

$$C = S_0 N\left(\frac{rT + \frac{\nu^2 T}{2} + ln\frac{S_0}{K}}{\nu\sqrt{T}}\right) - Ke^{-rT} N\left(\frac{rT - \frac{\nu^2 T}{2} + ln\frac{S_0}{K}}{\nu\sqrt{T}}\right) \tag{2.2.9}$$

Proof. We begin with $C(S,T) = max(S_t - K, 0)$ since the amount yielded by the call option will be equal to the difference between the price of the stock on the expiration date and the strike price. We then have

$$\begin{aligned} C(S,0) &= e^{-rT}\mathbf{E}[C(S,T)] \\ &= e^{-rT}\mathbf{E}[max(S_t - K, 0)] \\ &= e^{-rT}\int_K^\infty \frac{1}{\sqrt{2\pi T}\nu x}(x-K)e^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\nu^2 T}}dx \\ &= e^{-rT}\int_K^\infty \frac{1}{\sqrt{2\pi T}\nu}e^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\nu^2 T}}dx - e^{-rT}\int_K^\infty \frac{1}{\sqrt{2\pi T}\nu x}Ke^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\nu^2 T}}dx \end{aligned} \tag{2.2.10}$$

The first integral term of the last line is similar to that from equation (2.2.8) and as such we now have

$$e^{-rT}\int_K^\infty \frac{1}{\sqrt{2\pi T}\nu}e^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\nu^2 T}}dx = e^{-rT}S_0 e^{\mu+\frac{\nu^2 T}{2}}\int_B^\infty \frac{1}{\sqrt{2\pi}}e^{\frac{-t^2}{2}}dt \tag{2.2.11}$$

where

$$B = \frac{ln\frac{K}{S_0} - \mu - \nu^2 T}{\nu\sqrt{T}}$$

Since this integral meets the requirements for that of a C.D.F. of a normal random variable and since by equation (2.2.8) we have our mean $\mu = rT$, we now have

$$\begin{aligned} S_0\left(1 - N\left(\frac{ln\frac{K}{S_0} - rT - \nu^2 T}{\nu\sqrt{T}}\right)\right) &= S_0 N\left(-\frac{ln\frac{K}{S_0} - rT - \nu^2 T}{\nu\sqrt{T}}\right) \\ &= S_0 N\left(\frac{ln\frac{K}{S_0} + rT + \nu^2 T}{\nu\sqrt{T}}\right) \end{aligned} \tag{2.2.12}$$

Now let $y = \frac{ln\frac{x}{S_0} - \mu}{\nu\sqrt{T}}$ where $dy = \frac{dx}{x\nu\sqrt{T}}$. Then

$$\begin{aligned}
-e^{-rT}\int_K^\infty \frac{1}{\sqrt{2\pi T}\nu x} K e^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\nu^2 T}}dx &= -e^{-rT}\int_{B+\nu\sqrt{T}}^\infty \frac{1}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}dy \\
&= -e^{-rT}K\left(1-N\left(B+\nu\sqrt{T}\right)\right) \\
&= -Ke^{-rT}N\left(-B-\nu\sqrt{T}\right) \\
&= -Ke^{-rT}N\left(\frac{ln\frac{S_0}{K}+rT-\frac{\nu^2T}{2}}{\nu\sqrt{T}}\right)
\end{aligned} \tag{2.2.13}$$

Hence $e^{-rT}\int_K^\infty \frac{1}{\sqrt{2\pi T}\nu}e^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\nu^2 T}}dx - e^{-rT}\int_K^\infty \frac{1}{\sqrt{2\pi T}\nu x}Ke^{\frac{-(ln\frac{x}{S_0}-\mu)^2}{2\nu^2 T}}dx = S_0N\left(\frac{ln\frac{K}{S_0}+rT+\nu^2T}{\nu T}\right) - Ke^{-rT}N\left(\frac{ln\frac{S_0}{K}+rT-\frac{\nu^2T}{2}}{\nu\sqrt{T}}\right)$.

□

2.3 Issues with Implied Volatility in the Black-Scholes Model

When using the Black-Schloes model for forward options pricing the price of a stock, the time till the expiration of the option, and the risk-free interest rate are all known quantities, however the implied volatility cannot be viewed as such. This is why in the analysis of call options as they relate to the Black-Scholes model the market price of a call option must be used to solve for the implied volatility, just as Macbeth and Merville did in their *An Empirical Examination of the Black-Scholes Call Option Pricing Model*. In their article, Macbeth and Merville found that "on any given day different market prices of options written on the same underlying stock yield different values of ν^2 and that these implied variance rates, for the same option, change through time."[4] This would support the empirical tests conducted by Black and Scholes in their *The Pricing of Options and Corporate Liabilities*, in which they found that "the actual prices at which options are bought and sold are deviate in certain systematic ways from the values predicted by the formula" [5]. We will therefore test the accuracy of implied volatility in the Black-Scholes model using observed options pricing.

In order to test whether or not the implied volatility calculated from the Black-Scholes model is an accurate predictor of actual future volatility, data on call options for the VIX issued on

6/1/16 gathered from the Chicago Board Options Exchange (CBOE) will be tested. The data collected is divided into three different segments based on what date the options expire, although all were issued on 6/1/16. Since there was no difference in the price of the underlying asset, the VIX, between the different call options offered, the stock price S will be held constant. While Lauterbach and Schultz point out that a "possible deficiency in the Black-Scholes model arises from its assumption of a constant default free interest rate,"[6] for the purposes of this test we will hold the interest rate the time the options were quoted and all future expectations of interest rates to be constant. It is worth noting that the data regarding the implied volatility, ν, for the call options quotes is only attainable through the Black-Scholes model because the data for the observed market prices is available.

Since this test concerns an extremely reputable model in the finance world which factors into the decision-making processes of people who are employing it in an attempt to make money, a level of significance of 5% or $\alpha = 0.05$ will be used. The single variable linear regression model we will use to test the claim is as follows:

$$\nu_{BS} = \beta * \nu_{actual} + \epsilon$$

Where ν_{BS} is the implied volatility from the Black-Scholes model and ν_{actual} is the actual volatility for the same period. Specifically, we will be testing the claim that $\beta = 1$.

Since all three of the subsets of data we are testing have below 50 observations we can infer that our estimator, the single variable linear regression model, will have a t-distribution. Since we have three subsets of call option expiration dates within the observations we will run three separate regressions and compare the results. While all three subsets were quoted on 6/1/16, one subset expires on 6/15/16, one on 8/17/16, and one on 10/19/16. We will now assume that $\beta = 1$ and test the claim for the three different subsets.

The actual volatility for the VIX from 6/1/16-6/15/16 was calculated to be $\nu_{actual} = 2.7653$. The regression analysis for the set of implied volatilities, which vary based on strike price, from options expiring on 6/15/16 is as follows:

```
. reg shortimpvol sactualvol, noconstant
```

Source	SS	df	MS
Model	77.6184238	1	77.6184238
Residual	19.2587873	37	.520507765
Total	96.8772111	38	2.54940029

Number of obs	=	38
F(1, 37)	=	149.12
Prob > F	=	0.0000
R-squared	=	0.8012
Adj R-squared	=	0.7958
Root MSE	=	.72146

shortimpvol	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sactualvol	.5168316	.0423234	12.21	0.000	.4310763	.6025868

Figure 2.3.1.

Due to the high t-statistic and corresponding small p-value that is far below our level of significance of $\alpha = 0.05$, we can safely conclude that there is enough evidence to reject our claim that $\beta = 1$ for the first subset of data with an expiration date of 6/15/16.

The actual volatility for the VIX from 6/1/16-8/17/16 was calculated to be $\nu_{actual} = 3.4442$. The regression analysis for the set of implied volatilities, which vary based on strike price, from options expiring on 8/17/16 is as follows:

```
. reg medimpvol mactualvol, noconstant
```

Source	SS	df	MS
Model	27.5379143	1	27.5379143
Residual	1.32878595	32	.041524561
Total	28.8667002	33	.874748491

Number of obs	=	33
F(1, 32)	=	663.17
Prob > F	=	0.0000
R-squared	=	0.9540
Adj R-squared	=	0.9525
Root MSE	=	.20378

medimpvol	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
mactualvol	.2652258	.0102992	25.75	0.000	.244247	.2862045

Again, due to the high t-statistic and corresponding small p-value that is far below our level of significance of $\alpha = 0.05$, we can safely conclude that there is enough evidence to reject our claim that $\beta = 1$ for the second subset of data with an expiration date of 8/17/16.

The actual volatility for the VIX from 6/1/16-10/19/16 was calculated to be $\nu_{actual} = 2.8781$. The regression analysis for the set of implied volatilities, which vary based on strike price, from options expiring on 10/19/16 is as follows:

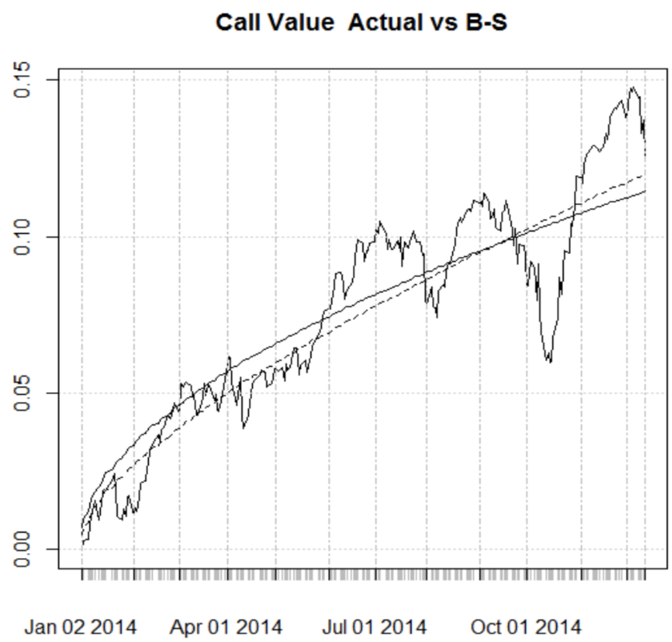
```
. reg longimpvol lactualvol, noconstant

      Source |       SS       df       MS              Number of obs =      33
-------------+------------------------------           F(  1,    32) =  903.28
       Model |  19.7101093     1  19.7101093           Prob > F      =  0.0000
    Residual |  .69826112     32  .02182066            R-squared     =  0.9658
-------------+------------------------------           Adj R-squared =  0.9647
       Total |  20.4083705    33  .618435469           Root MSE      =  .14772

------------------------------------------------------------------------------
  longimpvol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  lactualvol |   .2685186   .0089344    30.05   0.000     .2503199    .2867174
------------------------------------------------------------------------------
```

Testing our final subset gave us yet another high t-statistic and corresponding small p-value that is far below our level of significance of $\alpha = 0.05$, we can safely conclude that there is enough evidence to reject our claim that $\beta = 1$ for the third subset of data with an expiration date of 10/19/16.

Since tests conducted on all three subsets lead us to reject the same claim for all three, it is reasonable to conclude that the implied volatility calculated from the Black-Scholes model is not an accurate predictor of actual volatility for the underlying security for the period spanning the quote and expiration date for that option.

We see additional evidence of this in the analysis shown earlier in the chapter. In the analysis of call options pricing of 480 equities over a 240 day trading period, the graph of the mean log returns for the forward call option prices calculated using the Black-Scholes model for 2 different values of risk-free interest rate and the mean actual price for call options for the stocks as observed in the market is shown as

The researcher in this analysis points out that the differences in variances between the two calculations using the Black-Scholes model do pose an issue in practical application of the Black-Scholes model. He acknowledges that the annual volatilities used in the Black-Scholes calculations



were what he calls "peek-into-the-future" volatilities that were calculated from observed price history which was "a misrepresentation of the market, just to match the call option price."[10]

3
Allan Variance

3.1 What is Allan Variance?

In 1945, a physics professor at Columbia University named Isidor Rabi first posited that "a clock could be made from a technique he developed in the 1930's called atomic beam magnetic resonance."[8] Five years later the National Institute of Standards and Technology(NIST), then known as the National Bureau of Standards(NBS), completed development of the world's first atomic clock, measuring oscillations in an ammonia molecule. As the technology continued to progress in the 1950s, more thorough analysis of the atomic clocks was conducted with both academic and commercial goals. During this exploration it was discovered that both atomic clocks, and crystal oscillators "did not have a phase noise consisting only of white noise, but also of white frequency noise and flicker frequency noise."[9] This posed a major issue in any analysis of oscillations in atomic clocks as the statistical tools of the time were unable to adequately capture this noise. The tools of the time were inadequate to the degree that "traditional statistical tools such as standard deviation as the estimator will not converge."[9]

In February 1966, physicist David W. Allan published *Statistics of Atomic Frequency Standards*, in which he outlined methodology he had developed in order to better estimate frequency stability in oscillators, particularly oscillations atomic clocks. In his own words, he was able to show "a relationship between the expectation value of the standard deviation of the frequency fluctuations for any finite number of data samples and the infinite time average value of the standard deviation, which provides an invariant measure of an important quality factor of a frequency standard."[7] In seeking to solve the issue of estimating white frequency noise and flicker frequency noise in atomic clocks, Allan developed the general M-sample variance, which led to the more commonly used 2-sample variance. What is known as the Allan Variance, also known as AVAR, is the special 2-sample case of the general M-sample variance The formula for the general M-sample variance is given by:

$$\sigma_y^2(M,T,\tau) = \frac{1}{M-1}\left\{\sum_{i=0}^{M-1}\left[\frac{x(iT+\tau)-x(iT)}{\tau}\right]^2 - \frac{1}{M}\left[\sum_{i=0}^{M-1}\frac{x(iT+\tau)-x(iT)}{\tau}\right]^2\right\} \quad (3.1.1)$$

where $x(t)$ is the phase angle, M is the number of frequency samples used in the variance, T is the time between each frequency sample, and τ is the time length of each frequency estimate. The Allan Variance case is then given by:

$$\sigma_y^2(\tau) = \mathbf{E}[\sigma_y^2(2,\tau,\tau)] \quad (3.1.2)$$

The 2-sample, Allan Variance gives "a means to meaningfully separate many noise-forms for time-series of phase or frequency measurements between two or more oscillators." [9] In more simple terms, Allan Variance captures variation between points on time series to better predict frequency oscillations. Applying Allan Variance to a simple time series, with τ being equal to 3 lengths of τ_0 would look like

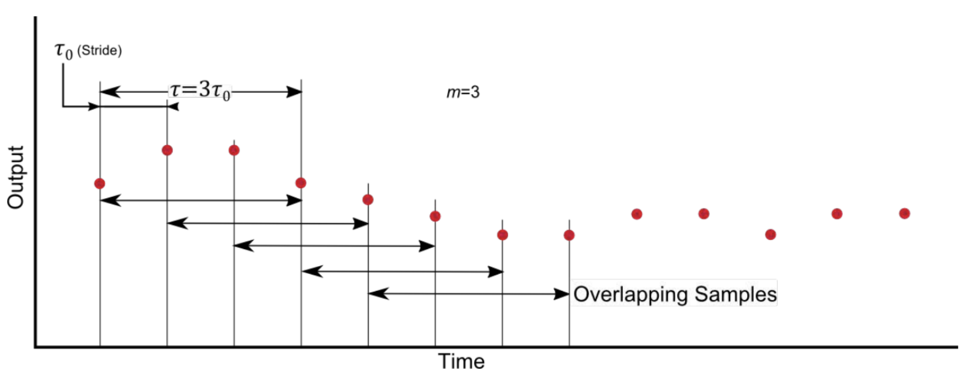


Figure 3.1.1.

where you are analyzing variation between units of time of length $3\tau_0$.

Currently, Allan Variance is used extensively across a number of fields since "it is a deterministic robust, quantity to analyze non-stationary time series for characterizing their variability at different observation intervals." [11]

3.2 Financial Analysis Using Allan Variance

In attempting to better measure annual volatility in equities, the relationship between Allan Variance and taking standard variance will be explored. Incorporating Allan Variance will perhaps give us a better estimate of variance in stock pricing as it relates to modeling "noise" observed in inter-day fluctuations observed in the market. Allan Variance should be useful in this regard since "the mathematical formalism that defines the AVAR is applicable top any other time series since no *a priori* assumptions are made regarding the origin and nature of the input time series...the AVAR analysis is focused on the variance of the fluctuations at different observation intervals." [11]

Beginning the exploration into the relationship Allan Variance plays in describing financial movements, two groups of six stocks were analyzed first. The groups were split between stocks of large cap technology companies and those of large cap financial companies. The tickers for the financial companies include JPM, BBT, BAC, BLK, GS, COF, and the tickers for the technology companies include AMZN, FB, EA, AAPL, AMD, ATVI. The variance, Allan Variance, and the proportion of Allan Variance over variance was calculated for each of the stocks in the two groups. The first test that was conducted was a single factor ANOVA between the Allan Variances of the two groups of stocks.

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	6	38.1609148	6.36015247	119.062793		
Column 2	6	279.196303	46.5327171	11675.1265		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4841.50485	1	4841.50485	0.82099833	0.38620009	4.96460274
Within Groups	58970.9467	10	5897.09467			
Total	63812.4515	11				

Figure 3.2.1.

With a p-value of just under 0.4, we cannot confidently conclude that Allan Variance alone can be used in accurately describing fluctuations in stock prices.

A second test was then conducted with the same data set in which another single factor ANOVA was run, but this time looking between the proportion of Allan Variance over actual variance for each of the stocks in the two groups.

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	6	3.82345566	0.63724261	0.19091512		
Column 2	6	1.48497649	0.24749608	0.00940353		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.45570707	1	0.45570707	4.54982162	0.05872063	4.96460274
Within Groups	1.00159327	10	0.10015933			
Total	1.45730034	11				

Figure 3.2.2.

Running an ANOVA on the proportion of the two variances gives us a much lower p-value of around 0.05 which, depending on the confidence interval, gives us a strong reason to conclude that perhaps there is some relationship that should be further explored.

As such the two groups of six financial and technology industry stocks were extrapolated out to twenty five stocks in each group. The tickers for the new group of financial stocks now include JPM, BBT, BAC, BLK, GS, COF, C, AMG, ALL, AXP, AMP, AON, AJG, AIZ, BK, CBOE, SCHW, CB, CINF, CME, CMA, DFS, ETFC, FITB, and FRC. The tickers for the new group of technology stocks include AMZN, FB, EA, AAPL, AMD, ATVI, AKAM, ADS, APH, AMAT, ADSK, ADP, AVGO, BR, CDNS, CSCO, CTXS, CTSH, FFIV, FIS, FLT, FLIR, FTNT, IT, and INTC. A single factor ANOVA was run again on the proportion of Allan Variance over actual variance for the stocks in the two groups.

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	25	13.249521	0.52998084	0.10447756		
Column 2	25	11.8301343	0.47320537	0.22085233		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.04029317	1	0.04029317	0.24770656	0.62096713	4.04265213
Within Groups	7.80791742	48	0.16266495			
Total	7.84821059	49				

Figure 3.2.3.

With our worst p-value so far of over 0.6, we cannot safely conclude from this test that there is variance between the groups that can be accurately described using the proportion of Allan Variance to actual variance. This is also easy to see graphically by plotting the Allan Variance versus the proportion of Allan Variance over Actual Variance.

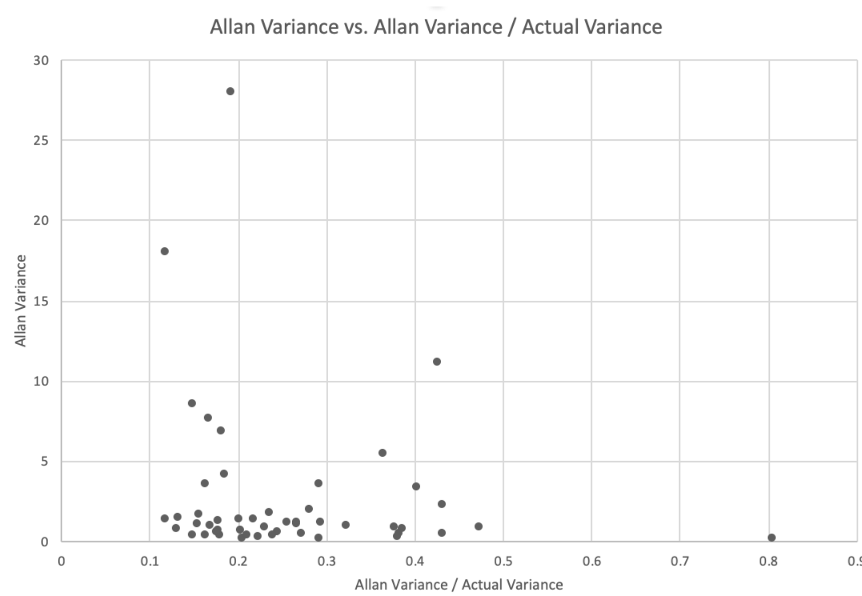


Figure 3.2.4.

Clearly the spread is not one that can be modeled easily, especially with a simple linear regression.

Despite the inconclusive tests that were run in this project, there is still reason to explore this relationship further. The second ANOVA that was conducted yielding promising results, in addition to the fact that a limited number of tests using a limited data set were run, are both potential signs that there could in fact be more to Allan Variance's use in financial predictions that warrant further exploration. While movements in financial markets in general are extremely difficult to predict in general, the underlying math that drives Allan Variance was developed in attempting to solve similarly difficult problems that deal with similar issues when viewed with a purely mathematical perspective.

Bibliography

[1] Carmen Reinhart and Kenneth Rogoff, *Is the 2007 U.S. Sub-Prime Financial Crisis So Different? An International Historical Comparison* (2008).

[2] Nassim Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House, New York, NY, 2007.

[3] Lars Nielsen, *Understanding N(d1) and N(d2): Risk-Adjusted Probabilities in the Black-Scholes Model*, INSEAD (1992).

[4] Larry Merville and James Macbeth, *An Empirical Examination of the Black-Scholes Call Option Pricing Model*, The Journal of Finance **34** (1979), 1173–1186, available at `www.jstor.org/stable/2327242`.

[5] Fischer Black and Myron Scholes, *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy **81** (1973), 637–654, available at `www.jstor.org/stable/1831029`.

[6] Beni Lauterbach and Paul Schultz, *Pricing Warrants: An Empirical Study of the Black-Scholes Model and Its Altenatives*, The Journal of Finance **45** (1990), 1181–1209, available at `www.jstor.org/stable/2328720`.

[7] David Allan, *Statistics of Atomic Frequency Standards*, IEEE **54** (1966), 221–230, available at `https://ieeexplore.ieee.org/document/1446564`.

[8] *A Brief History of Atomic Clocks at NIST* (2010), available at `https://www.nist.gov/pml/time-and-frequency-division/time-services/brief-history-atomic-clocks-nist`.

[9] *Allan Variance*, Iowa State Engineering, available at `http://home.engineering.iastate.edu/~shermanp/AERE432/lectures/Rate%20Gyros/Allan%20variance.pdf`.

[10] Matthew Krznaric, *Comparison of Option Price from Black-Scholes Model to Actual Values*, IdeaExchange@UAkron (2016), available at `https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1410&context=honors_research_projects`.

[11] Ricardo Hernadez-Perez, *Allan Deviation Analysis of Financial Return Series*, Elsevier (2012).