

Volatility Forecasting and Interpolation

Volatility Forecasting and Interpolation

Levi Turner

May 3, 2016

Advised by David Finnoff

Abstract

Forecasting volatility is important to financial asset pricing because a more accurate forecast will allow for a more accurate model to price financial assets. Currently the VIX is used as a measure of volatility in the market as a whole, but a major issue with this is that it is calculated based on manually traded options on the S&P 500. Another method of forecasting volatility is that of solving for volatility from the Black-Scholes model in option pricing, but this method is not consistent across prices; for different strike prices, a different volatility will be found, creating what is known as a volatility smile. I will develop a method which calculates a similar measure of volatility to the Black-Scholes method and the VIX, but using electronically traded options on the SPY ETF which tracks the S&P 500. I will also be incorporating the mathematical model developed by Britten-Jones and Neuberger in their 2000 paper, which is another variation from the method in which the VIX is calculated. The method developed will provide a smoother and more accurate forecast of volatility over any given time frame, with a 30 day forecast being the industry norm. The method will also have the ability to forecast volatilities for individual assets, not simply the whole market.

Volatility plays an important role in financial markets, as an indicator of returns and the magnitude of possible changes in prices for securities. Volatility is defined in this paper as an annualized standard deviation of the percentage change in prices; the standard deviation must be annualized in order to compare the different methods. It is inaccurate to compare the standard deviation of daily percentage price changes to the standard deviation of monthly percentage price changes. One can see historical examples where fast, large changes in volatility were accompanied by a financial crisis; a recent example is the crisis of 2008, where the VIX peaked at well above average at 50. Even though volatility is a basic parameter of instruments, such as expected return (the amount of return seen by the security on average over a given time frame) and the beta associated with a stock (a measure of risk when running a regression of returns against a market index), there is still quite a bit of work involved in forecasting and predicting volatilities in the performance of financial securities. One of the methods used is to evaluate the variance in the percentage changes of prices over a previous period of time, but there is no guarantee that future periods will abide by this same parameter estimate. Another method, which is preferable to the backwards looking estimate from changes in the price is to use an ARCH or GARCH model to forecast future volatility. A major detraction from this method is that the future is outside of our data set, one would need to extrapolate into new periods, which can be dangerous under the best of circumstances.

One of the major uses of volatility forecasting is the ability to accurately and precisely value an option. An option is a financial derivative which provides the holder the right to purchase or sell the underlying at a specific price. For example, if an investor were to purchase a call option on stock XYZ with a strike price of \$15, then he could purchase the stock for \$15

even if the stock price is above that value at let's say \$25; in the event the price of the stock falls below \$15, the holder of the call option is under no obligation to pay \$15 if the stock is below this value, at let's say \$10. The process is similar for a put option, except that with a put option, the right of the holder is to sell, not to buy at a given price. In essence, a call option is a bet on the security going up beyond a certain threshold, while a put option is a bet on the price of the security falling below a threshold. A major difference between simply going long a stock or short a stock and purchasing a call or put is that options have a time frame before they expire and become worthless. Due to this, pricing an option can be mathematically difficult and the price determined by various formulas and methods can be extremely sensitive to the projected volatility of the underlying security over the life span of the option.

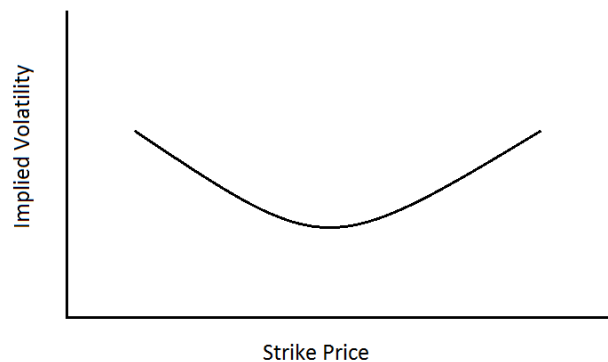
With the introduction of the Black-Scholes model, one can look at the price of an option on a security and use computer software to solve for the implied volatility of the option given all of its other parameters which are known. In traditional Black-Scholes option pricing, the volatility is considered a known, and the final price is calculated using their famous model. Using numerical methods, one can work from the price of an option, which is known from financial exchange data and is readily available, backwards to the volatility which traders are placing on the underlying security. This is called the implied volatility of the underlying security, which is often used as an indicator of future volatility moving forward throughout the option's life span. This method has a few issues from a forecasting perspective. First, the volatility is not known or calculated and then incorporated into the Black-Scholes model to arrive at an option price, but rather numerical methods are used to arrive at what the volatility would need to be in order to arrive at the current price of the option. Second, an instrument will have a

single volatility over a given time frame, but volatility smiles do exist, such as in the table below for options of the S&P 500 ETF. A volatility smile is when options at different strike prices will yield differing implied volatilities of the underlying security over the same time period, which is impossible. In the example detailed above, the value of \$10 is the strike price for our option.

Date of Expiration	Strike Price	Implied Volatility
May 6, 2016	\$180.00	44.95%
May 6, 2016	\$190.00	21.53%
May 6, 2016	\$200.00	16.36%
May 6, 2016	\$204.00	14.34%
May 6, 2016	\$204.50	14.07%
May 6, 2016	\$214.00	10.16%
May 6, 2016	\$225.00	13.58%

All data taken from Yahoo Finance on Apr. 5, 2016

Graphically, this looks like the following:



From our sample of seven strike prices from a single date of expiration, we can already see a large disparity between some of these volatilities. Intuition tells us this is a major problem; arguments have been made that one should use those options close to the stock price, such as using the implied volatility of an option with strike price of \$50 when the underlying stock is trading at \$50.23, but that does not utilize all the information provided by the option chain. Another issue with using the implied volatility is that forecasting implies a change in the volatility, but the Black-Scholes model assumes constant volatility in its formula.

This issue is addressed by the 2000 paper “Option Prices, Implied Price Processes, and Stochastic Volatility” by Mark Britten-Jones and Anthony Neuberger in The Journal of Finance. In their paper, they introduce a model which does not use any previous information regarding volatility and is consistent with all available option prices. Their first major result is the following equation, but before that some variables will be defined. S_t is the price of a stock price at a time t , where time t is the time until expiration of an option on that stock. dS_t is the incremental change in the random stock price from time t_1 to time t_2 , which is required to complete the integral on the left hand side. K is the strike price of an option, such as \$10 or \$50 from the examples above. C is the price of a call option, put options will not be considered in this paper; the price C of a call option is a function of the time to expiration and the strike price at which it is issued. dK is the incremental change in the value of the strike price from $K=0$ to $K=\infty$, which completes the integral on the right hand side.

$$E_0 \left[\int_{t_1}^{t_2} \left(\frac{dS_t}{S_t} \right)^2 \right] = 2 \int_0^\infty \frac{C(t_2, K) - C(t_1, K)}{K^2} dK.$$

What this says, is that the expected value of the volatility between two points in time t_1 and t_2 is defined by the integral of the left hand side. A simplified model when looking at volatility from current time 0 towards a future time t_2 is the equation following this paragraph. The simplification can be seen clearly by knowing that $C(0, K) = \max(S_0 - K, 0)$. This equality is a result of option pricing which says that when there is no time remaining in the life span of the option, the price of the call option is either 0 because the stock price is below the strike price, or the price of the call option is the difference between the current stock price and the strike price due to the stock price being above the strike price.

$$E_0 \left[\int_0^{t_2} \left(\frac{dS_t}{S_t} \right)^2 \right] = 2 \int_0^\infty \frac{C(t_2, K) - \max(S_0 - K, 0)}{K^2} dK,$$

A nontrivial task is seeing what role $\left(\frac{dS_t}{S_t}\right)^2$ is playing in the above equation, in other words: seeing how $\left(\frac{dS_t}{S_t}\right)^2$ works out to be a volatility measure. The algebraic steps are as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB$$

This first step is an equality coming from Ito's Lemma and its application in the Black-Scholes paper "The Pricing of Options and Corporate Liabilities" in 1973. Here, μ is the expected return of the stock S_t over time, this is sometimes referred to as the drift and σ is the standard deviation of this movement.

$$\left(\frac{dS_t}{S_t}\right)^2 = (\mu dt + \sigma dB)^2$$

$$\left(\frac{dS_t}{S_t}\right)^2 = \mu^2 dt^2 + 2\mu\sigma dt dB + \sigma^2 dB^2$$

$$dt^2 = 0 \text{ and } dt dB = 0 \text{ and } dB^2 = dt$$

The previous step is another result from stochastic calculus where dt is the change in the determined variable and dB is the random variation. A determined variable is a standard variable, where it changes with a known pattern and any value can be found at any moment. For a variable to follow random changes, such as dB above, the changes are wholly unknown and cannot be found for any given moment not in the past. Simplifying with the zeroes inputted, we achieve:

$$\left(\frac{dS_t}{S_t}\right)^2 = \sigma^2 dt$$

$$E_0 \left[\int_0^{t_2} \sigma^2 dt \right] = 2 \int_0^\infty \frac{C(t_2, K) - \max(S_0 - K, 0)}{K^2} dK$$

One now sees the familiar σ^2 which is the symbol used to denote squared volatility, where squared volatility is the square of the standard deviation σ .

Two issues present themselves from the Britten-Jones model for volatility; first, the integral assumes constant strike prices, but the set of available strike prices are not continuous. The strike prices are a relatively small subset of discrete values, some form of interpolation must be used to create a function which can be integrated. For the purposes of this project, a straight line interpolation was used; this created a piecewise function which can be integrated across. For the strike prices contained within the bounds provided, but something needed to be done regarding those from zero to the first strike price and from the last strike price to infinity. For the low end, a point which was at $(0, S)$ where 0 is the theoretical strike price and S is the value of the stock at the time use. The high end of strike prices proved to be much more difficult as the value needs to tend to 0 as the strike price tends towards infinity. This ruled out a continuous linear function to append to the last of the strike prices in the set; after a few trials, it was found to be easiest to set a value of zero from the last strike price to infinity. This is not theoretically correct, as there will be option prices for strike prices beyond the set, but the solution was minimally affected much by this simplification.

The second issue presented in the Britten-Jones model is that there are not option prices for every time t that one may desire to forecast volatilities. This problem can be solved

in a similar manner to the strike price issue, the option price for a time t_i which does not have explicit information can be interpolated by using points in time which are provided. Similar to the strike price interpolation, a straight line interpolation can be used for the times; in this project, that ended up not being a major issue as the object of interest was a forecast for thirty day volatility. There is data for May 6, 2016 options, which is 30 days after the date at which the data was acquired for these options.

A third, but relatively minor issue from a computational standpoint presents itself in the above Britten-Jones model. As mentioned at the beginning of this paper, the annualized standard deviation is the object of interest in this forecast, but the above model provides a forecast for the squared volatility over a period of time from time zero to future time t_2 . Annualizing this value is not difficult, the issue arises from Jensen's inequality, which is the following:

$$E_0 \left[\sqrt{\int_0^{t_2} \sigma^2 dt} \right] \leq \sqrt{2 \int_0^\infty \frac{C(t_2, K) - \max(S_0 - K, 0)}{K^2} dK}$$

If the above held as strictly equal, then a forecast of volatility could be found and not squared volatility; from a theoretical standpoint, there is no reason to assume this a is a strictly equal statement, but for the purposes of this paper, it will be considered to be so for the sake of interpretability of the computer program output.

Linear interpolations were used in this project, but some more advanced and exotic interpolations can be used. Cubic splines are one such option to obtain different results from the straight line interpolation. Another possibility is to find an interpolation which matches the

theoretical shape that option prices are determined to follow with respect to strike prices and expiration timelines.

The price C of an option was determined to be the average of the bid and ask prices due to the fact that the last price did not always reflect the full information in the market. Once the interpolation was completed, the above integral was split into two parts as follows:

$$2\left(\int_0^{\infty} \frac{C(t_2, K)}{K^2} dK - \int_0^{\infty} \frac{\max(S_0 - K, 0)}{K^2} dK\right)$$

This is done in part to integrate the $\max(S_0 - K, 0)$ function; the integral containing this function can be simplified down to:

$$\int_0^{\infty} \frac{\max(S_0 - K, 0)}{K^2} dK = \int_0^{S_0} \frac{S_0 - K}{K^2} dK$$

Resulting in:

$$2\left(\int_0^{\infty} \frac{C(t_2, K)}{K^2} dK - \int_0^{S_0} \frac{S_0 - K}{K^2} dK\right)$$

This can be performed because after S_0 , the value of the function will be zero.

Performing this simplification greatly improves the programming capabilities for solving for the volatility. The integral containing $C(t_2, K)$ is relatively simplistic as the functional form for this interpolation is linear, which is easily handled by a computer.

I chose to look at the firms Walmart and International Business Machines, as well as the stock market as a whole for this project. Walmart was chosen due to its standing at the top of the Fortune 500; IBM was chosen due to the fact that it is also a member of the Fortune 500, but IBM is in a different industry from Walmart. Walmart is a general retailer, while IBM is a technology company; the two of these provide glimpses into different aspects of the market.

The evaluation of the ETF on the S&P 500 provides a forecast for the volatility of the entire market, similar to the VIX. The choice of the companies is not important, any company with options chain data on Yahoo Finance can be evaluated using the method and program developed. Using my program, the following volatilities were calculated:

Walmart	International Business Machines	SPY (ETF)
14.51%	19.31%	15.36%

All data taken from Yahoo Finance on Apr. 5, 2016

We can compare these results to other volatilities readily available online; although not a perfect comparison, the forecasted volatilities for Walmart and IBM will be compared to the implied volatility of the closest at-the-money option in their respective chains. For the volatility calculated for SPY, we can compare to the VIX, which is a 30 day forecast of the volatility in the S&P 500. For various reasons, the calculated values should not be exactly those to which they are compared. For the firms, as mentioned earlier, the implied volatility comes from the Black-Scholes, whereas the calculated volatility comes from the prices of all options in the chain. For the SPY ETF, there are three reasons why the values will differ from the VIX; first, the VIX is a forecast of volatility on the S&P 500, not the SPY ETF which tracks the S&P 500. Secondly, the calculation of the VIX is not using the same method and model which was used in the programming here. And thirdly, the VIX is calculated off of manually traded options, while the program here uses electronically traded options; the options trade in different mediums and exchanges.

Starting with Walmart, we see that the calculated value is 14.51%, while the implied volatility of the option closest to being at-the-money is 20.31%. Here, there is a difference in volatilities, with the calculated volatility forecast being 5.8% lower than the implied volatility.

For IBM, we have a volatility forecast of 19.31%, but an implied volatility of 26.38%; in this case, our forecast differs from the implied volatility by more than the Walmart scenario, with the calculated IBM value being 7.07% lower than the implied volatility. Finally, for the SPY, we have a forecast value of 15.36%, while the VIX is similar at 15.42% at the time of this paper and the SPYIX is at 15.74%. The SPYIX is a brand new method of forecasting volatility launched by Bats Global Markets in the beginning of April. The major difference between the VIX and SPYIX is that the SPYIX derives its forecast from electronically traded options on the SPY ETF, and not manually traded options on the S&P 500 in the case of the VIX. Due to the extreme newness of the SPYIX, I will be continue comparing the calculated values of the SPY against the value of the VIX, which is still the industry norm. Here, we see the calculated value being 4.24% higher than that of the VIX.

The accuracy of this forecasting model can be greatly improved by using non-linear interpolations, as mentioned, or by using a much harder method of extracting values of C when K is not included in the data set. The accuracy would increase towards a better forecast due to the function being integrated tending towards a more theoretically correct shape and model. This is the direction which I see this project progressing: one can use a closed form function for C , such as Black-Scholes, and hold all variables constant except for K . From this, we can integrate across all K from zero to infinity using the model above. This method is much more complex and much more computer intensive, using numerical integration and many steps, but this process may yield even more accurate forecasts of volatility over a given timeframe from present day 0 to expiration date t_2 . The new model would look akin to this:

$$E_0 \left[\int_0^{t_2} \left(\frac{dS_t}{S_t} \right)^2 \right] = 2 \left(\int_0^\infty \frac{S e^{-\delta t_2} N(d_1) - K e^{-r t_2} N(d_2)}{K^2} dK - \int_0^{S_0} \frac{S_0 - K}{K^2} dK \right)$$

$$d_1 = \frac{\ln \left(\frac{S}{K} \right) + \left(r - \delta + \frac{\sigma^2}{2} \right) t_2}{\sigma \sqrt{t_2}}$$

$$d_2 = d_1 - \sigma \sqrt{t_2}$$

The $N(x)$ function is the standard normal distribution function. The above equation would have a σ on the right hand, but as shown earlier, so does the left hand side. With all other variables held constant and integrated over K , this is one equation with one unknown. The above equation would require advanced numerical methods, quite a bit of computer time, and a powerful computer set up to run the calculations. This was originally a direction intended for this paper, but it was soon apparent that this was beyond the immediate capabilities to code and probably beyond the current numerical methods used in this paper. It is still not a certainty that this method would work for forecasting volatilities; interpolation may be the most effective and efficient method. In addition to the two firms and market wide ETF analyzed above, the major sector ETFs were also analyzed, with results below.

	Forecasted Volatility
XLY	14.71%
XLP	14.34%
XLE	14.58%
XLF	13.66%
XLV	14.57%
XLI	14.43%
XLK	14.22%
XLU	14.32%

All data taken from Yahoo Finance on Apr. 5, 2016

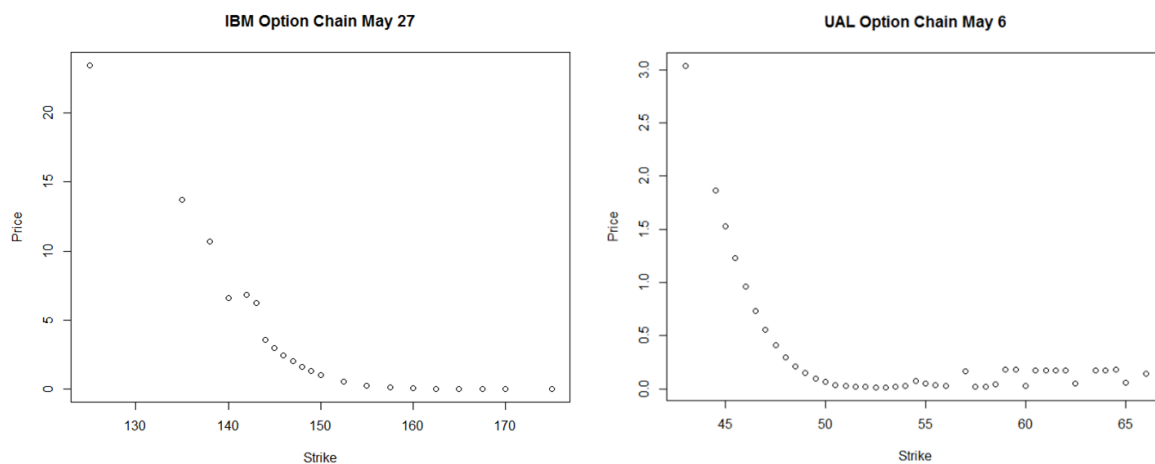
One interesting thing to point out from this table is that even though the model forecasts a volatility of 15.36% for the market wide SPY ETF, all of the major sector ETFs are below this value by at least 0.65%. For your curiosity, the sector ETF information is summarized in a table at the end of this paper. They are not discussed in depth, but they have a forecasted and reality value associated with them.

As of close of business May 3, 2016 the volatility in the percentage of the price changes of WMT is 17.56%, for IBM it is 25.45%, and for the SPY it is 10.37%. It needs to be noted that there are still three trading days of data to come before the volatilities will be comparable; the forecast ends May 6, 2016 whereas the historical data only runs until May 3, 2016 currently. This may sound trivial, but given that there were only 23 trading days between the initial and terminal dates, these three days comprise 13% of the overall set of available days. Examining the Britten-Jones forecast of 14.51% for WMT and the reality of 17.56%, we can see that the Britten-Jones method missed the mark here, but so did the other measure of the implied volatility from the at the money option. With a value of 20.31%, the implied volatility falls above reality, and we see that the reality is straddled by the implied volatility and the Britten-Jones method in the WMT case. For IBM, the predicted value is 19.31%, while the implied volatility was 26.38%; the reality is that the volatility has been 25.45%, so in this case the implied volatility was much closer than the Britten-Jones prediction.

Things get a little more interesting when comparing the various SPY predictions to the reality. All of the predictions, those from VIX, SPYIX, implied volatility, and the Britten-Jones prediction, hovered at approximately 15%. Due to the consistency of these values, there was an expectation for the reality to be approximately this value, but that is not the case. As of May

3, 2016, the volatility has been only 10.36% since April 5th. This was the largest surprise among these three assets due to the consistency of the various volatility forecasts.

One of the largest problems with the Britten-Jones method of forecasting volatilities is that when volumes are low on the option chain, errors can arise within the shape of the curve, which then can transfer into errors in the equation being integrated. An example of such errors can be seen in the charts below:



In both the IBM option chain and the UAL option chain, we can see numerous points which do not conform to the general downward motion of the curve. These are simply illustrations of the errors that occur, UAL was used here due to the number of errors in the tail of the data. There can also be errors where the data points are much lower than where they should be, which present similar problems.

From a single method forecast perspective, my recommendation is to use the Britten-Jones method which I have developed above instead of the VIX, SPYIX, or implied volatility for the following reasons. First, the Britten-Jones method is independent of all historical information; it does not rely on any historical performances to forecast the volatility. Second,

the Britten-Jones method is a method which is considered asset specific; this method can be used to forecast a volatility on any asset which has options trading for it. Lastly, the Britten-Jones method is consistent across all option strike prices; in other words, this model uses all of the information provided by the market in order to arrive at its forecast. There are not differing volatility forecasts for different strike prices.

This is not to say that the Britten-Jones method is perfect, it has issues, as noted above. With more research into the area surrounding volatilities, a multi forecast method could be created to forecast volatilities. The first to be examined will be a dual-method forecast where the Britten-Jones method is combined in some fashion with the implied volatility from an at the money option. This could potentially be more robust as it would incorporate the highly quantitative forecast from Britten-Jones and the less quantitative forecast from the implied volatility. Alternatively, the Britten-Jones model could be expanded to include put options, which it does not currently. This could potentially be done using put-call parity, but an issue arises when the strike price is infinity, because a put option does not tend to zero as a call option does.

In conclusion, the Britten-Jones model can be used to forecast volatilities over any specified time frame, is independent of all historical observations, and is consistent with the information provided within the option chain for a security. Some choices must be made in the form of interpolation across strike prices and times, but it appears that a simple linear interpolation works well given my current computational constraints. In addition, it should be noted that the volatilities above are a forecast, which means that they are not deterministic. The random fluctuations in the market could cause the forecasts to be wildly off from the

realized volatilities once data is available. This can easily be seen by the comparisons of the forecasted values to reality in this paper. As mentioned, the linear interpolation is not the only one available, another choice could be made. The linear interpolation allows for an understanding of the model and a springboard into more complicated functional forms for forecasting volatilities in the future.

Appendix of Tables

For your ease of comparison, the methods and values are summarized in the below tables.

	SPY (ETF)	WMT	IBM
Britten-Jones	15.36%	14.51%	19.31%
Implied Volatility	14.34%	20.31%	26.38%
VIX	15.42%		
SPYIX	15.74%		
Reality	10.36%	17.56%	25.45%

	Forecasted Volatility	Reality
XLY	14.71%	11.79%
XLP	14.34%	12.25%
XLE	14.58%	23.37%
XLF	13.66%	15.83%
XLV	14.57%	14.32%
XLI	14.43%	10.34%
XLK	14.22%	12.36%
XLU	14.32%	14.78%

References

Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (1973): 637-54. Web. 8 Apr. 2016

Britten-Jones, Mark, and Anthony Neuberger. "Option Prices, Implied Price Processes, and Stochastic Volatility." *The Journal of Finance* 55.2 (2000): 839-66. Print.

"The CBOE Volatility Index-VIX". Chicago Board Options Exchange (2014). White Paper.

Figlewski, Stephen. "Forecasting Volatility". New York University Stern School of Business (2004). Online.

Turner, Matt. "A New 'Fear Index' is Coming to Wall Street." *Business Insider*. 8 Mar. 2016. Web. 5 Apr. 2016. This article speaks about the launch of the SPYIX.

RStudio was used to manipulate the data and perform calculations, the code is available upon request. All data was taken from Yahoo Finance.