

Variance in Volatility: A foray into the analysis of the VIX and the Standard and Poor's 500's Realized Volatility

Arthur Kim*
Duke University

April 24, 2013

Abstract

This study finds that the AR models map the VIX and Realized Volatility time series' better than MA models do, and find the lags of greatest correlation between the two time series' to be between 11 and 16 days, with a correlation coefficient of approximately 0.54.

Keywords: VIX, Realized Volatility

*Arthur Kim is a senior Economics and Statistics double major at Duke University with his B.S. expected in May 2013

1 Introduction

While the volatility of a stock or an index has been well documented and followed, the variance/volatility of the volatility itself is not followed as rigorously. In general, the variance of the volatility helps predict the accuracy with which we can predict returns on a portfolio. If we are able to predict the portfolios returns more accurately, it gives us an opportunity to be able to invest more intelligently. In a more applicable sense, a volatility variance model is important because it gives us a more immediate chance to make monetary gains by more accurately modeling and predicting the movement of the VIX. It is possible to bet on the VIX, and those who can predict the variance of the Standard and Poor's 500 more accurately will have a chance to increase their wealth by buying or selling VIX options through the CBOE. When trading VIX futures, one is betting on the discrepancy between what the market expects the value to be and what you think the value will be in a month. Essentially, you are trying to gauge market angst and make a move on it. The VIX moves fairly quickly, so there is a lot of potential to gain (and also to lose) wealth. I will seek to model the variance between to the VIX and the realized volatility values of the Standard and Poor's 500 in an effort to better model the VIX.

In a greater sense, should people become more knowledgeable about the true volatility values of the variance, we will be able to predict markets better, and hopefully, in the future, increase the efficiency of the markets as a whole. Market efficiency and its calculation will not be a large part of this paper, but it is an area that could be improved by the results.

I will calculate the values of the realized volatility for the Standard and Poor's 500 and model the difference between the models with a number of different possible regressions, including autoregressive, moving averages, and autoregressive moving average models. My goal is twofold. Firstly, to find what models best map the difference in the variance of volatility, and to see which lags of the times series' are best correlated with each other in order to better forecast the VIX using the Standard and Poor's 500.

In section 2, the terms and equations necessary for the rest of the paper will be defined and explained. In section 3, we will look at the data itself, and in section 4, we will technically analyze it and produce figures and graphs to better explain it. Finally in section 5, we will discuss the significance of the correlations and in section 6, discuss the conclusions and further research.

2 Equations and Definitions

2.1 Volatility - Realized and Implicit

Realized Volatility is volatility calculated from the average square of the log returns of prices.

$$RV_{t+1}^m = (1/T) \sum_{j=1}^T R_{t+j}^2 \quad (1)$$

where R is the log returns of the prices.

$$R_{t+1} = \log(x_{t+1}/x_t) \quad (2)$$

Implied volatility is, under the Black-Scholes formula, the market's estimate of the constant volatility parameter (Mayhew, 1995). It is calculated from option pricing formulas by inverting the formula and calculating for the volatility. By doing this, one is able to calculate, by using the market price and the other data points, what the market assesses to be the volatility of the asset.

Looking at the Black-Scholes formula, we can see how volatility can be implicitly calculated. In the Black-Scholes equation,

$$w_2 = rw - r x w_1 - (1/2)v^2 x^2 w_{11} \quad (3)$$

where r is the annual risk free (continuously compounded) rate, w is the price of the derivative, v is the standard deviation, x is the price of the stock, and the subscripts denote the variable with respect to which it has been partially differentiated (x being 1 and t (time) being 2). So, for example, w_2 means price of the derivative partially differentiated with respect to time, w_1 means price of the derivative partially differentiated with respect to the price of the derivative, and w_{11} means price of the derivative partially differentiated twice with respect to the price of the derivative. The Black-Scholes equation is usually written as:

$$(\partial V / \partial t) + 1/2 \sigma^2 S^2 (\partial^2 V / \partial S^2) + r S (\partial V / \partial S) - r V = 0 \quad (4)$$

In this case, the V is the price of the derivative, S is the price of the stock, r is the annualised risk-free interest rate (continuously compounded), and σ is the volatility of a stock's returns. As we can see, if we have all the other values from this equation (which can be data collected from the market), we can calculate the variance (and thus, the volatility) by inverting the formula and solving for σ^2 . This is how implied volatility works, and how one can calculate the volatility from option pricing.

2.2 VIX

The VIX is the expected movement in the Standard and Poor's 500 over the next 30 days. It's calculated by the Chicago Board Options Exchange (CBOE) from option prices. It is implicit and to some point uses the assumption that large investors are risk-neutral. It is possible to buy and trade the VIX. The VIX is the annualized 30 day expected variance calculated with 'out-of-the-money' SPX option strips. The formula can be found online, but the strip of options has been fairly elusive. It was recently changed (Sept. 22, 2003) from calculating from 'at-the-money' values of the Standard and Poor's 100 to the 'out-of-the-money' puts and calls of the Standard and Poor's 500 that are at least one week until expiration. Investopedia lists 'out-of-the-money' to mean that if the option were to expire today, it would be worth no money (ie, option strike price is higher than asset market price for call, or lower than asset market price for a put). The VIX is usually one of the more volatile future contracts (example online being that from May to November 2004, a fairly low volatility period, the VIX futures daily variation from 0.06

*insert equation

2.3 J.P. Morgan RiskMetrics

The J.P. Morgan RiskMetrics model was developed by the J.P. Morgan firm in the late 1980's as a tool in risk management. It's a method of calculating volatility for indexes

and goods, and I will input the standard and Poor's 500 data into this model to calculate a derived/explicit measure of volatility. The equation is as follows:

$$\sigma_{t+1}^2 = (1 - \lambda) \sum_{\tau=1}^{\infty} \lambda^{\tau-1} R_{t+1-\tau}^2, \quad 0 < \lambda < 1 \quad (5)$$

and can be simplified down to

$$\sigma_{t+1}^2 = \lambda \sigma_t^2 + (1 - \lambda) R_t^2 \quad (6)$$

The J.P. Morgan RiskMetrics system, however, has its own critics. It heavily uses a the VaR (Value at Risk) system that, while still widely used, took a lot of the blame for the 2007/2008 financial crisis. In addition, according to Jamie Dimon (CEO of J.P. Morgan), a change in their VaR calculation methods was also partly to blame in masking a 2 billion dollar loss in May 2012. He said that "In the first quarter [J.P. Morgan] implemented a new VaR model, which [J.P. Morgan] now deemed inadequate, and went back to the old one that [J.P. Morgan] used for the past several years, which we deemed to be more adequate." Though VaR seems to have its critics, and can be calculated in a number of different manners, it still seems to be the benchmark method in the financial world. The VaR system calculates the worst expected loss over a given horizon at a given confidence level under normal market conditions, and the RiskMetrics uses it to calculate volatility for goods.

2.4 Possible Regressions to fit

1. Autoregressive AR(1)

$$\sigma_t^* = \phi \sigma_{t-1}^* + \epsilon_t \quad (7)$$

where ϵ_t stands for white noise and $\epsilon_t \sim No(0, \sigma^2)$

2. Autoregressive AR(2)

$$\sigma_t^* = \phi_1 \sigma_{t-1}^* + \phi_2 \sigma_{t-2}^* + \epsilon_t \quad (8)$$

where $\epsilon_t \sim No(0, \sigma^2)$

3. Moving Average MA(1)

$$\sigma_t^* = \mu + \epsilon_t + \rho \epsilon_{t-1} \quad (9)$$

where $\epsilon_t, \epsilon_{t-1} \sim No(0, \sigma^2)$

4. Autoregressive Moving Average AR(p)MA(q)

$$\sigma_t^* = c + \epsilon_t + \sum_{i=1}^p \phi_i \sigma_{t-i}^* + \sum_{i=1}^q \rho_i \epsilon_{t-i} \quad (10)$$

where all $\epsilon_t \sim No(0, \sigma^2)$

5. Autoregressive Integrated Moving Average AR(p)I(d)MA(q)

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t \quad (11)$$

where L is the lag operator, ϕ_i are the parameters from the AR part of the model, d are the number of 'differencings,' θ_i are the parameters from the MA part of the model, and ϵ_t is the error term. The ARIMA is a general model that can be adapted to form AR, MA, ARMA models by changing the values of p , d , and q . For example, by setting $d = 0$, one can create a AR(p)MA(q) model.

2.5 Risk Variance Premia

The Variance-Risk Premium is the compensation risk-premium an investor receives for time variant volatility. It's calculated

$$\text{VRP}_t = \mathbb{E}_t^Q[\sigma_{t+21}^2] - \mathbb{E}_t^P[\sigma_{t+21}^2] \quad (12)$$

I will model the $E_t^Q[\sigma_{t+22}^2]$ with the VIX data and the $E_t^P[\sigma_{t+22}^2]$ portion with the output from the J.P. Morgan RiskMetrics model when the Standard and Poor's 500 data is inputted.

3 Data

3.1 Data Description

My data will be taken largely from online sources such as Yahoo! Finance. The time series data found on such sites for indexes such as the Standard and Poor's 500, as well as the VIX is accurate, and is downloaded conveniently in a .csv format. The data comes in a number of columns, with headers as to the starting and closing prices for the day, as well as the highest and lowest. When analyzing the data, each day's closing price will be the one looked at. The VIX, as mentioned before, comes in units of annualized percent, whereas the Standard and Poor's 500 Index comes in a measure of dollars as pricing for the weighted option strip of 500 goods that the index is calculated from. With both indexes, the closing data will be used as to avoid confounding effects from overnight/overseas trading.

The strength of this type of financial data lies in that there are no problems with its accuracy or reliability, and that it has no bias.

When comparing values between the realized volatility and the VIX, the realized volatility will be multiplied by 100 to put the measures in similar units such that they will be comparable. The data for the Standard and Poor's 500 that will be used will be for the period during which the VIX has been extant and recorded: from January 2, 1990 to present (which I took to be January 25th, at the time of writing).

3.2 Raw Data Summary

Table 1: Raw Data Summary

Index	Obs	Mean	Med	Std Dev	Var	Max	Min
VIX C Price	5814	20.49	18.99	8.16	66.66	80.86	9.31
S&P500's C Price	15839	417.02	122.70	474.86	225491	1565.15	16.66

4 Data Modeling

4.1 AR(5)

The first model we will work with is the AR(5). The equation for the model is defined as follows.

$$\sigma_t^* = \phi_1 \sigma_{t-1}^* + \phi_2 \sigma_{t-2}^* + \phi_3 \sigma_{t-3}^* + \phi_4 \sigma_{t-4}^* + \phi_5 \sigma_{t-5}^* + \epsilon_t \quad (13)$$

This model was used for both the VIX and Realized Volatility datasets, with the following results:

Table 2: AR(5) on VIX

Index	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	Intercept	AIC
Coefficients	0.8687	0.0150	0.0395	-0.0042	0.0676	20.4498	21354.26
S.E.	0.0131	0.0174	0.0174	0.0174	0.0131	1.4704	n/a

Table 3: AR(5) on Realized Volatility

Index	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	Intercept	AIC
Coefficients	1.0255	0.0681	-0.0535	0.0156	-0.0634	15.9902	16647.46
S.E.	0.0131	0.0188	0.0188	0.0188	0.0131	0.0002	n/a

4.2 MA(5)

The next model we will work with is the MA(5). The equation for the model is defined as follows.

$$\sigma_t^* = \mu + \epsilon_t + \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \rho_3 \epsilon_{t-3} + \rho_4 \epsilon_{t-4} + \rho_5 \epsilon_{t-5} \quad (14)$$

This model was used for both the VIX and Realized Volatility datasets, with the following results:

Table 4: MA(5) on VIX

Index	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	Intercept	AIC
Coefficients	1.3108	1.3878	1.1798	0.7751	0.3523	20.4391	25495.13
S.E.	0.0153	0.0227	0.0192	0.0140	0.0121	1.7098	n/a

Table 5: MA(5) on Realized Volatility

Index	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	Intercept	AIC
Coefficients	1.6563	2.0583	1.8880	1.2245	0.4815	15.9970	22769.92
S.E.	0.0130	0.0188	0.0225	0.0184	0.0097	0.1881	n/a

4.3 ARMA(3,3)

The next model we will work with is the ARMA(3,3). The equation for the model is defined as follows.

$$\sigma_t^* = c + \epsilon_t + \sum_{i=1}^3 \phi_i \sigma_{t-i}^* + \sum_{i=1}^3 \rho_i \epsilon_{t-i} \quad (15)$$

This model was used for both the VIX and Realized Volatility datasets, with the following results:

Table 6: ARMA(3,3) on VIX

Index	ϕ_1	ϕ_2	ϕ_3	ρ_1	ρ_2	ρ_3	Intercept	AIC
Coefficients	0.5660	0.9874	-0.5601	0.2983	-0.7177	-0.0252	20.4499	21344.64
S.E.	0.0599	0.0220	0.0577	0.0609	0.0450	0.0197	1.6345	n/a

Table 7: ARMA(3,3) on Realized Volatility

Index	ϕ_1	ϕ_2	ϕ_3	ρ_1	ρ_2	ρ_3	Intercept	AIC
Coefficients	0.9020	0.9719	-0.8777	0.1102	-0.7907	0.0553	15.9844	16563.27
S.E.	0.0166	0.0139	0.0158	0.0211	0.0198	0.0136	1.2958	n/a

5 Correlations

The correlations from the VIX and the monthly volatility values were calculated, and

Table 8: Correlations

Index	0	1	2	3	4	5	6	7	8	9	10
Positive Lags	0.286	0.253	0.232	0.205	0.185	0.174	0.174	0.176	0.173	0.158	0.136
Negative Lags	0.286	0.311	0.337	0.349	0.359	0.364	0.390	0.403	0.421	0.436	0.451

Table 9: Correlations

Index	11	12	13	14	15	16	17	18	19	20
Positive Lags	0.122	0.117	0.127	0.135	0.146	0.162	0.165	0.167	0.157	0.146
Negative Lags	0.455	0.452	0.447	0.455	0.453	0.454	0.443	0.429	0.411	0.392

It could be seen from the correlations tables that the range with the strongest correlations were in the negative 11-16 lags. A correlation of 0.45 should be considered to be moderately strong, and when creating an instrument to forecast the VIX using the Standard and Poor's 500, I believe that forecasting from that range will yield the best results.

6 Conclusions and Further Studies

This study found that the VIX and the Realized Volatility time series were more likely to follow an autoregressive than a moving average. While the ARMA models fit the time series' better than both the AR and the MA models, the AIC from the ARMA models were not that much lower than the AIC from the AR models, indicating that while adding the MA models added an extra dimension that helped model the time series, the brunt of the modeling was most likely done by the AR process.

The moderately strong correlations between the VIX and the realized volatility of negative lag 11-16 were fairly encouraging for the prospects of finding a model to forecast the VIX better in the future. My one caveat with my results would be that I worked with the data from 1990-2013, rather than from 2003-2013. The VIX was derived from the Standard and Poor's 100 until 2003, when it was changed to the Standard and Poor's 500, and though there doesn't seem to have been a large difference, it may be that one time series could better predict the other if data from 2003-2013 had been used.

When I began the thesis I had started with the idea of using the JP Morgan RiskMetrics system to calculate volatility values for the Standard and Poor's 500, but the system proved to be too dense. The technical manual attempted to both sell the product while explaining the workings of the product, and it made the entire process much more difficult than it could have been.

I regret that I did not have the time to work further on using a Gumbel driven AR model (geared towards modeling time series' with extreme values) to fit the series, as I do feel it would have done an excellent job, but the idea came too late in the process to implement. I myself am very interested in continuing this path of research, and would like to explore it further during my time in graduate school.

The Risk-Variance Premia would also be a better topic for further discussion, as though I was able to calculate it for some values, I was not able to draw any relevant conclusions from it. Though I was unsuccessful while looking at it as a side-project, I do feel that a full study of the topic would yield great results.

7 Referencing

References will be included with the bibliography. (author?) [4, 6, 3] (author?) [5, 1, 7] (author?) [2]

8 Figures

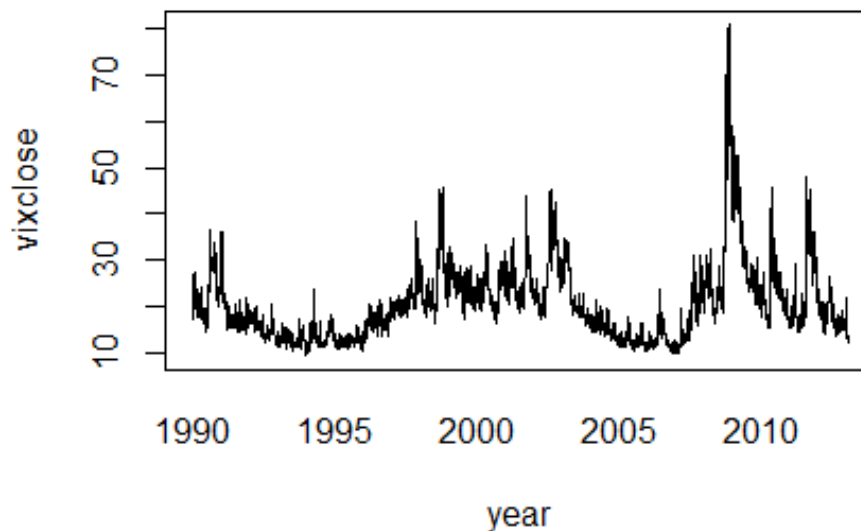


Figure 1: Plot of the VIX's closing values over time

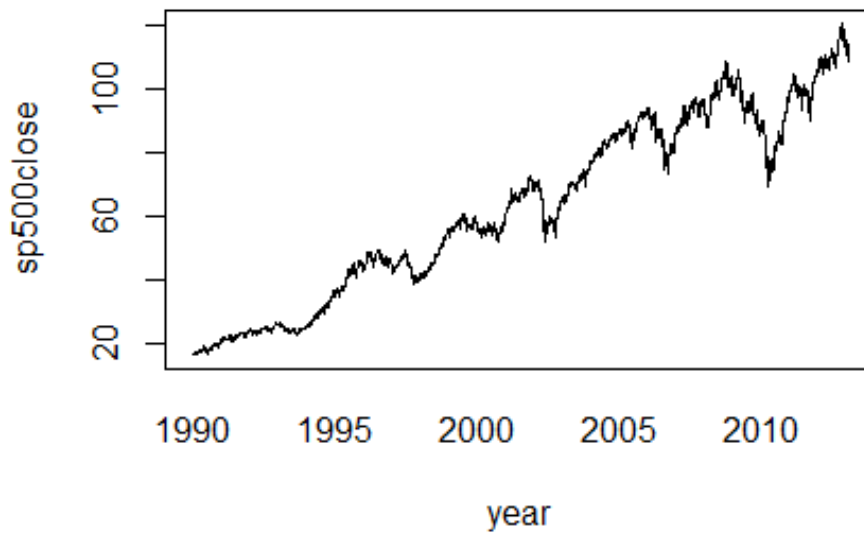


Figure 2: Plot of the Standard and Poor's 500's closing values over time

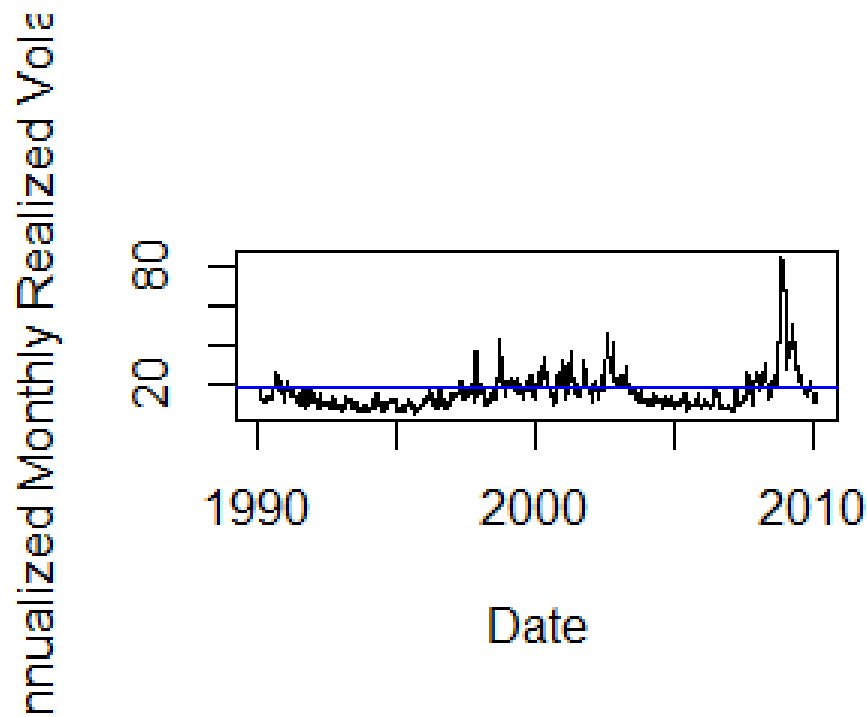


Figure 3: plot of the annualized volatility over time

vix1[, 2] & mvol1[, 3]

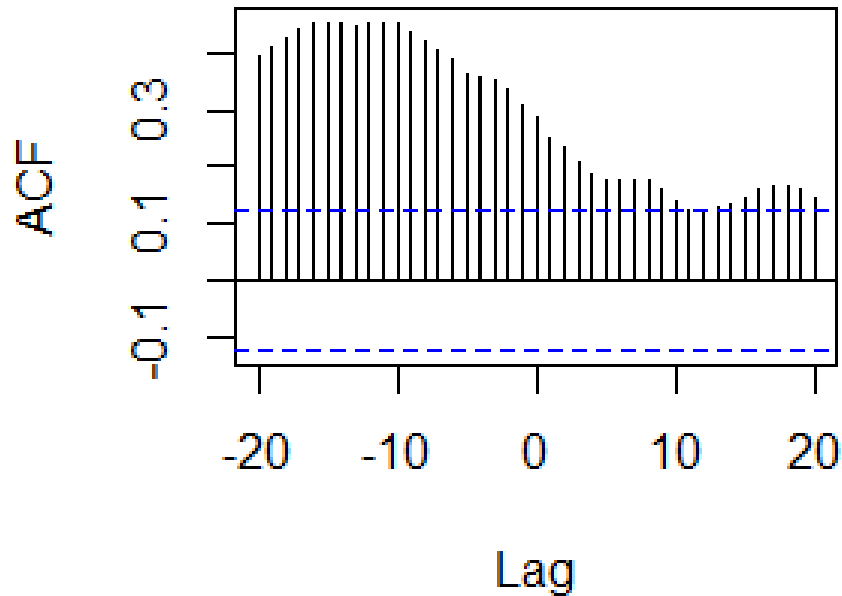


Figure 4: This is a plot of the correlations between the time series' by lag.

9 Acknowledgements

Much thanks to Professor Wolpert and Dr.* Jouchi Nakajima, as well as Professor West and Professor McElroy.

References

- [1] Romain Berry. Value-at-risk: An overview of analytical var.
- [2] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 1973.
- [3] Tim Bollerslev, George E. Tauchen, and Hao Zhou. Expected stock returns and variance risk premia. *AFA 2008 New Orleans Meetings Paper; Review of Financial Studies, Forthcoming; Duke Department of Economics Research Paper No. 5*, 2008.
- [4] Peter F. Christoffersen. *Elements of Financial Risk Management*. Elsevier, 2012.
- [5] Stewart Mayhew. Implied volatility. *Financial Analysts Journal*, 1995.
- [6] Robert E. Whaley. *Understanding VIX*, 2008.

[7] Christopher Whittall. Value-at-risk model masked JP Morgan 2 bln loss, May 2012.