Katarzyna Bieniek
Professor Stanley
Econophysics
14 May 2018

The Dependence Between Volatility Magnitude and Time Until the Return to Standard Volatility

Introduction

In economic terms, volatility describes the tendency of prices to change rapidly and unpredictably. This can be clearly seen in the stock market everyday. The prices go up and down, quickly, with seemingly random intervals and intensities. Even though randomness seems to govern this volatility, there are patterns that arise. Economics explains the reason for volatility as the response of consumers to their anticipations of future prices [1]. From an econophysics perspective, unlike the autocorrelation function of returns versus time lag, which is only short range correlated of about 4 minutes, the autocorrelation function versus the time of day for volatility is approximately linear on a log-log scale making it long range correlated based on a power-law [5]. Due to this correlation, and volatility clustering, several models, including ARCH and GARCH have been created to try to describe it [3].

Volatility clustering is the idea that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes", as first noted by Mandelbrot [2]. This has been already shown, but this made me interested in taking it a step further: is the time it takes for the large changes to go back to small changes dependent on the size of the change?

Methods

Volatility can be calculated several different ways including absolute value of returns, or log difference between consecutive returns. Unfortunately for long ranges simply doing the absolute values of returns is not ideal because as prices rise with natural inflation of the market, a higher change in price does not necessarily mean that the change in price is as impactful. This can be fixed by simply plotting the returns on a log-log plot. For this investigation, I chose to look at daily data from the S&P 500 between

1950 and 2018, so absolute value of returns would not work without a log-log scale. Instead, I took the absolute value of the difference between opening prices of consecutive days and divided by the price of the earlier day in order to obtain the percent price difference between consecutive business days.

Next, we need to have some sort of baseline for what we consider “normal volatility”, an area to which large fluctuations should return after some amount of time. To do this, I split up the S&P 500 data into groups of five years. The data between January 3, 1950 and December 31, 1954 was used only to create the baseline for the data going between January 3,1955 and December 31, 1959, which was the first set of data. Taking the mean plus one standard deviation of the data between 1950 and 1955 created the baseline or the "normal volatility". This became the baseline for the next five years of data (first set of data). Then the mean plus one standard deviation of the first set of data was the baseline for the second set of data and so on.

Assuming volatility values are normal (which they are not. They more closely resembles the student t distribution), this baseline would encompass approximately 84% of the data under this line, making the remaining approximate 16%, peaks. The next step is to calculate the amount of time, in days that it takes the volatility to go back under the set baseline for 7 consecutive days. This choice of 7 days is arbitrary but it made sense since that is a week. To explain further how the days count, let us assume you have one spike that is above the threshold, followed by one below, three above, and eight below. This would mean that our time until relaxation would be twelve days. Using this method, with the help of Python, the beginning, end, length, magnitude of the of the first point above the threshold, as well as maximum magnitude within each interval above the threshold were calculated. Using this data, scatter plots of Volatility spikes versus Time Until Stabilization were created for each five-year period and for the data as a whole (Fig 1,2,5,6). For each of these, the linear correlation coefficient (r), and the coefficient of determination (r^2) were calculated (Table 1). These two values tell us the following: “the greater the absolute value of a correlation coefficient, the stronger the linear relationship. The coefficient of determination indicates the extent to which the dependent variable is predictable. An R^2 of 0.10 means that 10 percent of the variance in Y is predictable from X” [4].

Results

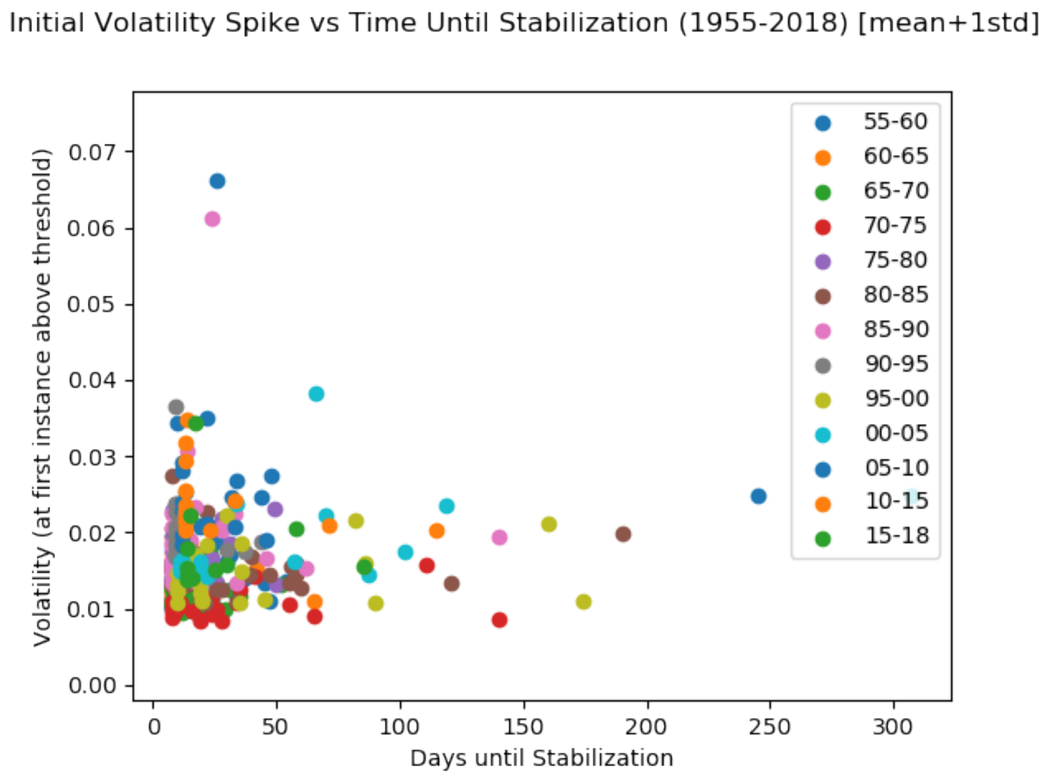

Figure 1. Graph of the initial spike in volatility versus the time until volatility returns back to "normal". The different colors represent the different ranges of data

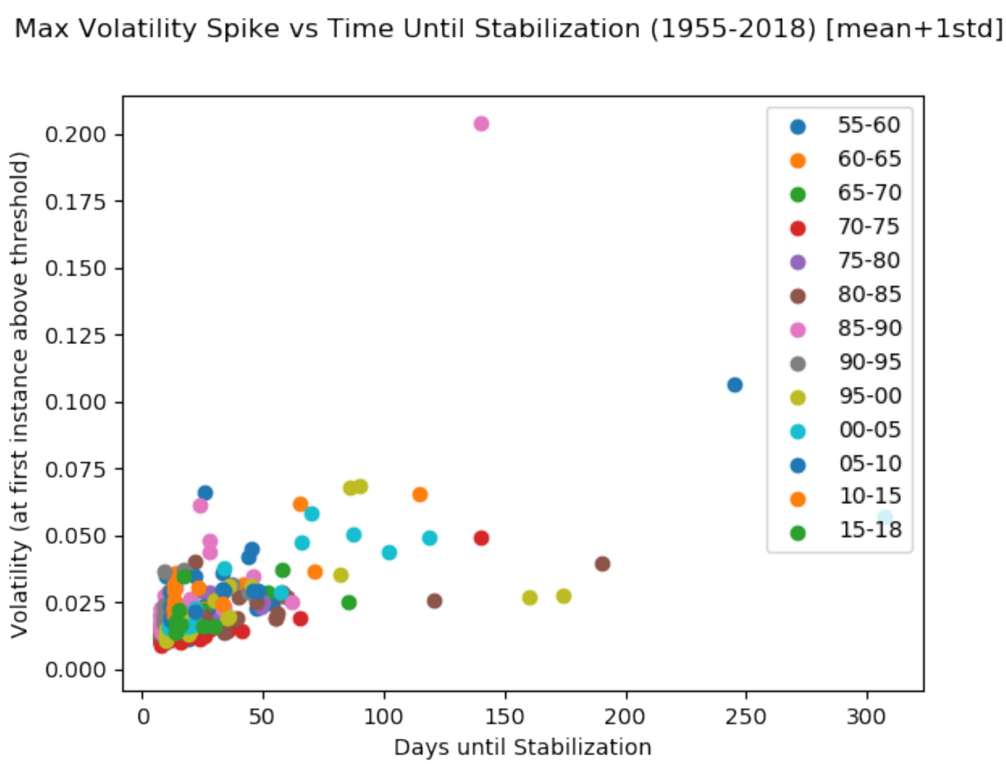

Figure 2. Graph of the maximum spike in volatility versus the time until volatility returns back to "normal". The different colors represent the different ranges of data

Simply looking at the first time the volatility is above the baseline and the time it takes for the volatility to get back to normal, the correlation is very weak (Fig 1). The R-value is only .097, suggesting that while the correlation is positive, the linearity of the data is weak. Only about 1% (R^2= 0.00947) of the change in number of days until stability can be explained by the magnitude of the first spike in volatility, suggesting that there is no relationship between the two.

When looking at the number of days until stability and the maximum magnitude of volatility on that interval, however, the relationship becomes much stronger (Fig 2). The R-value becomes .613, which suggests a stronger positive linear relationship and the R^2 value is .376, which means that now approximately 37.6% of the changes in length until return to "normal" can be explained by maximum volatility. This relationship is significantly stronger than the one with first peak volatility. This also suggests that most of the time, the initial point above the threshold is not the highest point, which possibly suggests that it takes some time for volatility to build up to its maximum point (as seen in Fig 3).

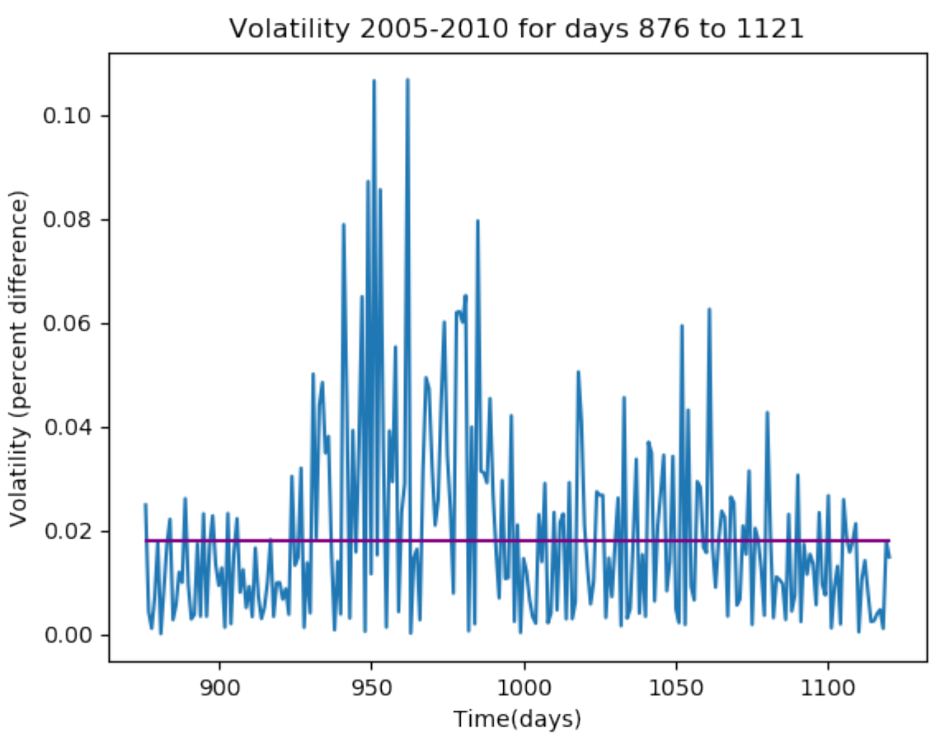

Figure 3. Example of in initial point above the threshold not being the maximum point and volatility taking time to build up

Years	Correlation Coefficient (R)	Coefficient of Determination (R^2)
1955-1960	0.558344093079	0.311748126277
1960-1965	0.863066598421	0.744883953311
1965-1970	0.795822519364	0.633333482327
1970-1975	0.949725283344	0.901978113822
1975-1980	0.560397618474	0.314045490792
1980-1985	0.651637747868	0.424631754447
1985-1990	0.893162354121	0.797738990819
1990-1995	0.470754146424	0.221609466375
1995-2000	0.531509544991	0.282502396417
2000-2005	0.748289429409	0.559937070165
2005-2010	0.972835822736	0.946409537998
2010-2015	0.919399407487	0.845295270488
2015-2018	0.465598641349	0.216782094826

Table 1. Shows all the R and R^2 values for maximum volatility magnitude versus time until relaxation for each 5 year period

The correlation was also calculated for each five-year period and is given in Table 1. As can be seen, sometimes the correlation is very strong (Fig 5), while other times it is weak (Fig 6). Plotted against time, it appears that this correlation fluctuates with time, where periods of low correlation between maximum volatility magnitude and time until relaxation are followed by periods of very high correlation (Fig 4).

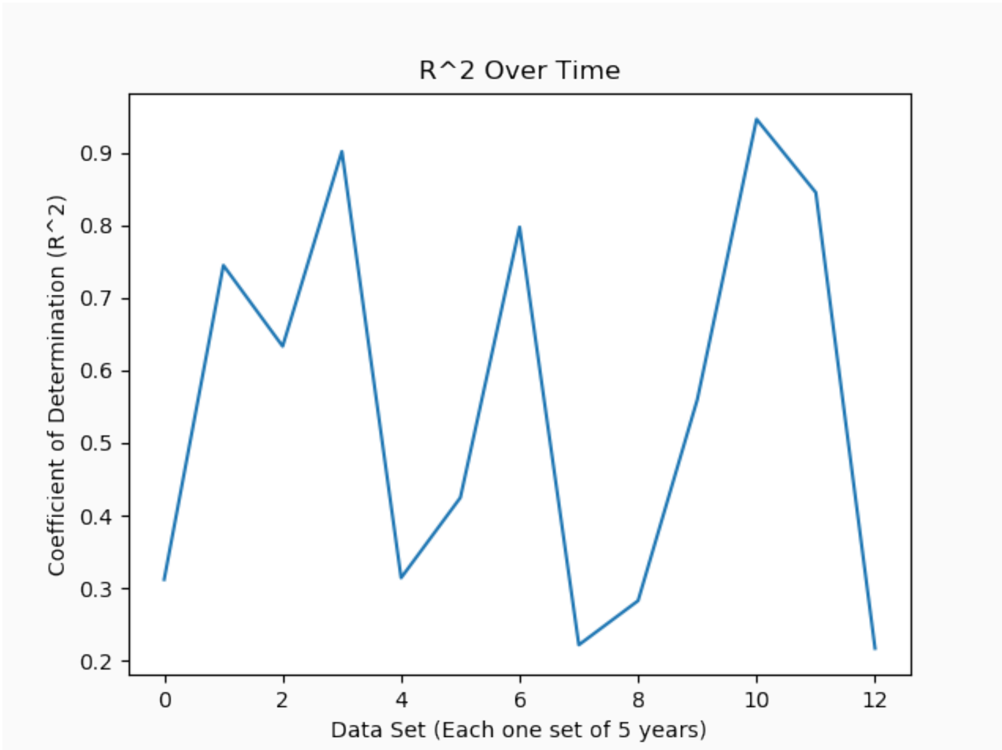

Figure 4. Graph of R^2 values for maximum volatility magnitude versus time until relaxation for each 5 year period

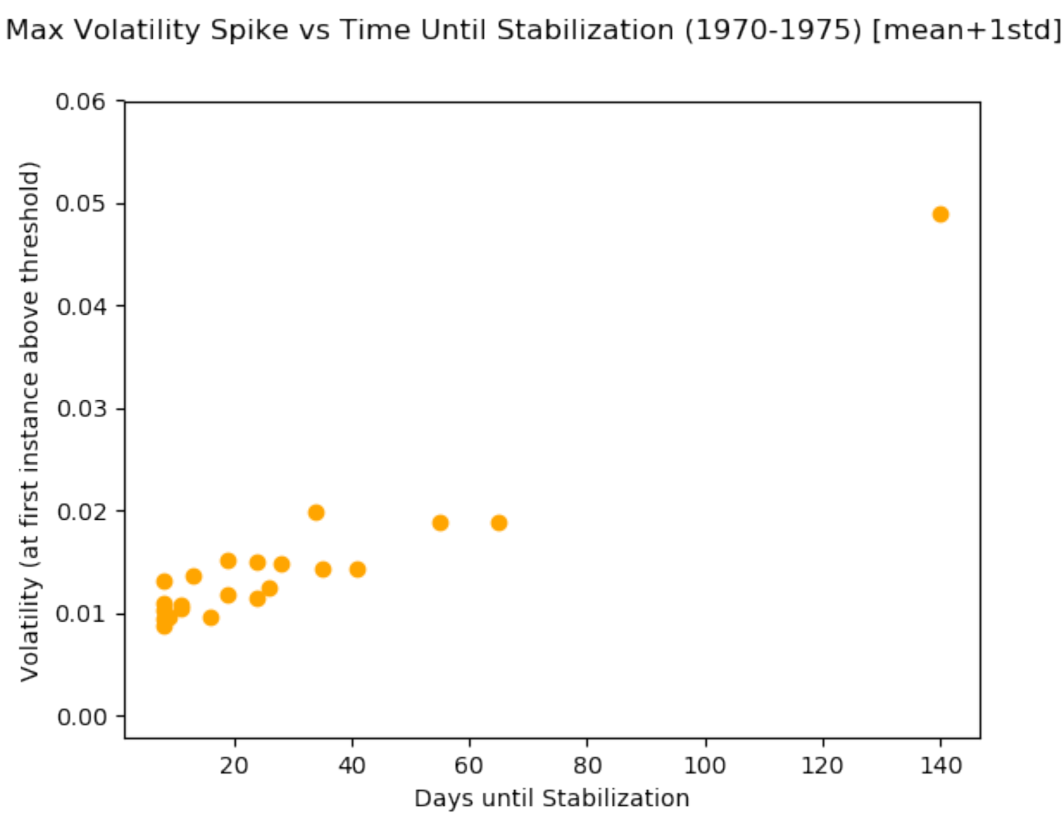

Figure 5. An example of a 5 year period with a strong correlation. R= .950, R^2 = .90

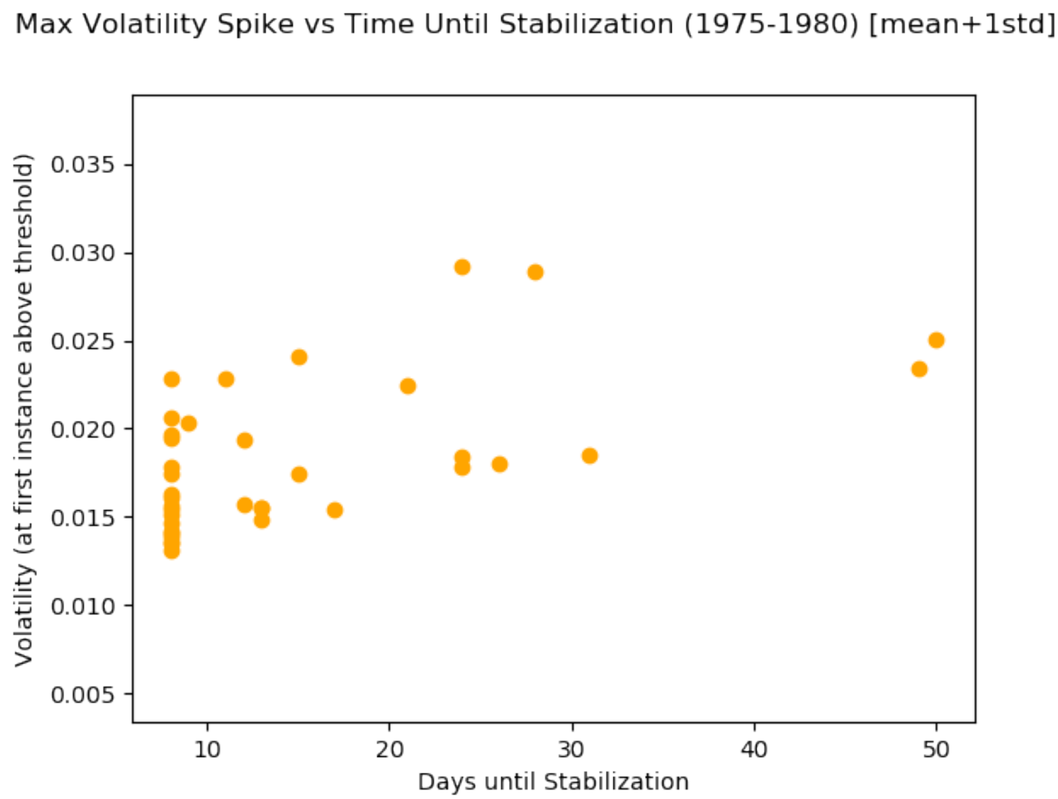

Figure 6. An example of a 5 year period with a weak correlation. R= .560, R^2 = .314

Continued Work

This data only comes from one stock, the S&P 500. This work would need to be repeated on other stocks in order to see if the same findings hold true. Furthermore, there are several areas which analysis could be improved upon. Regardless of what the normalcy period is set to (here it was 7 days), problems can arise. When the period is set to 7 days, if the next spike is 8 days after the last spike, then it could still be related to the previous increase in volatility, but it is counted as being a part of a different set. On the other hand, if this value is set to be too high, then everything will end up being included in one set.

Another area of concern is that the threshold was only updated once for every 5 year period. By updating the mean only every 5 years, at the beginning of the new 5 year period, the average is fresh, but at the end, it is already 5 years old. It might be better to have a 5 year running average instead as the threshold. The idea of breaking up the data into 5 year periods creates another problem: if there are peaks that go from the end of one

5 year period to the beginning of another 5 year period, this group gets cut off into two with this method and neither set has the correct relaxation length.

Conclusion

It appears that there is some correlation between maximum magnitude of volatility and the number of days until stability. This correlation itself seems to fluctuate with time. Furthermore, it can be seen by comparing Max Volatility Spikes and Initial Volatility Spikes that the maximum volatility is rarely the initial pass across the line of "normalcy". More data and a few changes to method of analysis are needed in order to make firm conclusions, but it does look like there are patterns that arise within volatility that go beyond its simple autocorrelation.

Works Cited

[1] Krugman, Paul R., and Robin Wells. Economics. 2nd ed., Worth Publishers, 2009.

[2] "Volatility Clustering." Wikipedia, Wikimedia Foundation, 21 Feb. 2018, en.wikipedia.org/wiki/Volatility_clustering

[3] Lux, Thomas, and Michele Marchesi. "Volatility Clustering in Financial Markets: A Micro-Simulation of Interacting Agents." IFAC Proceedings Volumes, vol. 31, no. 16, 1998, pp. 7–10., doi:10.1016/s1474-6670(17)40450-2.

[4] "Coefficient of Determination." Stat Trek, stattrek.com/statistics/dictionary.aspx?definition=coefficient_of_determination