Christopher van Hoecke.
PY 531
Eugene Stanley
4 May 2017

Research Project Write Up.

Dynamics of Skewness in ETF of the S&P500 returns.

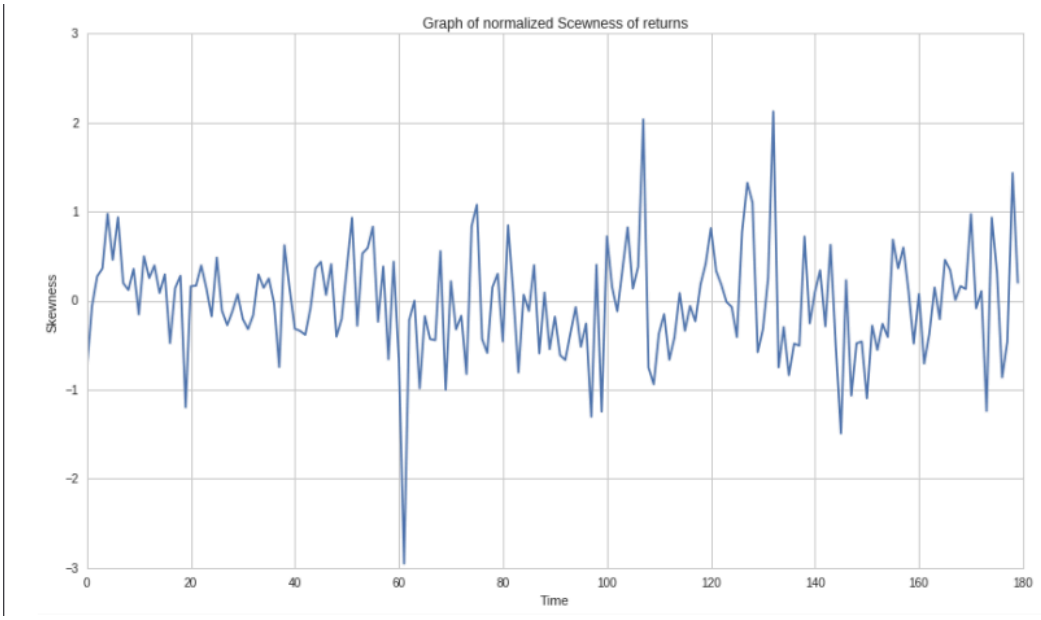
The talk that I gave earlier in class involved researching how skewness in market returns (which means how much from the normal the mean of the returns are deviated), mainly in the SPY ETF, which is the ETF for the S&P500 move with time. To do this, we took data going back as far as 2002, and got monthly skewness values for daily returns of stock prices. Using these skewness values, we attempted to find a model that would help accurately predict future values of Skewness. The hope would be that if we could predict such moments, an investor would be able to invest in the markets with the knowledge that a month from time t, the probability of having returns x on his investment would be significantly greater than they are now. This would allow an investor to make smarter decisions ahead of others thus beating competition and speculations.

We started off by finding the skewness values and the mean returns. From there we plotted the skewness over time to try and find some sort of a pattern to the data. While none was obvious to find, we did however show to explicit outliers in the data. Thus, enforcing the importance of our experiment. Finding and predicting when such outliers would occur would greatly impact market decisions such as stock picking. To continue our research, we set about finding probabilities of negative and positive skewness, and found there were more instances of negative skewing than positive skewing, counter to what behavior economist believe about the markets. This can be

attributed to the fact that our data is monthly and not yearly. Next, we set about finding confidence intervals, which posed a challenge as our data was not normalized. As a result, a definition of an outlier was set at ‘a skew value greater than 1’. From there we managed to find the instances of ‘outlier’ in the data. Which we found occurred at least once a year.

We found a total of about 8% of data were outliers in data going back to 2002.

Here in this graph, you can see the skewness moving as a function of time:



We next tried to find instances of greatest skewing and attached them to the average prices at the time. And found there was a delay of a couple of days between the skewness value and the mean returns value. While this is interesting, it does not conclude anything as it is not a successful model for determining when such outliers would occur.

We thus set out next to checking for autocorrelation and running a regression in the hopes of finding a conclusive model. The autocorrelation proved ineffective as it simply indicated that

our data was random (though this would motivate trying to model skewness with a random walk or Brownian motion) and the regression's error was too great to give it any significance.

In conclusion, we attempted to find a way to predict the skew dynamics in returns of SPY the S&P 500 ETF index. To start with, we set upon finding the skew monthly values. This is because we needed enough data points to have valuable skew values, while at the same time needing the right amount of data points. Once that was done, we set about finding the monthly means. This was for comparison from monthly returns to skew values. A linear regression was done to try and find an equation that best fitted the monthly skew values. Along with scatter plot of skew values and monthly values. We attempted to find the autocorrelation in the data, to find if the previous month's skew value could impact the present one. This was based on the paper earlier sited, indicating that markets have a natural tendency to be positively skewed. We find that on a monthly scale, there are more negatively skewed Data points, and that there isn't much correlation between previous and present skew values. While our linear regression seemed to provide a somewhat similar plot, we were unable to take the error into consideration given the normalization. We ultimately concluded that while it is not possible following this research to accurately predict skewing movements, it would be interesting to further it by running Multiple linear regression model on three variables (variance perhaps) and to try and find more accurate instances of outlying skewness in the returns in other assets as well. Another interesting note would be to take minute data, and find the daily skew values. This would give the added advantage of providing us with more data points, and we could perhaps see the transitions from positive to negative skewing more accurately.