A Deep Learning Approach to Forecasting S&P 500 Volatility

Kylie Taylor

16 May, 2019

1 Abstract

In financial trading being able to accurately predict future states of securities and indices is invaluable and widely sought after. The purpose of this study is to forecast the volatility of the S&P 500 index through the implementation of machine-learning software and neural networks. This study uses 17,430 observations of the S&P 500 index spanning from January 1, 1950 to April 24, 2019 to train and test a Structural Time Series (STS) model using software TensorFlow Probability. The focus of this study is localized to five major economic crises: "Black Monday", "Dot Com Bubble", 2002 Downturn, 2008 Financial Crisis, and European Credit Crisis. Using local linear effects, seasonal effects, and Variational Autoencoding, the model attempts to minimize the error of forecasts of the ten days after the first five days of the crisis. The RMSE of these forecasts lie withing 0.2449 and 3.341 units of volatility for all five series. While the models preform well at forecasting a random walk like the S&P 500, there is room for improvement in the methods used when developing the neural network and implementation of more data.

2 Introduction

Often times, feelings of distress and panic root from uncertain future events. This is particularly true for the financial sector. The financial sector has made many strides of improvement in documentation over the last few decades, through the ability to collect and store data. One method financiers use to measure how the overall market is preforming is to create an "index". An index is a collection of many securities (stocks, bonds, other investments...) used to represent an entire market. Two examples of stock market indices are the Dow Jones Industrial Average (DJIA or "Dow Jones") and the Standard and Poor's 500 (S&P 500) index. The S&P 500 is a collection of 500 different stocks that are estimated to best represent the stock market as a whole. The S&P 500 was founded in 1860 by Henry Varnum Poor and began tracking prices in 1923. The S&P 500 includes all stocks in the DJIA, and 470 more. Each stock in the S&P 500 is weighted according to total market values of outstanding shares, implying companies with many shares and high market value will have a higher weight and represent a larger proportion of the total stock market. This analysis will use data collected from the S&P 500.

In particular, this analysis will focus on the S&P 500 behavior around five prominent economic crises. These five crises are "Black Monday", "Dot Com Bubble", stock market downturn of 2002, financial crisis of 2008, and the European debt crisis. I will give brief descriptions of each.

The crisis known as "Black Monday" occurred on October 19, 1987, obviously on a Monday. Originating in Hong Kong and spreading across all global markets, this crash is the largest one day drop in U.S. stock market history, where both the Dow Jones and S&P 500 both fell over 20% in one day. Causes of this crash are attributed to program trading and illiquidity. Now there are restrictions put on trading when the market drops over 7% and trading suspended for the day when the market drops 20% because of this crash.

The burst of the "Dot Com Bubble" occurred on March 10, 2000, at the turn of the millennia, lasting until October 9, 2000. Three main factors led to the expansion of this bubble and ultimately its downfall: popularity of internet and internet companies, lower interest rates and capital gain taxes, and speculative investing in internet companies. In a series of events with increased interest rates, a recession in Japan and failure of a merger between Yahoo! and Ebay, the bubble burst. This resulted in a severe decline in the value of nearly every internet company.

The stock market downturn on 2002 occurred on October 9, 2002. This downturn occurred as a correction to the nearly decade long bull market and market inclines after the September 11 attacks. This was a much less dramatic decline, as it took several months to hit its low in October 2002.

The financial crisis of 2008, or commonly known as the "Housing Bubble", occurred on September 16, 2008. This infamous crisis was triggered by the downfall of major financial institutions, due to the abuse in issuing sub-prime mortgage loans and credit default swaps that were issued to insure those loans. In this crisis, the severe decline of the Icelandic krona nearly resulted in government bankruptcy.

The last crisis is the European debt crisis, occurred on April 27, 2010. More commonly associated with the labor market crisis in Greece, this economic shock was caused by several European Union members being unable to bail out over-indebted banks, and/or repay government debt. Another contribution is the structure of the euro zone as being currency union, but not fiscally union, ultimately limiting the ability for European leaders to respond.

The goal of this analysis is to determine how well a deep-learning framework preforms in forecasting S&P 500 volatility ten days following the first five days of these five crises. The forecasts will be made through the use of a Long Short Term Memory (LSTM) network using machine-learning software, TensorFlow Probability. A LSTM is a type of neural network that belongs to a class of neural networks called Recurrent Neural Nets (RNN). In the most simple terms, a RNN is a neural net with loops among the nodes, allowing information to persist or for the network to "remember". A LSTM does the same thing and is also able to remember longer-term information. The standard LSTM contains four interacting layers, whereas the the standard RNN only contains one.

3 Data and Methods

The data used in this analysis is the S&P 500 index values collected daily from January 1, 1950 to April 24, 2019. This data was obtained from Yahoo! Finance and has 17,430 observations. There are seven variables included in the original data set: Date, Open, High, Low, Close, Adjusted Close, and Volume. All prices are reflected as real 2010 U.S. dollar values. I will only be using Date and Close values for the duration of this project. Below is a plot of all 17,430 Close prices.

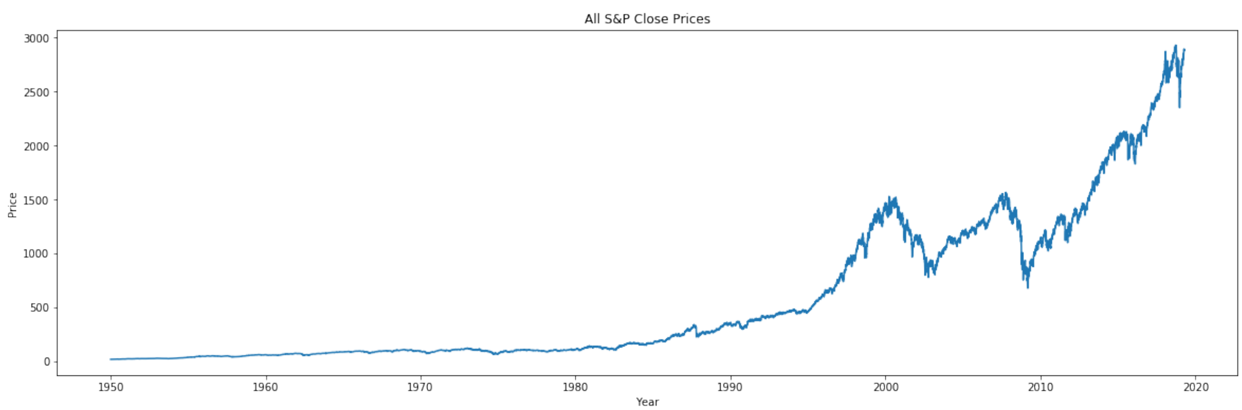

Figure 1: S&P 500 Index

Simply by inspection, it is clear that the data collected from Jan 1950 to nearly 1980 has very little variation, likely due to data collection abilities. Another obvious observation is that there is a clear upward trend. To combat these issues, I am only conducting analysis after 1985 and I will use the intra-day volatility of Close price, instead of absolute Close price. The calculation of the Close price intra-day volatility follows

$$rate_t = \frac{Close_t}{Close_{t-1}} - 1 \tag{1}$$

$$vol_t = log(rate_t^2) \tag{2}$$

I calculate the daily rate of return of Close price, then calculate the volatility of the daily rate of return using the above transformation. I need to take the square of the rate of return, because the log of a negative value is undefined. This transformation makes the series stationary and will now allow me to make slightly longer horizon forecasts, which I chose to be 10 days. A plot of the intra-day volatility of Close price follows

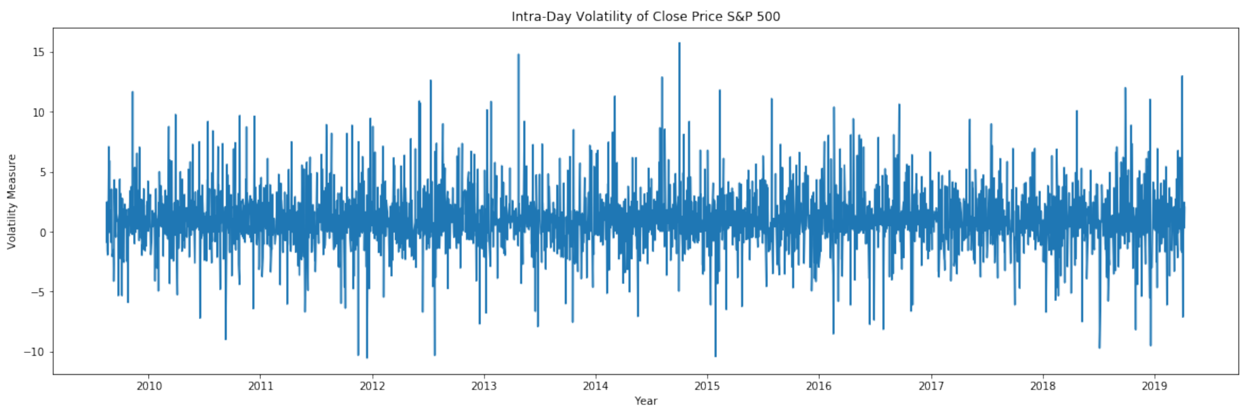

Figure 2: Intra-day Volatility of S&P 500 Index

Once making these transformations, I began building the model for forecasting the values using machine learning platform, TensorFlow Probability. TensorFlow Probability (TFP) is the March 2019 update of TensorFlow that combines deep-learning software and probabilistic models. To be specific, I used a package in TFP called Structural Time Series (STS) that uses a Bayesian framework for analyzing time series data. The best comparison is to R's BSTS package, except employing the computational power of neural nets.

A general step-by-step process I took when forecasting the volatility after a crisis follows:

1. Specify training data (n=205) and testing data (n=10).

2. Build the STS model by specifying a local linear trend and daily 'seasons'.

3. Feed the model into a Variational Autoencoder and build posterior distributions of the data.

4. Minimize variational loss by iterating over 200 variational steps/iterations.

5. Print the loss values.

6. Feed into STS forecast function.

7. Calculate out-of-sample RMSE.

8. Plot the 10 forecast estimates against the testing data.

The neural net aspect of this model is introduced at the Variational Autoencoder step. An encoder network takes in an input, and converts it into a smaller, dense representation, which a decoder network can use to convert it back to the original input, resulting a vector of size n. An autoencoder network generates encodings specifically useful for reconstructing its own input. A Variational Autoencoder uses the same general framework as an autoencoder, except it will output two vectors of size n: a vector of means, μ, and another vector of standard deviations, σ. The goal of this VAE is to minimize σ, or loss. This loss is calculated by incorporating Kullback–Leibler divergence and likelihood into the loss function. By minimizing the KL loss I am inherently maximizing the likelihood that my mean and variance of the forecasts are close to the true mean and variances of the testing data. The exact equation I am trying to maximize is

$$L(\theta) = logp(x) - KL(q_\theta(z|x)||p(z|x)) \tag{3}$$

Where θ is each iteration of the VAE (200 in total), x is the data, or the volatility of Close price, and z are the parameters, or the past values of the volatility. Since the goal is to maximize the likelihood, this can be done by minimizing the KL divergence portion of the function. In practice, we normally take the estimate of $logp(x)$ as $E_{q_\theta}(logp(z|x))$. This is precisely what I am doing when I calculate the loss at each θ iteration.

4 Results

In this section I will present the output from my TFP STS forecasts for each of the five crises. Before I present the visualizations of the forecasts, I will discuss the loss values I estimated. The loss function at the last iteration of the VAE are 510.29 for "Black Monday" forecasts, 474.81 for "Dot Com Bubble" forecasts, 464.08 for 2002 Downturn forecasts, 476.73 for 2008 Financial Crisis forecasts, and 507.87 for European Credit Crisis forecasts. These numbers are a measure of error rates, but cannot be directly compared to out-of-sample RMSE. The out-of-sample RMSE values are 0.2449, 0.4539, 0.6866, 3.341, and 2.889, respectively. These RMSE values seem acceptable in comparison to the units of volatility, as they are not abnormally large. This reveals that the TFP STS model did an acceptable job of forecasting volatility. The performance of the models may be easier to see in the following graphs.

The forecasts for "Black Monday" are plotted below. The orange dotted line are the forecasts that the model makes, and the solid blue line is the observed series. The light orange bands are the 95% confidence intervals of the forecasted values.

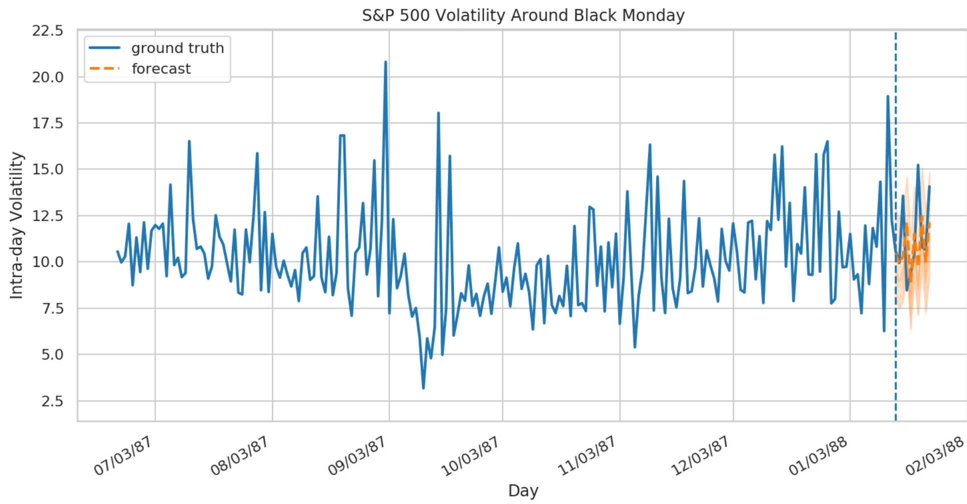

Figure 3: Forecasts of Intra-day Volatility After Black Monday

We see that the observed series lies within the 95% bounds of the forecasts most of the time. The forecasts also appear to follow a similar oscillation to the observed values.

The next following graph is of the forecasts made for the days following the "Dot Com Bubble" collapse.

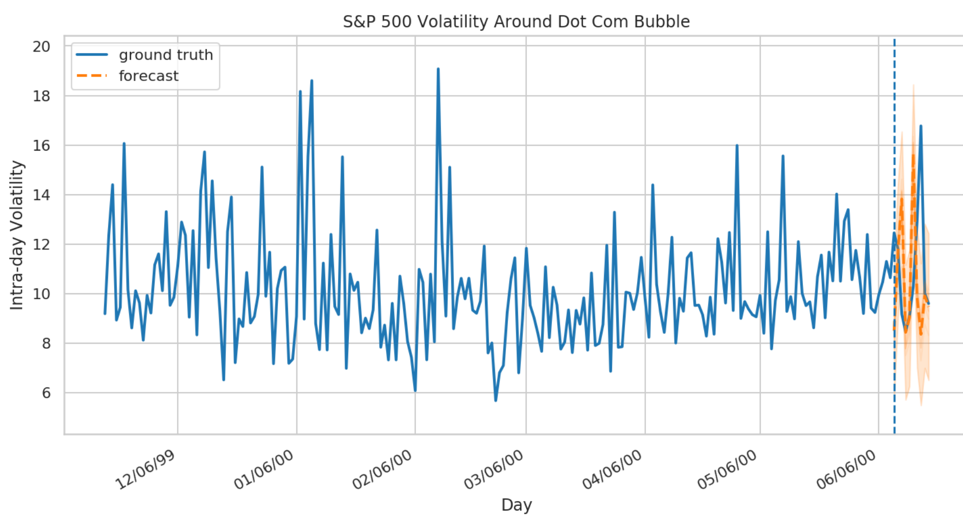

Figure 4: Forecasts of Intra-day Volatility After Dot Com Bubble Collapse

We see that the observed series does not fall within the 95% bounds of the forecasts as well as the series above. This is likely due to the larger swings in the observed series, which the forecasts (orange dotted line) appears to follow relatively well.

The next graph is of the forecasts following the 2002 economic downturn.

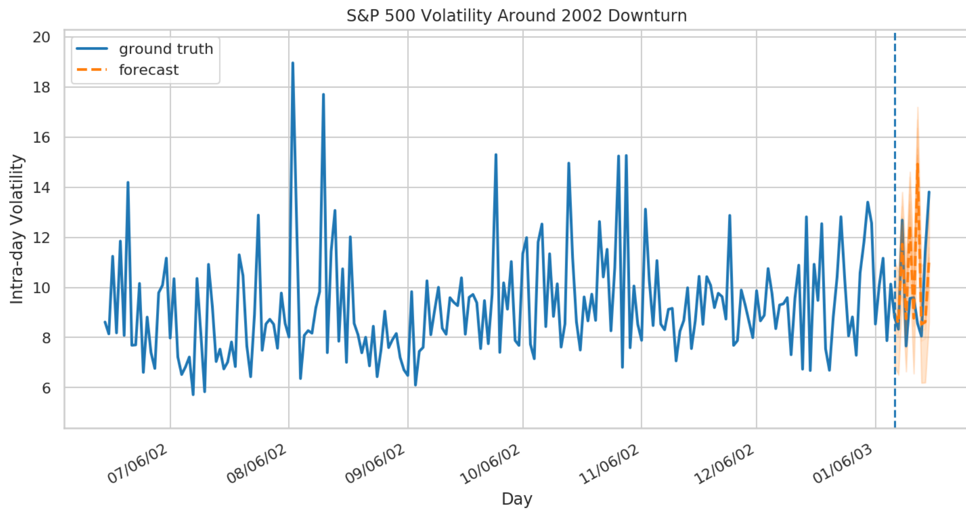

Figure 5: Forecasts of Intra-day Volatility After 2002 Downturn

With close inspection, the observed series appears to fall within the 95% bounds of the forecasts, even though the forecasts do not appear to estimated the observed values very well. The forecasts make large predictions where the actual values are lower. This model appears to have not preformed as well as the last two.

This graph presents the forecasts after the famous 2008 Financial Crisis.

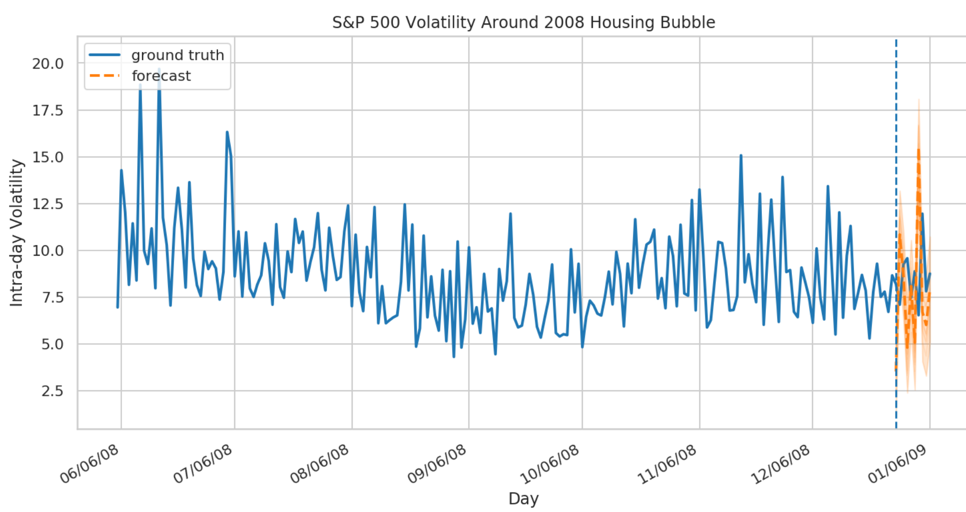

Figure 6: Forecasts of Intra-day Volatility After 2008 Financial Crisis

While the observed series is contain within the 95% bounds of the forecasts, the forecast values are far more extravagant in their estimations than the actual values are. This is also revealed by the largest out-of-sample RMSE of the five forecasts, 3.341 units.

The last plot in this section is for the forecasts following the European Credit Crisis and Greek labor market crash.

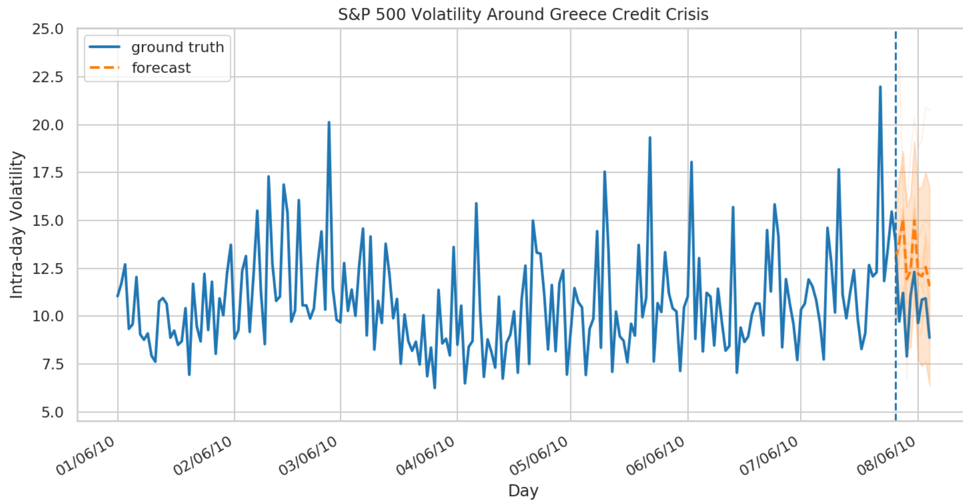

Figure 7: Forecasts of Intra-day Volatility After European Credit Crisis

The forecasts of this series would be very good if the estimates were shifted down a by approximately 4 units of volatility, since the forecasts appear to follow the same path as the observed values. The observed series lies within the lower bound of the 95% confidence interval of the forecasts.

5 Concluding Remarks

While the forecasts of S&P 500 volatility following economic crises using Bayesian Structural Time Series (STS) models are fascinating, there is room for improvement, as with nearly every statistical model. Some findings that really stood out to me was that the testing data fell within the 95% confidence interval of the average forecast for almost all forecasts of the five series. This is not terrible since the STS model was willing to make forecasts that had large deviations from the average of the series, unlike a typical ARMA model that makes much more reserved forecasts. Ideally, the observed series would like within a 10% bound of the forecasts, but this is hard to attain for a random walk like the S&P 500.

The natural next step would be to makes improvements to the underlying neural structure. One way would be to compare the out-of-sample KL loss or RMSE from different ways of calculating the VAE loss, or to optimize using varying methods of gradient descent. After working out any kinks in the STS model, I think it would be very interesting to incorporate an NLP aspect into the forecasts. This could be done by reading in daily articles pertaining to the S&P 500 and other financial markets, identifying words that are linked to fluctuations in the market, and then using those findings to update forecasts of any stock option or index. There are an endless amount methods that could be implemented to make this model better, especially as new research and software in deep-learning and neural networks emerges.