A Regression Model to Predict Stock Market Mega Movements and/or Volatility using both Macroeconomic indicators & Fed Bank Variables

Timothy A. Smith[#1], Alcuin Rajan

Embry Riddle Aeronautical University
600 S. Clyde Morris Blvd. Daytona Beach, Fl 32114 U.S.A

Abstract *In finance, regression models or time series moving averages can be used to determine the value of an asset based on its underlying traits. In prior work we built a regression model to predict the value of the S&P 500 based on macroeconomic indicators such as gross domestic product, money supply, produce price and consumer price indices. In this present work this model is updated both with more data and an adjustment in the input variables to improve the coefficient of determination. A scheme is also laid out to alternately define volatility rather than using common tools such as the S&P's trailing volatility index (VIX). As it is well known during times of increased volatility models like the Black-Scholes will be less reliable, hence, this work can be used to identify such times in a forward moving timeframe rather than using trailing economic indicators*

Keywords — *Partial differential equations, regression analysis, stochastic, financial mathematics. AMS classification: 35K10.*

I. Introduction

During the time period of the stock market crash of 2008, often referred to as the depression of 2008, it was observed that the commonly used stock market prediction models, such as the famous Black Scholes Stochastic Partial Differential Equation [1],

$$\frac{\partial X}{\partial t}+\frac{1}{2}\sigma^2 S^2\frac{\partial^2 X}{\partial S^2}+rS\frac{\partial X}{\partial S}-rX=0$$

demonstrated limitation in their ability to predict during rapidly changing times of volatility [2-3]. For example, in non-rapidly changing times of volatility, the famous Black Scholes Formula - using the standard notation for the normal distribution

$$N(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\tau^2/2}d\tau$$

along with the call option maturity date T, the strike price k and the risk free interest rate r - will accurately predict the fair price of an option from which future valuations can be obtained. Namely given the stock's price today as x_0, the Black Sholes Formula will give the fair price of the option as

$$x_0N\left(\frac{\ln\left(\frac{x_0}{k}\right)+\left(r+\frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right)-ke^{-rT}N\left(\frac{\ln\left(\frac{x_0}{k}\right)+\left(r-\frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right)$$

Now, a value of concern in this model is the value of σ "volatility" which has been debated for many years. Starting from the original famous pioneering work [1] of Fisher Black, along with Merton Sholes, some concern was noted as to exactly how to define σ. In general, it is taken that σ^2 is the variance of the assets under consideration but it is not so obvious as to how one can obtain that value practically in the real work. Moreover, it is observed that a change in σ can drastically change the output of this option's fair price. For example, if one were to study an option with stocks value of 100, maturity date of one year, and strike price of $10 along with the risk free interest rate 5% one would obtain the options value to be $14.39 provided that a minimal value of volatility was inputted as 10%. But, if that volatility was to increase to a more common measure of turmoil times, say 20%, then the options value would jump up to almost $19. But, if one halved the volatility to 5% then the options value would be almost unchanged and stay just north of $14. Obviously the model is telling us something about volatility! When studying the S&P 500 many experts in the financial management and prediction field commonly use the VIX index which is basically the variance of yesterday's S&P movement; however, that is clearly a lagging or trailing economic indicator which may not be the best value to use. In this paper we develop a new linear regression model which can be used to determine how far the "market" is from the "economy," and it is expected that the results of such a model can be used to define a value of σ perhaps as sort of an implied volatility rather than lagging economic indicator

II. Regression Model

Various economic indicators allow predictions of the future performance of an economy to be drawn. In a prior study [4] four major indicators were utilized to build a model; namely: the consumer price index (CPI), producer price index (PPI), gross domestic product (GDP), and money supply (M) and the model was found to work quite well with a coefficient of determination near 0.8. Furthermore, in another prior study [5] a numerical scheme was constructed to effectively create an implied volatility that was shown to improve the performance of the Black Sholes model during the crash of 2008, in a sense mathematically predicting the event. At that time, and based on common belief in market investment firms [6], that the model was optimal and adding extra variables would not improve the model. However, this would not necessarily be a static result as with time moving forward and the growing availability of data it was likely that a more optimal model could be created. The purpose of the present study was to do exactly that.

In the prior case [6], the independent variables were the x_1=GDP and x_2=PPI and x_3=CPI and x_4=M; the dependent variable will be the actual value of the S&P 500. With readily available monthly data, we can construct a MLR model from January 1990 to July 2013 by following a regression model of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots$$

Because one may desire to compare the residues of the model with the VIX it may be required to restrict the dates of analysis to the number of months of data we have for the VIX. The calculation methodology changed for the VIX in 1990, and we may need to restrict from going back further, however the regression model itself can be constructed for whatever time period its variables are readily available for (roughly since 1960). The governing equation of the model over this restricted time frame is

$$y_i = 1948.181 + 0.287x_1 + 16.581x_2 + 9.891x_3 + 0.074x_4.$$

In this model i = 1,…, n corresponds with the month in question out of n total months starting in 1990.

In the present study various other variables were attempted to be added into the model. For example it was considered to add in the variables of unemployment rate, interest rate, value of international stock markets, value of exchange rates of various international currencies, along with various consumer confidence polls. This process was done on a one by one basis followed with a proper statistical model analysis. Furthermore, a variable which showed any mathematical reason for improvement in the model was kept, while ones that did not were not kept. After in depth analysis it was found that the variables the z_1=GDP and z_2=FFR (fed funds rate) and z_3=M and z_4=U (unemployment rate) were the most useful. And, should one desire to utilize a four variable model it appears that an extremely improved model would be

$$y_i = 965.0635 + 349.0002z_1 - 12.1984z_2 + 186.5665z_3 - 177.406z_4.$$

Where one should note that in this model the variables are normed, using the usual Z transform, and that the variables indices are referring to different variables than in the prior model. In addition, it appears that in this model the CPI & PPI (which one would expect some inter relationship) have been absorbed by other variables? While this model is not the statistically correct model to keep, one may desire to keep the FFR variable as it is observed that sudden changes in that variable do have a strong effect on the stock market! For example when studying the data on a finer time range of the data, such as during the years 2000 – 2017, one will see the effect of the FFR does carry more weight in the model which does make sense due to the events after the crash of the real estate market during that time.

	Coefficient	*St Error*	*t Stat*
Int	965.0635	6.662222	144.8561
GDP	349.0002	32.39814	10.77223
FFR	-12.1984	14.47617	**-0.84265**
MS	186.5665	28.78833	6.480629
UI	-177.406	8.24829	-21.5083

Fig. 1 The data analysis of the updated 4 variable model .

However, if one choses to take a statistically proper approach that will remove the FFR variable and doing so they will obtain the model

$$y_i = 965.0635 + 362.1807z_1 + 183.392z_3 - 173.762z_4.$$

Here, the variables have not been relabelled so that the reader can keep track of what quantities they represent while avoiding a more proper, but lengthy, statistical notation such as $Z_{1.134}$.

	Coefficients	*Standard Error*	*t Stat*
Intercept	965.0635	6.659789	144.909
GDP	362.1807	28.36177	12.77003
MS	183.392	28.53035	6.427962
UI	-173.762	7.021221	-24.7482

ANOVA

	df	*MS*	*F*
Regression	3	41499537	2333.34
Residual	397	17785.47	
Total	400		

Fig. 2 The data analysis of the optimal 3 variable model.

Thus, we have now found not only a statistical optimal model which, as one can see above has a very solid F statistic along with an amazing multiple R^2 value of 0.95, but we have also created a computationally speaking easier (better) model! In addition, these results are also quite interesting as they do show what really drives the market is truly the large macro-economic indicators: the amount the US produces, the amount of money floating around and the rate of people unemployed. Moreover, it is also essentially telling us that things the government creates, such as indices like PPI or rates like FFR, do not have as much effect on the overall market as one might think. However, it is still the belief of the author that FFR is a variable worthy to keep track of due to that fact that so many things in current society are conducted on borrowed money and sudden changes in FFR could have drastic, albeit often human emotionally driven, effects on the market.

III. CONCLUSIONS & SUGGESTION FURTHER STUDY

In this work a prior regression model has been updated to both improve it statistically, with a significant increase in both the coefficient of determination (R^2) along with the F stat (P value) of the model, and at the same time simplified it computationally speaking due to the reduction in the number of inputs. In a common sense approach one can see, as outlined in the graph below, that this model predicted, due to extremely high deviations above, the "market top before market crash." In addition, it showed a turning point which represented the "market bottom."

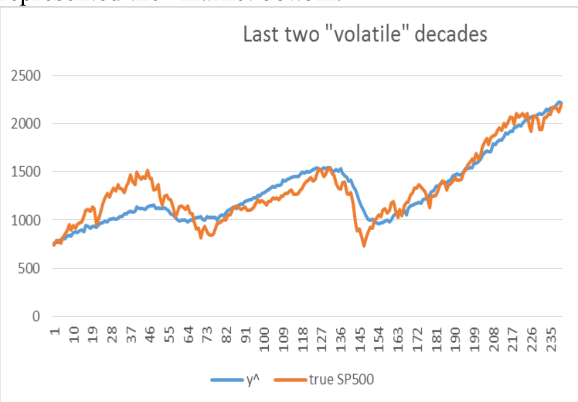

Fig. 3 A graphical representation of the models performance.

While this information could be very useful for financial traders to predict long term market trends, which could lead to a very profitable return on investment, this is not the mathematical purpose of our study. As it is well know a lot of interest has been given as of late to the topic of market volatility. It is expected that this model can be used to predict an alternate measure of volatility which then could be inputted into other financial models, such as the famous Black Sholes formula, and from that it is expected that those models would perform better in volatile times. In the prior study [5] a scheme was conducted to utilize a similar model to analyse volatility along with the VIX; furthermore, in that work statistically significant results were developed, hence validating the idea. It is suggested to reconstruct the study but using this new model, and it is expected that not only should the results be reproduced but since this new model is improved the significance of those results should increase. Moreover, it is suggested that either a regime switching method to be used for actual market predictions or perhaps a simple implementation of volatility can be driven from this. For example, if one were to define the deviation of this model output y^ from the true SP500 value as Δy, perhaps the volatility $\boldsymbol{\sigma}$ can be defined in a pricewise manner such as outlined in the chart below. Of course much study and data analysis would be needed to accurately define this function!

$$\begin{bmatrix} \Delta y & \sigma \\ <100 & <0.05 \\ 100-200 & 0.05 \\ 200-300 & 0.1 \end{bmatrix}$$

REFERENCES

[1] Black, F. and Scholes, M. "*The Pricing of Options and Corporate Liabilities,"* Journal of Political Economy 81 (3), 1973.

[2] Colander, David and Föllmer, Hans and Haas, Armin and Goldberg, Michael D. and Juselius, Katarina and Kirman, Alan and Lux, Thomas and Sloth, Birgitte, "The Financial Crisis and the Systemic Failure of Academic Economics." Univ. of Copenhagen Dept. of Economics Discussion Paper No. 09-03, (2009)

[3] Moyaert, T. & Petitjean, M . "The performance of popular stochastic volatility option pricing models during the subprime crisis." Applied Financial Economics. 21(14), 2011.

[4] Smith, T. at al, "An Economic Regression Model to Predict Market Movements." International Journal of Mathematics Trends and Technology, 2014

[5] Smith, T. et al, "A regression model to investiage the performance of the Black-Sholes using macroeconomic predictors." International Journal of Mathematics Trends and Technology, 2013

[6] Park, Sam. "*Reducing the Noise in Forecasting the SP 500*," Wentworth, 2005.