S&P500 with Vector Autoregression

& Conditional Forecasting for Stress Testing

Kevin Amrelle

April 25, 2017

TABLE OF CONTENTS

ABSTRACT 3

CHAPTER 1: INTRODUCTION 3

CHAPTER 2: MATERIALS AND METHODS 4

2.1 Literature Review 4

2.2 Data 5

2.3 Statistical Tests 6

2.3 Forecast Models 7

CHAPTER 3: RESULTS 9

3.1 Conclusion 9

REFERENCES 9

APPENDIX 10

**ABSTRACT**

This study utilizes Dickey-Fuller Generalized Least Squares, Pearson Correlation test and the Granger Causality test to build a framework for a vector autoregression and conditional forecast. From this study, I find that Non-Farm Payroll and PCE are Granger causal to the S&P500. The S&P500 is forecasted more accurately with the inclusion of Non-Farm Payroll specifically because it is a great predictor which was discovered from the conditional forecast. The S&P500 is predicted to increase over the next two-years as predicted from this VAR forecast using macroeconomic variables.

**CHAPTER 1: INTRODUCTION**

The primary purpose of this paper is to forecast the S&P500 and generate the effects of higher interest rates, faster inflation and rising unemployment. By doing this I can test which variable effects the S&P500 the most. In addition to this there will be standard statistical tests run on my data as well as a conditional forecast and VAR. For the first part of this study I will apply a vector autoregression which will provide the relationship between the dependent and independent variables. I will use a Granger Causality framework for this part with the intention on explaining the relationship of the S&P500 with the Real GDP, PCE, and Non-Farm Payroll.

For the final part, I will do a sensitivity test on the S&P500 regressand with my the three regressors mentioned above. Modeling this relationship is very important because the scenarios created will be real scenarios from historical data which are likely to happen again over time. An example of this is rising unemployment or higher inflation both of which will probably occur within the next decade and these results will help show their impact on the S&P500.

There is an abundancy of research available on how macroeconomic models can impact indexes such as the S&P500. This paper will not incorporate recent or potential legislation but rather historical data on key indicators. Since the financial recession greater emphasis has been placed on stress-testing and understanding the relationship between key variables. In this paper the sensitivity analysis will be useful because it will show which parameters effect S&P the most and this process is helpful because it will prevent time from being spent on the non-sensitive ones within the model.

**CHAPTER 2: MATERIALS AND METHODS**

**2.1 Literature Review**

These topics have been discussed frequently over time. There has been significant literature in developing methods for forecasting individual stock prices and indices. There are many methods available. For this paper, the primary forecasting method will be vector autoregression. In the pursuit of forecasting stock values or indexes it is important to remember the Efficient Market Hypothesis. This hypothesis states that all the information within the stock market has already been observed and acted upon and therefore there are no arbitrage opportunities available. If this is true then one can describe the stock market as a random walk process where all returns are random and impossible to accurately forecast. This is not true according to much of the current literature which state a strong relationship between the stock market and macroeconomic variables such as GDP, interest rates, inflation, and nonfarm payroll.

In Marcellino, Porqueddu and Venditti (2015) they mention the importance of these macroeconomic variables on short-term forecasting models. They explain how they evaluate their point and density out-of-sample forecast accuracy of their models by studying the impact of macroeconomic information releases. Stock and Watson (2012) ties back in with the previously mentioned author because they study how macroeconomic time series variables provide improved forecast performance relative to the standard univariate autoregressions and small vector autoregressions. The main issue revolving around these macroeconomic variables is that they not updated frequently enough.

In Mukherjee and Bose (2014) they examine the stock market movements between India, Asia, and the United states. In this paper, MB uses a series of techniques such as vector autoregression, Granger causality, cointegration, and vector error-correlation models. From this they discovered that there is an information leadership from the U.S. market and that U.S. indices directly affect the Asian and Indian markets. There is a bi-directional causality relationship between these three nations. This is important because in this study, I will use VAR and Granger causality to discover similar relationships between my regressors and regressand.

**2.2 Data**

This research paper has a focus on analyzing the effects of changes in GDP, Non-Farm Payroll, and PCE on the S&P500 index as well as the tests and forecast methods aforementioned. The independent variables were pulled from Fred which is a database within the Federal Reserve Bank of St. Louis. This database provides the quarterly percent change values necessary.

The independent variable the S&P500 was extracted from Yahoo Finance. This variable was manually converted into am average monthly and then quarterly series. Following these procedures, it was converted into percent changer per quarter.

The start date of 7/1/1985 was selected to keep the dataset uniform and to also be able to include two structural breaks. One on July 1990 for the 1990-1991 recession and the second the financial recession on January 2008 for the 2008-2009 financial crisis.

**2.3 Statistical Tests**

For the first step of this paper I will perform a unit root test for stationarity within the dataset. Using the Dickey-Fuller Generalized Least Squares (DF-GLS) test, I will determine whether this data series is either stationary or non-stationary. The reason I will use the DF-GLS test is because it has a 75% test accuracy. The first thing the DF-GLS test will test for is if there is a zero-mean, single-mean, or a deterministic trend present. The DF-GLS test will determine if these model variables are invariant to time at the 5% level.

The null-hypothesis of this test is that the data series is non-stationary or a mean-diversion. The alternative hypothesis is that the dataset is stationary. Within my dataset, I determined that for all of these variables we will reject the null hypothesis of non-stationarity since the p-values in Table are all less than 0.05. I will also determine that there are not any deterministic trends present within my dataset.

The following test will be the Pearson Correlation test. This test measures the linear relationship between two variables to show its direction and strength. The range of values for the coefficients in this output is between -1 and +1. -1 will indicate that the two variables have perfect negative correlation, +1 will show perfect positive correlation and 0 will show perfect neutrality or no correlation between the two variables. The output for this test is in Table 2.

The next test is the Granger Causality test. This statistical test will check for the presence of a Granger causal relationship between any of the variables. A Granger causal relationship exists when one time series aids in the forecast of another. For this test, the null hypothesis is that there is no Granger causality. The alternative hypothesis is Granger causality exists within the data series.

In Table 3, I performed tests on 8 combination sets of these variables. When Pr>Chisq is greater than 0.05 we fail to reject the null hypothesis, which states that the independent variables do not granger cause the dependent variable. During the breakdown of the causal relationships the S&P500 and the independent variables we can see that there is a causal relationship between our dependent and Non-Farm Payroll and PCE. This relationship does not appear to exist between the S&P500 and Real GDP. Following this step, I will use the SIC/SBC criteria to determine the lag order by selecting the one with the lowest Schwarz Bayesian Criterion (SBC) which was lag 1 for this dataset.

**2.4 Forecast Models**

Before narrowing the forecast down the VAR model, I will implement an AR and ARIMA model. The purpose of this is to test the relative accuracy of these forecast by observing their Root Mean Squared Forecast Error (RMSFE). To minimize the RMSFE is to choose the forecast method which has the smallest deviation from the actual values. The easiest way to do this is through recursive techniques.

The Autoregressive Model(AR) is a stochastic process which estimates future values using lagged values. This process assumes that the historical values influence current and future values. For this paper, we will these lags 1-4 within the AR model and we will select the model with the lowest SBC. For this test it happened to be model AR(1).

Following this model, I perform an Autoregressive Integrated Moving Average model (ARIMA) which is a slightly more complicated stochastic process. The first step of this model is to determine p, d, and q. The autoregressive order is represented by p, d shows the degree of differencing, and q captures the moving average order. To do this we use the SCAN correlation method which will determine the correct value for the forecast. In this case it was the ARIMA (0,1) model.

I incorporated in a vector autoregression model to forecast the S&P 500 in a 2-year ahead forecast. In order to create an accurate forecast for the S&P500 returns in the long run, I had to establish a VAR framework. First I had to determine how many and which variables to include. After choosing the model with the lowest SBC, the model with the S&P500 dependent variable and Non-Farm Payroll dependent variable, I had to determine the appropriate number of lags to include. This was done in a similar procedure with the selection depending on minimizing the SBC. For this VAR model, it was determined to be lag 1. The values for the VAR forecast with the criteria aforementioned are available in Table 7.

For the final part of this paper I provided a conditioned forecast which helped show the S&P500’s sensitivity to changes in Real GDP, Non-Farm Payroll, and PCE. A conditional forecast is conducive for this purpose because it shows what will happen if only one specific independent variables changes. In my conditioned forecast, I learned that S&P is most strongly impacted by changes in Non-Farm Payroll and that changes in Real GDP and PCE have negligible effects. Therefore S&P500 is most sensitive to Non-Farm Payroll. Non-farm payroll has a 1:5 relationship with S&P500. Therefore a 1% increase in Non-Farm Payroll will lead to a 5% change in the S&P500. For the other two variables, a change of 10% is unlikely to cause a change greater than 1% in the S&P500.

**CHAPTER 3: RESULTS**

**3.1 Conclusion**

Within this study, I utilized several statistical tests such as the Dickey-Fuller Generalized Least Squares, Pearson Correlation test and Granger Causality test. These tests helped us understand the relationships between the variables selected for this dataset. Following this framework, I implemented a vector autoregression forecast to forecast two years ahead in time. After this test, I ran a conditional forecast to gain understanding of the sensitivity between the regressors and regressand. The results show how there is a Granger causal relationship between the S&P500 and Non-Farm Payroll and PCE. The primary things to note about these results is that the S&P500 is supposed to increase during the 2016 and 2017 years and that Non-Farm Payroll is a great indicator of the S&P500.

**REFERENCES**

Alexander, Carol and Elizabeth Sheedy, 2008, Developing a Stress Testing Framework Based on Market Risk Models, Journal of Banking and Finance 32, 2220-2236.

# Chu, Quentin C., Wen-liang Gideon Hsieh, and Yiuman Tse, 1999, Price Discovery on the S&P 500 Index Markets: An Analysis of Spot Index Index Futures, and SPDRs, International Review of Financial Analysis 8, 21-34.

# Gujarati, Damodar N., and Dawn C. Porter,2009, Basic Econometrics, 5th edition. New York, NY: McGraw-Hill/Irwin.

# Marcellina, Massimiliano, Mario Porqueddu, and Fabrizio Venditti, 2016, Short-term GDP Forecasting with a Mixed-Frequency Dynaic Factor Model with Stochastic Volatility, Journal of Business and Economic Statistics 23, 118-127.

Mukherjess, Paramita, and Suchismita Bose, 2008, Does the Stock Market in India Move with Asia?: A Multivariate Cointegration-Vector Autoregression Approach, Emerging Markets Finance and Trade 44, 5-22.

# Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H Stock, 2014, Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve, Journal of Economic Literature 52, 124-188.

# Silvia, John E., Azhar Iqbal, Sam Bullard, Sarah Watt and Kaylyn Swankoski,2014, Economic and Business Forecasting: Analyzing and Interpreting Econometric Results. Hoboken, NJ: Wiley.

# Stock, James H., and Mark W. Watson, 2007, Introduction to Econometrics 2nd Edition. Boston, MA: Pearson/Addison Wesley.

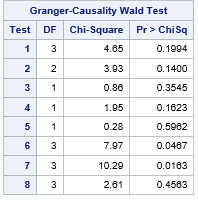
**APPENDIX**

**DF-GLS Results: Table 1**

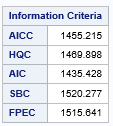
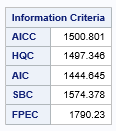
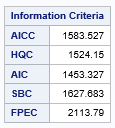
****

**Pearson Correlation Results: Table 2**

**Granger-Cause Results: Table 3**

****

**Table 7: Granger-Cause Lags**

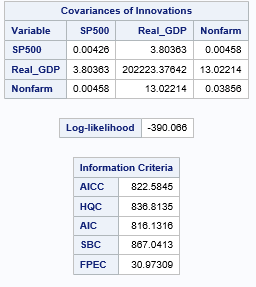


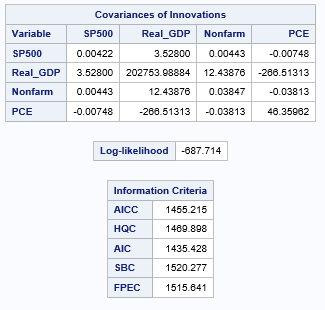
Lag (1)

Lag (3)

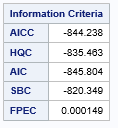
Lag (2)

**VAR: Table 4**





**VAR Lags: Table 5**

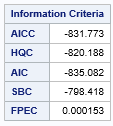
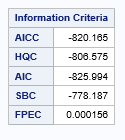
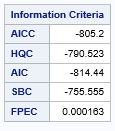


Lag (4)

Lag (3)

Lag (2)

Lag (1)



**VAR Forecast: Table 6**

