

12-1-2013

# Don't Judge the S&P 500 by its Cover: When Expectations Meet Regression

Daniel Freitag  
*George Fox University*

---

## Recommended Citation

Freitag, Daniel, "Don't Judge the S&P 500 by its Cover: When Expectations Meet Regression" (2013). *Student Scholarship - School of Business*. Paper 3.  
[http://digitalcommons.georgefox.edu/gfsb\\_student/3](http://digitalcommons.georgefox.edu/gfsb_student/3)

This Paper is brought to you for free and open access by the School of Business at Digital Commons @ George Fox University. It has been accepted for inclusion in Student Scholarship - School of Business by an authorized administrator of Digital Commons @ George Fox University.

Don't Judge the S&P 500 by its Cover:

When Expectations Meet Regression

Daniel Freitag

ECON 410

12/18/13

## 1. Introduction

The purpose of this paper is to investigate potential determinants of stock market prices utilizing regression techniques. In particular, this study aims to answer whether the health of the United States equities market is a representation of the well-being of the macro-economy. Accurately measuring the health of the stock market in its entirety is a difficult task. In light of this complication, financial analysts commonly refer to stock indices, such as the S&P 500 or the Dow Jones Industrial Average, as benchmarks representing trends in the equities market. For simplicity's sake, this analysis considers the S&P 500 index to be the United States' most precise baseline of equity performance.

There are two primary goals in this study. The first is to model the determinants of fluctuating stock prices in the short-run. For this reason, data is collected in monthly units intended to represent short-term fluctuations in the S&P price over time. The second is to model the expected impact of recessionary pressures on the performance of equities.

Investing in the equities market has become an integral component to the United States economy, especially in the context of retirement. It is an age where savings are no longer sufficient for the layman to obtain the capital necessary to comfortably retire because depository institutions today are offering exceptionally low – practically zero percent – interest rates on savings accounts. This trend has provoked individuals to seek higher returns elsewhere; they turn to the equities market.

There are two motives for researching the determinants of the S&P 500. First, a comprehensive understanding of the stock market could help mitigate the risk of investing. If an investor understands what causes the variation in a stock's performance, they could potentially make wiser investments, resulting in higher returns over time.

Second, it is important to question whether the S&P 500 is representative of a nation's economic health. If a statistical relationship were to present itself, the S&P 500 could be considered a snapshot of the United States economy at any given time. The intricacy of an entire economy would be encompassed into a single fluctuating stock quote that is readily available to the open public for analysis. My regression analysis provides results rejecting the hypothesis that the stock market is representative of economic well-being, at least in consideration of my chosen economic variables. However, statistical relationships were discovered between the S&P 500 and *individual* economic indicators.

## 2. Data Overview

The initial steps in this study were to identify potential determinants of the stock market's performance over time. All data collected are linked to the months between January 1955 and December 2010. The goal in using monthly units is to model short-term fluctuations in stock market prices.

The S&P 500, also known as the Standard & Poor's 500, is a stock market index comprised of 500 of the United States' largest companies taken from a wide variety of industries. Due to its unique weightings system, the S&P 500 is believed to be the best single gauge of the United States' equities market at a given point in time.<sup>1</sup> Yahoo Finance provides the historical data necessary to run a time-series regression on the performance of the S&P 500. The variable *S\_P\_Close* is the chosen regressand, which represents the closing price of the S&P 500 stock index at the end of each month considered in this study.

---

<sup>1</sup> S&P Dow Jones Indices. (n.d.). *S&P Dow Jones Indices*. Retrieved November 18, 2013, from <http://us.spindices.com/>

To represent the effects of monetary policy on the stock market's performance, I chose to regress the federal funds rate upon the S&P 500. The federal funds rate is defined as the rate at which depository institutions, such as banks, credit unions, etc., trade Federal Reserve funds.<sup>2</sup> When a depository institution has excess cash reserves, it can lend that surplus to other banks that are lacking sufficient cash reserves. These transactions are weighted by the Federal Reserve and are averaged to estimate the effective federal funds rate for a given time period. Given that the federal funds rate affects the rate at which banks are lending to firms, individuals, and organizations, it indirectly has the power to influence entrepreneurial activity within the stock market. Thus, the federal funds rate should be inversely proportional to the S&P 500 price. The variable *Effective\_FFR* represents the weighted average federal funds rate for the given time period, and the relevant data was collected from the Federal Reserve website.

Investor confidence in the stock market significantly impacts the variation in short-term market prices. One measure of investor confidence, or pessimism, is the average level of trading activity taking place in the equities market. Yahoo Finance reports the average "volume" of daily trades taking place over a month's time, and the variable *Average\_Volume* is included to estimate the impact of trading activity on stock market prices. Intuitively, a high volume of trading should be correlated with confidence in the market, as shown by increasing stock prices. However, trading activity can also increase in fear of economic turmoil as investors sell their assets to protect their savings, which would indicate a non-linear relationship between volume and equity prices. To test this theory, the variable *Average\_Volume* is included to statistically model the relationship between trading activity and stock market prices.

---

<sup>2</sup> Open Market Operations. (n.d.). *Board of Governors of the Federal Reserve System*. Retrieved November 18, 2013, from <http://www.federalreserve.gov>

An unemployed individual is defined by the Bureau of Labor Statistics as someone who does not have a job, has actively searched for work in the past four weeks, and is currently available for work.<sup>3</sup> The unemployment rate is then computed by estimating the total amount of unemployed individuals in the United States as a percentage of the total labor force. I include the variable *Unemployment\_Rate* to model the impact of unemployment shocks on the S&P 500 stock price as a result of investor uncertainty in the job market.

The true determinants of stock market prices are the market forces of supply and demand for stocks. These forces fluctuate on a daily basis, causing stock prices to be extremely volatile. Major variations in stock prices are commonly linked with the performance of macro-economic output, or more specifically a nation's GDP. Since 1929, the National Bureau of Economic Research (NBER) has mapped the United States business cycle over time.<sup>4</sup> Trends of economic growth are known as expansionary periods, whereas periods of economic contraction are defined as recessionary periods. I include a dummy variable, *recessionary\_period*, to model the impact of being in a recessionary state in the business cycle.

The NBER defines a recession as “a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales.”<sup>5</sup> There have been two major recessions in the past fifteen years, both the result of rapidly increasing asset prices commonly referred to in the economic community as “bubbles.” In January 2001 the tech (also known as the dot com) bubble burst, causing a recession lasting roughly eight months. As a result of a booming housing market, the economy recovered very quickly. Unfortunately, this expansionary

---

<sup>3</sup> U.S. Bureau of Labor Statistics. (n.d.). *U.S. Bureau of Labor Statistics*. Retrieved November 18, 2013, from <http://www.bls.gov/>

<sup>4</sup> The National Bureau of Economic Research. (n.d.). *The National Bureau of Economic Research*. Retrieved November 18, 2013, from <http://www.nber.org/>

<sup>5</sup> [www.nber.org](http://www.nber.org)

period marked yet another bubble that was just about ready to burst. In 2008, the real estate and equities markets plummeted, and the United States entered another recession. To model the impact of the two most recent recessions on our economy in relativity to normal recessionary periods, two dummy variables are included: *Rec\_Tech\_Bubble* and *Rec\_House\_Bubble*. Values of “1” are assigned to months falling within the two most recent recessions.

The Federal Reserve Bank of St. Louis provides CPI data representing fuel, electricity, and gasoline for the United States.<sup>6</sup> The explanatory variable *CPI\_Energy\_Index* is included to model the impact of energy prices on the price of the S&P 500 index. I gathered data from a CPI index representing the rising costs of fuel, electricity, and gasoline in the United States. These are common energy costs associated with doing business, which counteract profits that are linked to stock prices. This data set is intended to investigate a relationship between rising energy costs and stock market prices, and represents the percentage change (inflation) of energy prices relative to the preceding month. Theoretically, stock prices should positively correlate with inflation. Otherwise, the real value of the stock would decrease as a result of the deflated value of the dollar.

*Table 1: Summary Statistics*

<b><u>Variable</u></b>	<b>Mean</b>	<b>Median</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Std. Dev.</b>	<b>Variance</b>
<i>ΔS_P_Close</i>	2.753	3.065	-197.610	132.160	34.677	1,202.494
<i>ΔEffective_FFR</i>	-0.019	0.000	-6.630	3.060	0.604	0.365
<i>CPI_Energy_Index</i>	0.433	0.277	-17.975	11.468	2.667	7.113
<i>ΔUnemployment_Rate</i>	0.005	0.000	-0.700	0.900	0.181	0.033
<i>Recessionary_Period</i>	0.136	0.000	0.000	1.000	0.344	0.118
<i>ΔAverage_Volume</i>	8.67E+6	66,500.00	-1.66E+9	2.63E+9	2.93E+9	8.58E+18

<sup>6</sup> *The Federal Reserve Bank of St. Louis*. Retrieved November 18, 2013, from <http://research.stlouisfed.org/fred2/>

### 3. Methodology

Once I compiled my data, I was able to run my base regression utilizing robust standard errors and the ordinary least squares (OLS) method. The initial results suggest that *Unemployment\_Rate* and *Recessionary\_Period* are negatively correlated with the price of the S&P 500 index. However, these preliminary results fail to account for inconsistencies in the data. When running a time series regression, it is essential to test for non-stationarity in the considered variables. To do this, I utilized the Augmented Dickey-Fuller to test for unit roots in my data. Conclusive results were found suggesting that the variables *S&P\_Close*, *Effective\_FFR*, *Average\_Volume* and *Unemployment\_Rate* are non-stationary. To compensate, these variables were first-differenced to adapt the model to represent the change in the S&P 500 price relative to the preceding period as a function of the federal funds rate, unemployment, trading activity, being in a recessionary state, and the inflation of energy prices.

Table 2: ADF Test

<b>Augmented Dickey Fuller Test for Non-Stationarity with 1 Lag</b>		
	<b>P - Value with Constant</b>	<b>P - Value with Constant and Trend</b>
<b>S_P_Close</b>	0.86	0.511
<b>Effective_FFR</b>	0.18	0.02
<b>Unemployment</b>	0.76	0.99
<b>Average_Volume</b>	0.72	0.33
<b>Reessionary_Period</b>	6.79E-06	6.10E-05
<b>CPI_Energy_Index</b>	7.88E-29	6.94E-33
Note: A high p-value indicates that we cannot reject the null hypothesis of the ADF test that the data set is non-stationary		



**Model (1)**

$$\Delta S\&P \text{ Price} = B_0 + B_1 \Delta FFR + B_2 \text{ CPI} + B_3 \Delta \text{Unemployment} + B_4 \text{ Recession} + B_5 \Delta \text{Volume} + u_t$$

(Where  $\Delta$  denotes a change in the specified variable from the preceding period, and  $u_t$  is the stochastic error term)

The base regression was expanded to investigate diverging trends occurring during recessions. An interaction term was added between volume and recessions, and is notated by the variable *Volume\_Recessionary*. This measure was taken to assess whether trading activity affects the S&P price differently during recessions. Binary variables were then added to represent the recessions resulting from the two most recent asset bubbles: the tech and housing crises. A value of 1 was assigned to the recessionary periods following the bubbles' pop. The goal was to model the expected outcome of asset bubbles on the equities market. The underlying assumption is that we can generalize the impact of an economic meltdown on the United States equities market.

**Model (2)**

$$\Delta S\&P \text{ Price} = B_0 + B_1 \Delta FFR + B_2 \text{ CPI} + B_3 \Delta \text{Unemployment} + B_4 \text{ Recession} + B_5 \Delta \text{Volume} + B_6 \Delta \text{Volume\_Recession} + B_7 \text{ Rec\_Tech\_Bubble} + B_8 \text{ Rec\_Housing\_Bubble} + u_t$$

When analyzing the S&P 500 quote over the past 60 years, there appears to be varying trends in the data depending on what time period is being considered. In the years between 1955 and 1975, the price of the S&P 500 remained fairly constant. For this reason, the sample for this regression was shrunk to include only the months between January 1975 and December 2010 to make the model more representative of current trends. The late 1970s are when stock prices first began to consistently increase over time. This upward trend remained constant until 2001, when

the United States experienced the repercussions of the tech bubble bursting. Since 2001, the United States has experienced two major recessions, and the returns on the S&P 500 fluctuate much more drastically than in the past. To investigate whether a new trend exists in the model, I conducted a Chow test in GRETL to test for a break in the data following March 2001. The test added dummy variables representing a split in the data to compensate for any diverging trends that present themselves. These binary variables were added to the base model, providing a third regression to be used in this analysis.

*Table 3: Chow Test*

<b>Chow Test for Structural Break at Observation 2001:3</b>	
Null Hypothesis: No Structural Break	
Asymptotic Test Statistic: Chi-Square	9.41
P - Value	0.15

***Model (3)***

$$\Delta S\&P\ Price = B_0 + B_1 \Delta FFR + B_2 CPI + B_3 \Delta Unemployment + B_4 Recession + B_5 \Delta Volume + B_6 sd_{\Delta FFR} + B_7 sd_{\Delta Unemployment} + B_8 sd_{\Delta Volume} + B_9 sd_{Recessionary\_Period} + B_{10} sd_{CPI\_Energy\_Index} + u_t$$

(Where "sd" denotes "split dummy," representing new trend lines after the tested break)

When using time-series regression, it is important to consider whether causality is present within the data. To test for causality, I ran a vector autoregression (VAR) model with twelve lags representing one full year's worth of data. In the model, the variables  $\Delta S\_P\_Close$ ,  $\Delta Effective\_FFR$ , and  $\Delta Average\_Volume$  were assumed to be endogenous, and the variables

$CPI\_Energy\_Index$ , and  $\Delta Unemployment$  exogenous. The Granger Causality test results can be found in section 4 of this paper.

#### 4. Results & Interpretations

Several statistical relationships were discovered between the dependent variable – the change in the S&P price – and the chosen regressors. However, the economic and statistical significance of these relationships vary depending on the statistical model referenced. The Chow test in regression (3) suggest that, following March 2001, we can say with 85% confidence that a statistical break occurred in our estimated trend lines, and that volume displays a statistically significant diverging trend at a 95% confidence level. The relevant output from the four estimated models is displayed below:

*Table 4: VAR Results*

<b>VAR Granger Causality Tests</b>		
(Dependent Variable in Bold)		
	F-Test	P-Value
<b>d_S_P_Close</b>		
All lags of d_S_P_Close	0.46	0.94
All lags of d_Effective_FFR	0.88	0.56
All vars, lag 12	0.50	0.68
<b>d_Effective_FFR</b>		
All lags of d_S_P_Close	0.85	0.59
All lags of d_Effective_FFR	5.38	0.00
All lags of d_Average_Volume	0.99	0.46
All vars, lag 12	0.17	0.91
<b>d_Average_Volume</b>		
All lags of S_P_Close	1.01	0.44
All lags of d_Effective_FFR	0.75	0.70
All lags of d_Average_Volume	1.53	0.11
All vars, lag 12	1.1	0.35

Table 5. Output

	Dependent Variable: $\Delta$ S&P 500 Price					
Regression:	(1)		(2)		(3)	
<i>Const.</i>	4.57**	(1.83)	4.54**	(1.78)	4.07**	(1.62)
$\Delta$ <i>Effective_FFR</i>	-2.13*	(1.17)	-1.24	(1.21)	-1.91*	(1.12)
$\Delta$ <i>Unemployment</i>	4.12	(9.20)	5.19	(8.72)	7.28	(5.99)
$\Delta$ <i>Average_Volume</i>	-2.91E-8***	(6.22E-9)	-1.83E-8*	(1.03E-8)	1.14E-7	(7.18E-8)
<i>Recessionary_Period</i>	-14.38**	(6.36)	-5.32	(3.85)	-5.55*	(2.87)
<i>CPI_Energy_Index</i>	0.79	(1.20)	0.71	(1.02)	-0.35	(0.95)
<i>Volume_Recession</i>			-1.59E-08	(1.12E-8)		
<i>Rec_Tech_Bubble</i>			-2.23	(16.01)		
<i>Rec_Housing_Bubble</i>			-27.51*	(15.87)		
<i>sd_Δ_Effective_FFR</i>					41.57	(26.66)
<i>sd_Δ_Unemployment</i>					-22.39	(35.13)
<i>sd_Δ_Average_Volume</i>					-1.46E-07**	(7.25E-8)
<i>sd_Recessionary_Period</i>					-4.50	(12.45)
<i>sd_CPI_Energy_Index</i>					0.96	(1.57)
$R^2$	0.080		0.102		0.127	
<i>N</i>	432		432		432	
<i>Sample Range</i>	Jan 1975 - Dec 2010		Jan 1975 - Dec 2011		Jan 1975 - Dec 2012	
<i>F-Statistic</i>	8.23		17.09		6.68	
Notes: * Denotes significance at the 10% level, ** at the 5% level, and *** the 1% level sd denotes "split dummy" (Standard errors in parentheses)						

The results of the VAR model provide no statistical evidence of bidirectional granger causality problems within the data set, as shown by the relevant F-statistics. It was found, however, that lags of  $d\_Effective\_FFR$  do in fact granger cause  $d\_Effective\_FFR$  at time  $t$  (the current period under consideration). These preliminary results were unexpected because it was anticipated that the federal funds rate would display bidirectional causality in the data set. The Federal Reserve's usage of monetary policy depends on the current status of the economy; an indication I believed would translate to the stock market.

Two of the three specified models provide results suggesting that the federal funds rate is negatively correlated with stock market prices. For a 1% increase in the federal funds rate, the S&P price declines by about \$2.00, plus or minus its standard error. Relative to the long run average change in the S&P 500 price - \$2.75 – this value is economically significant in that a 1% increase in the federal funds almost fully counteracts this expected growth in stock prices. However, a noteworthy weakness to this result is that the federal funds rate rarely fluctuates by an economically significant margin. In this regard, the federal funds rate is not generally considered an economically significant determinant of fluctuating stock prices in the short-run. In light of this conclusion, monetary policy influencing the federal funds rate is not an economically viable option to influence stock market performance.

In all three models, trading activity appears to be negatively correlated with stock prices. This negative trend remains constant in each of the three model specifications, disproving the theory that volume positively correlates with improving stock prices in the absence of recessionary periods. In the second model, it was discovered that the interaction term is negatively correlated with the S&P 500 price at an 85% confidence level. These findings imply that, *ceterus paribus*, a high level of trading activity negatively affects stock market growth in

both expansionary and recessionary periods. Additionally, if we can accept an 85% confidence level, we can conclude that recessionary periods have a slightly larger negative impact on the S&P 500 price relative to expansionary periods. Marginally speaking, however, a one-unit increase in average trading activity is not economically significant. This finding suggests that volume describes only a minor component of fluctuating stock prices. In consideration of the diverging trend found in the third model, volume has become slightly more economically significant, but not enough to reject our previous conclusions.

This study provides no statistical evidence of a relationship between unemployment and the price of the S&P 500. Contrary to the expectations – that unemployment would be negatively correlated with stock prices due to uncertainty in the job market - the estimated coefficient for unemployment is positive, suggesting that higher unemployment yields better performance from stocks. This coefficient, however, remains highly insignificant in terms of its statistical relevance as a determinant of the S&P 500 price. This result unexpectedly disproves the theory that positive unemployment shocks cause investors to lose confidence in the market.

As anticipated, recessionary periods negatively correlate with S&P prices. The purpose of including recessionary periods was to statistically estimate the short-term impact of being in a recession on the S&P 500's performance. According to the base model, being in a recessionary state for a month causes the S&P 500 price to decline by an economically significant value of \$14.00. However, this figure encompasses every recession considered in my time series, including the statistical outliers of economic meltdowns, which overestimates the coefficient. Therefore, the result found in the third model - that being in a recession causes the S&P 500 to fall by \$5.50 - is speculatively a more reliable figure when considering the marginal impact of a generic recessionary period in the business cycle on stock prices.

The second model aimed to compensate for this weakness by including dummy variables representing the 2001 and 2008 recessions. *Rec\_Tech\_Bubble* proved to be statistically insignificant, while *Rec\_Housing\_Bubble* was significant at a 92% confidence level. This variation in statistical significance is likely attributed to the 2001 recession being a much shorter recession relative to recession beginning in 2008. The recession following the tech bubble lasted roughly eight months, whereas the recession following the housing bubble lasted a full eighteen. Under the assumption that we can generalize the impact of being in an economic meltdown, being in a recession for one month causes S&P 500 to fall by \$27.00, an economically significant value. On the other hand, this coefficient has an extremely high standard error. This finding suggests that accurately estimating the expected marginal impact of a market crash on stock market performance is statistically challenging. Although we may not be able to precisely measure that impact, this analysis quantitatively proves how drastically asset bubbles can negatively influence the equities market.

Unexpectedly, there is no statistical relationship between the inflation of energy prices and the change in the S&P 500 price. Therefore, it can be assumed that the inflation of energy prices is not a statistical or economic determinant of stock market prices. A weakness in this conclusion is that the energy CPI considered in this analysis does not encompass the inflationary pressures on other types of goods in the United States economy. To further investigate the impact inflation has on the price of the S&P 500, other CPI indices should be considered before adopting my conclusion that inflation does not statistically impact the price of equities.

## 5. Conclusions

A noteworthy weakness to my model is the presence of an extremely low  $R^2$  value in the regressions. This low  $R^2$  suggests that the macro-economy is not the primary force determining equity performance. As a result, we can conclude that the S&P 500 should not be considered a "snapshot" of the entire macro-economy. This paper provides conclusive evidence that the stock market is representative of only certain aspects of macro-economic health, such as interest rates and recessionary cycles, but fails to respond to shocks in unemployment and inflation, which are known to be leading indicators of economic well-being. In conclusion, there is more to the S&P 500 than meets the eye. This study has shown that intuition alone cannot determine stock market prices. The market responds to external forces affecting the supply and demand of stocks, some of which are considered in my model, but many that are left undiscovered. Regression has once again taught the valuable economic lesson that causation is not as obvious as one might think.



## References

Open Market Operations. (n.d.). Board of Governors of the Federal Reserve System. Retrieved November 18, 2013, from <http://www.federalreserve.gov>

S&P Dow Jones Indices. (n.d.). S&P Dow Jones Indices. Retrieved November 18, 2013, from <http://us.spindices.com/>

The Federal Reserve Bank of St. Louis. Retrieved November 18, 2013, from <http://research.stlouisfed.org/fred2/>

The National Bureau of Economic Research. (n.d.). The National Bureau of Economic Research. Retrieved November 18, 2013, from <http://www.nber.org/>

U.S. Bureau of Labor Statistics. (n.d.). U.S. Bureau of Labor Statistics. Retrieved November 18, 2013, from <http://www.bls.gov/>