Alex Bradford

Time Series

Fall 2014

Time Series Student Project – S&P 500 Index Prices

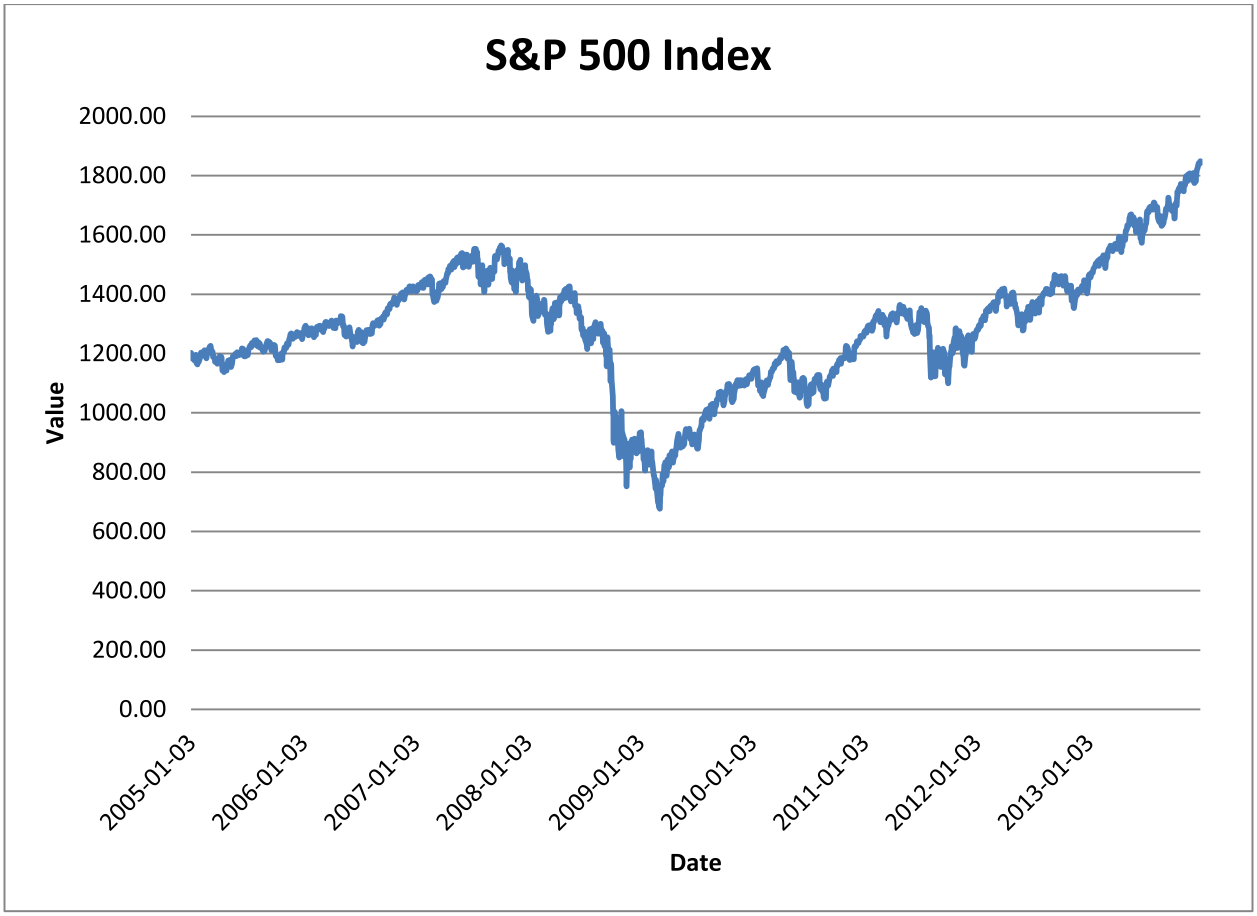
Introduction

The purpose of this project is to construct a time series regression model on the daily value of the S&P 500 Index. I will be investigating whether past values of this index fund could be used to predict future fund movements using 2265 daily market valuations over a nine year period.

Data

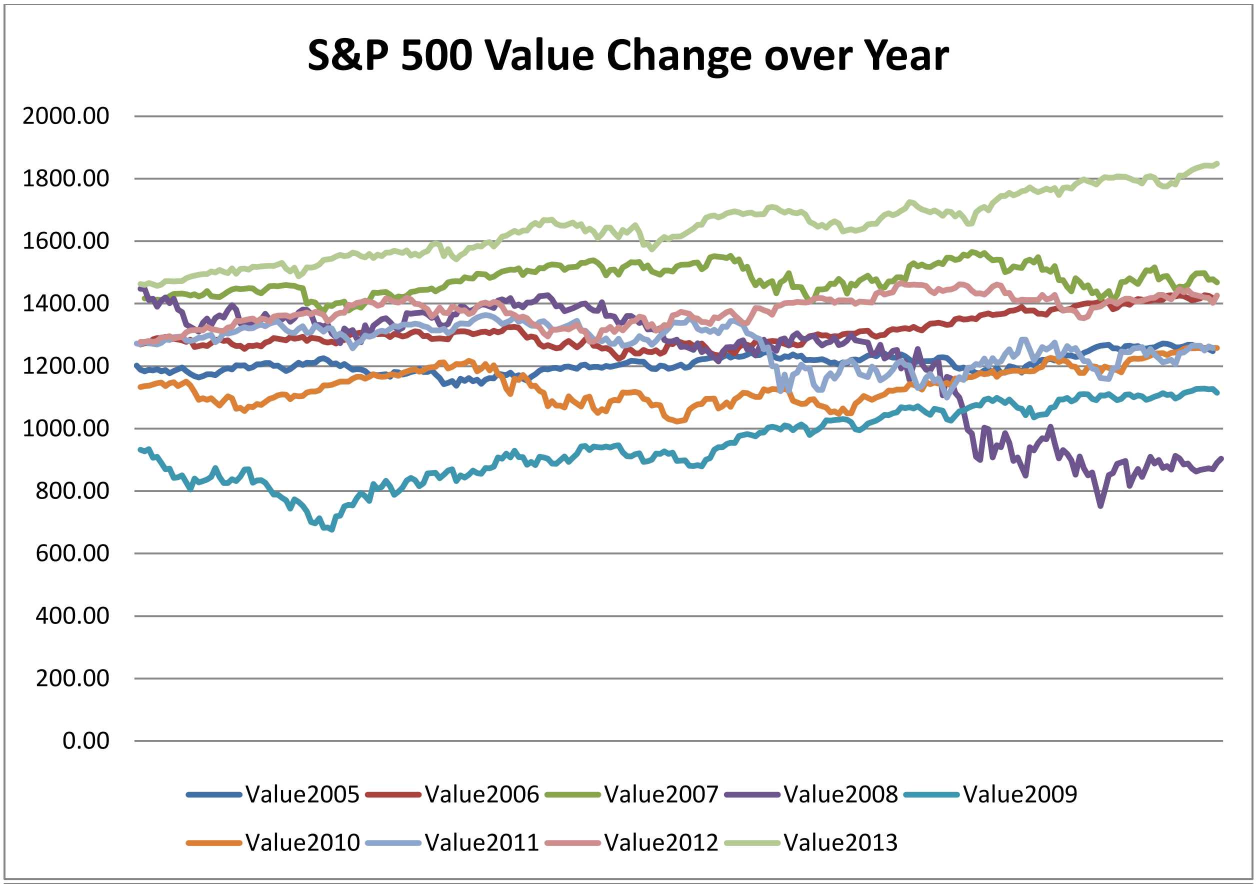
The S&P 500 is a stock market index of the 500 largest publicly traded companies zon the NYSE and NASDAQ stock exchanges. Each of the 500 companies listed on the index are weighted according to their relative market capitalizations (share price * shares outstanding), so larger companies have a larger influence on the price of the index fund. This index is commonly used as a measure of business investment when judging the health of the overall economy. When the S&P 500 Index increases over a comprehensive period of time, it means that businesses are growing and have an optimistic outlook on the future. Conversely, when the S&P 500 Index decreases over an extended period, it indicates that both businesses and consumers have a more cynical outlook on the future of the economy. I have gathered the S&P 500 Index data from the Federal Reserve Bank of St. Louis's economic research database.

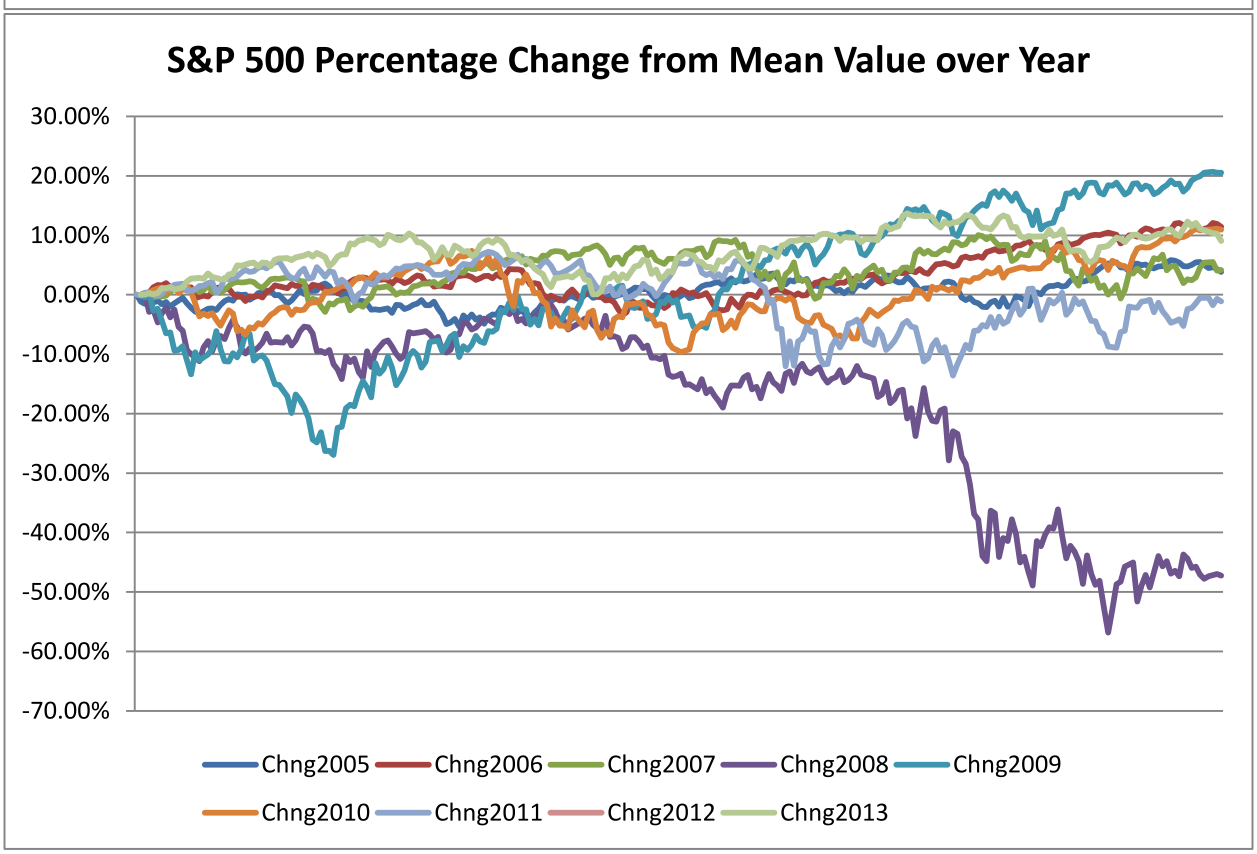
http://research.stlouisfed.org/fred2/series/SP500/downloaddata

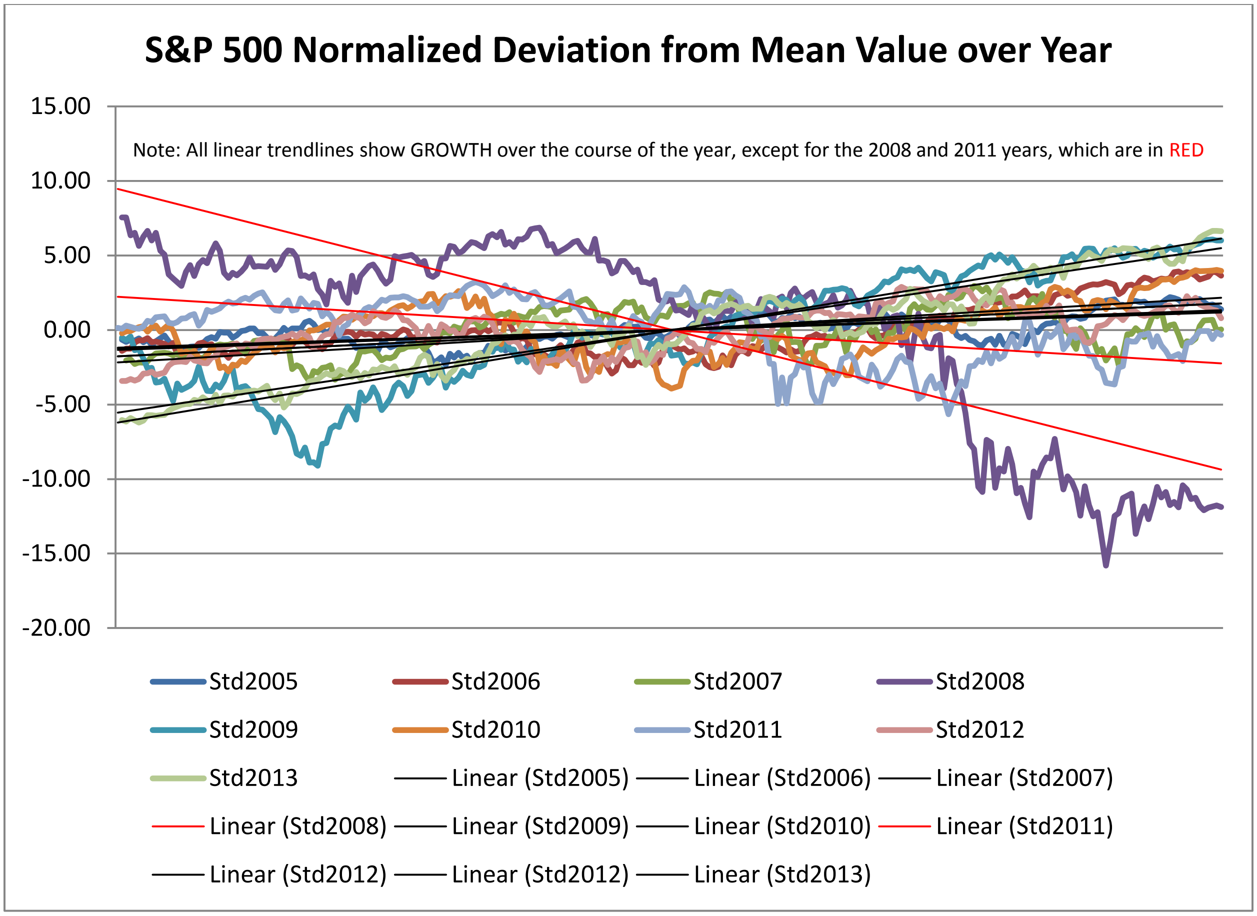


Analysis

Looking at the data, there is an upward trend over the course of each year in the S&P 500 Index's value. This makes sense as the S&P 500 Index is a good measurement of the state and size of the overall economy and the economy tends to grow over the course of each, individual year. The large exception to this observation, is the decrease in value in 2008. This can be explained as a direct result of the Financial Crisis where business and consumer confidence hit historic lows. 2011 also saw a decrease in the index fund's overall value due to widespread fear of a double-dip recession. We can view these phenomenon in greater detail in the next three graphs.


S&P 500 Value Change over Year
2000.00
1800.00
1600.00
1400.00
1200.00
1000.00
800.00
600.00
400.00
200.00
0.00
Value2005
Value2006
Value2007
Value2008
Value2009
Value2010
Value2011
Value2012
Value2013


S&P 500 Percentage Change from Mean Value over Year
30.00%
20.00%
10.00%
0.00%
-10.00%
-20.00%
-30.00%
-40.00%
-50.00%
-60.00%
-70.00%
Chng2005
Chng2006
Chng2007
Chng2008
Chng2009
Chng2010
Chng2011
Chng2012
Chng2013



In this last graph, I looked at the number of standard deviations the S&P 500 Index was away from it's mean value for a given year. In most years, the index's value fluctuates between 2-3 standard deviations from it's average yearly value, the large exception being 2008 due to the Financial Crisis, which also carried over to the first half of 2009. As we can see from the trendlines in the graph, the deviations from the average value transition from positive to negative over time, which suggests the value increases over a typical year. Again, the exceptions being 2008 and 2011. This data set should be appropriately modeled by a stationary time series model.

Correlations	*Y2005*	*Y2006*	*Y2007*	*Y2008*	*Y2009*	*Y2010*	*Y2011*	*Y2012*	*Y2013*
Y2005	1								
Y2006	0.466425	1							
Y2007	0.030923	0.063273	1						
Y2008	-0.63216	-0.88102	-0.2095	1					
Y2009	0.618355	0.689069	0.445649	-0.78414	1				
Y2010	0.176672	0.790447	-0.07267	-0.61056	0.361532	1			
Y2011	-0.49947	-0.40747	-0.07207	0.483164	-0.67905	-0.10669	1		
Y2012	0.320362	0.611171	0.175405	-0.56888	0.532514	0.450232	-0.55649	1	
Y2013	0.688905	0.71026	0.457347	-0.82862	0.870566	0.446939	-0.48779	0.576031	1

However, after more analysis, there is very little correlation between adjacent years worth of data. The correlation between the 2005 and 2006 values of the S&P 500 yielded 0.466 and comparing 2006 and 2007 yielded an even lower 0.063. The most highly correlated adjacent years worth of data are 2012 and 2013 with a value of 0.576. However, even this value is relatively low which implies that the daily movement of the S&P 500 index's value from the year prior does not influence this year's movement. This makes sense because what happened to the market one year prior should be completely independent from this year's market movement. Before delving into the regression analysis, it is worth noting that we are looking at very granular historical data – daily values, not monthly averages – so our R-Squared values are not optimally representative. I show the 2005-2006 and 2012-2013 regressions because they are the most highly correlated adjacent years' datasets. I also look into the 2008-2009 regression because these two years have the greatest absolute value of correlation.

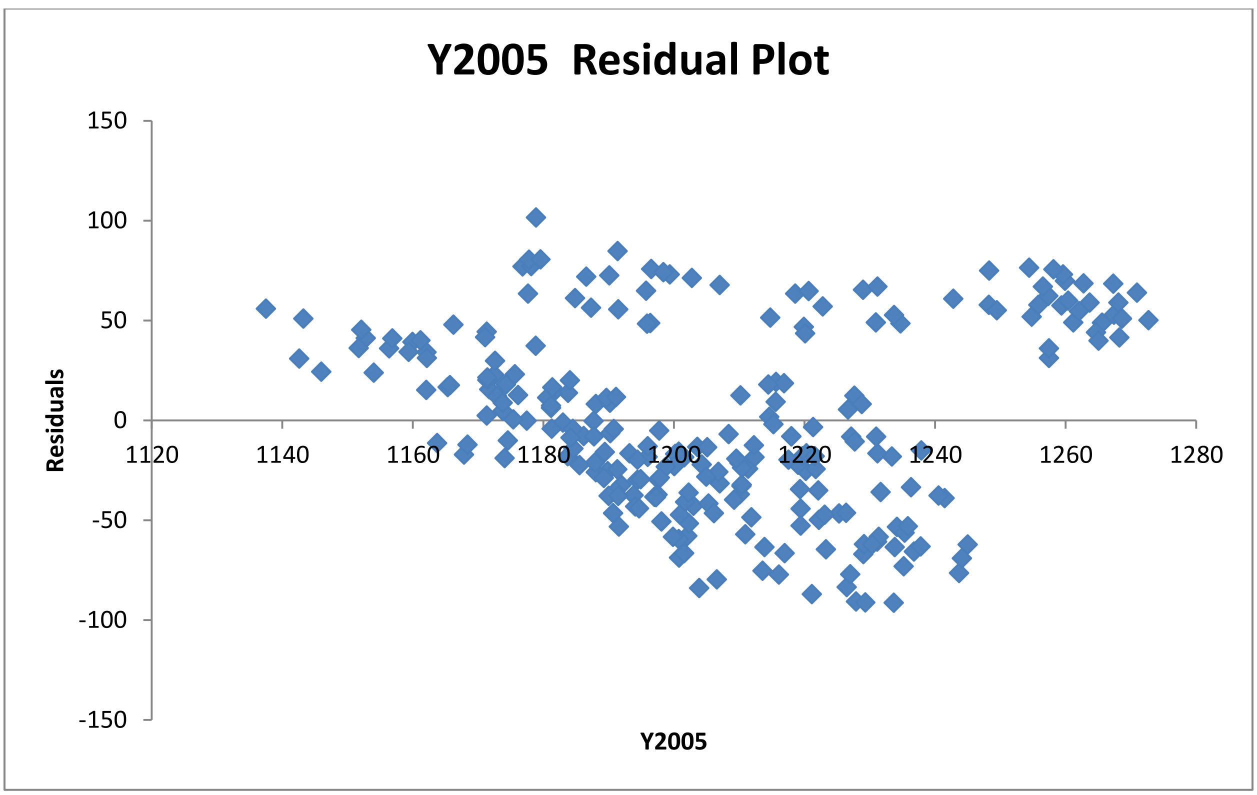
2006 Predictive Model Using 2005 Data

Regression Statistics	
Multiple R	0.466425425
R Square	0.217552677
Adjusted R Square	0.214519935
Standard Error	45.5607948
Observations	260

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	148905.7992	148905.7992	71.73465741	1.89157E-15
Residual	258	535552.7938	2075.786022		
Total	259	684458.593			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	337.8262929	114.8758444	2.94079486	0.003570914	111.6126208	564.039965
Y2005	0.805581552	0.095114121	8.46963148	1.89157E-15	0.618282694	0.992880411



Even though, these two years worth of data are more correlated than other years, the low R-Square value suggests this is a very weak predictive model.

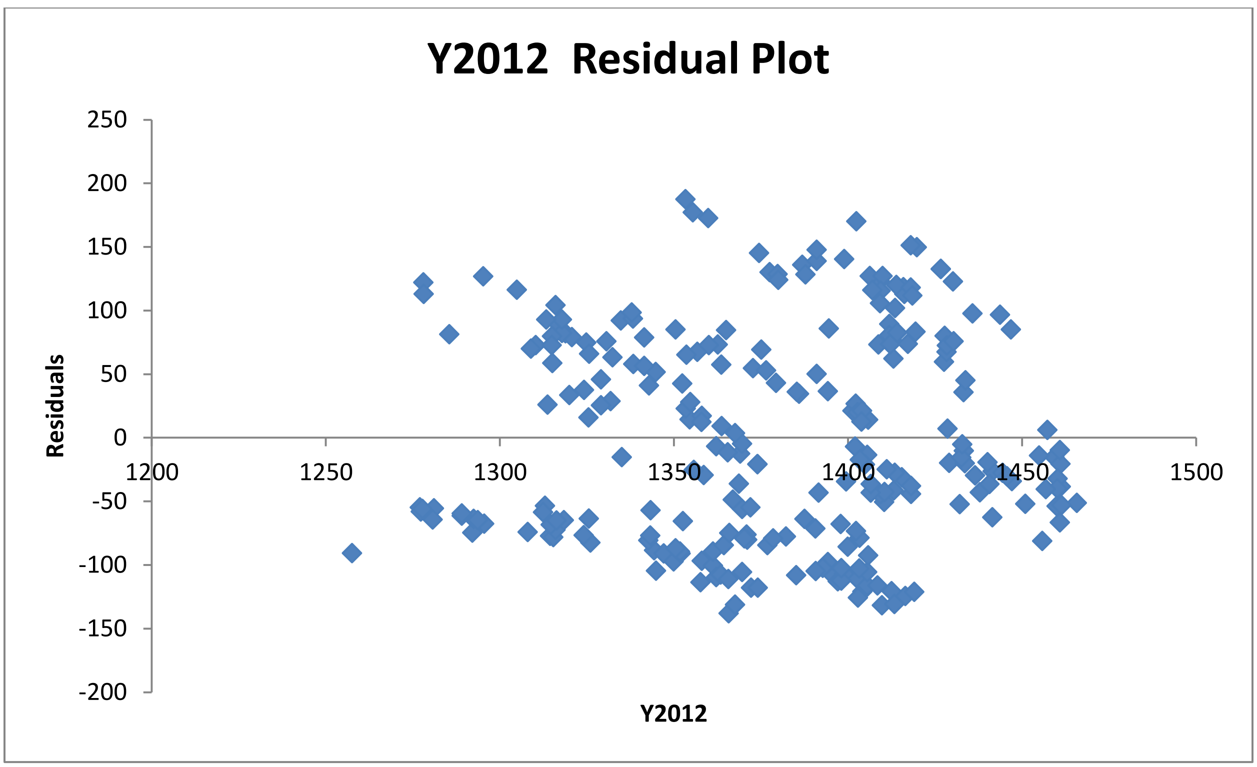
2013 Predictive Model Using 2012 Data

Regression Statistics	
Multiple R	0.576031408
R Square	0.331812183
Adjusted R Square	0.329222308
Standard Error	82.00692867
Observations	260

ANOVA

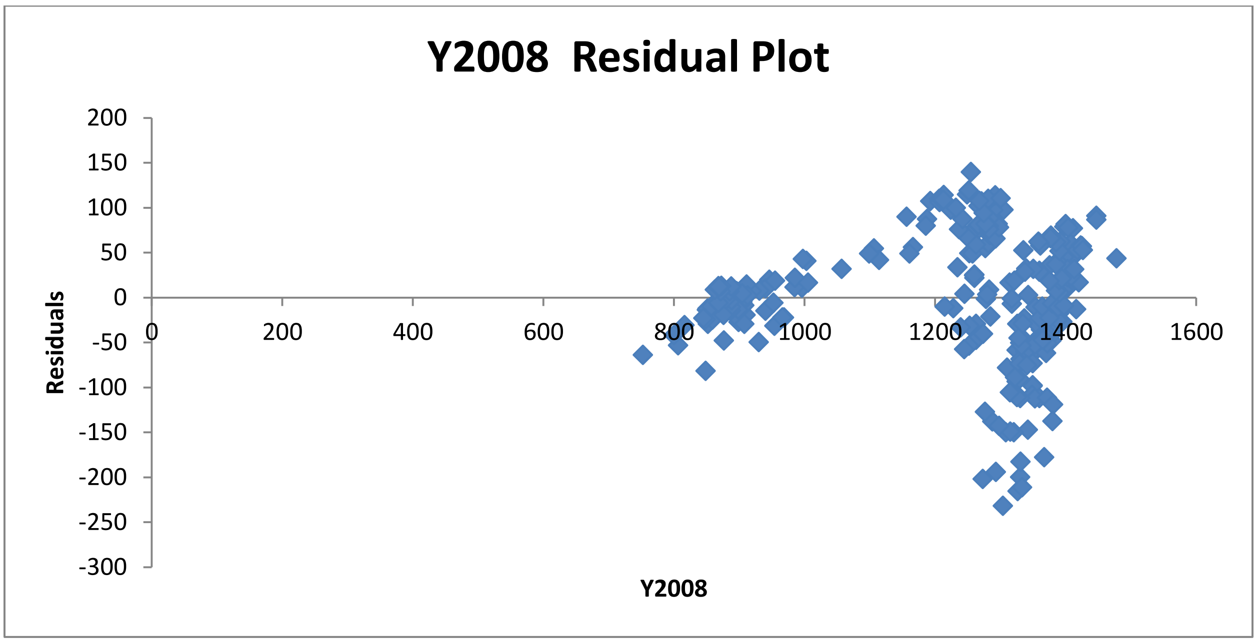
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	861617.6277	861617.6277	128.1189827	2.20737E-24
Residual	258	1735085.178	6725.136349		
Total	259	2596702.806			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-50.54357766	149.6129843	-0.337828818	0.735766904	-345.1616738	244.0745185
Y2012	1.227608384	0.108455881	11.31896562	2.20737E-24	1.014036913	1.441179854



The two most highly correlated years worth of data still has a low R-Square value which suggests this is also a weak model.

2009 Predictive Model Using 2008 Data						
Regression Statistics						
Multiple R	0.784136233					
R Square	0.614869631					
Adjusted R Square	0.613376878					
Standard Error	71.57981278					
Observations	260					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	2110455.019	2110455.019	411.9030274	2.20638E-55	
Residual	258	1321906.756	5123.669597			
Total	259	3432361.775				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1526.83857	28.91044846	52.81269063	2.4595E-140	1469.908075	1583.769066
Y2008	-0.473987324	0.023354428	-20.29539424	2.20638E-55	-0.519976896	-0.427997752



Ironically, the "least" correlated two years worth of data has the most "robust" predictive model according to regression analysis – this has the highest R-Square value of all of the adjacent years' data. However, this is an exception to the upward trend in S&P 500 Index value over the course of a year (this model has a negative beta value), and would definitely not be a good fit for any of the other years used in this study.

Even though the 2013 predictive model using 2012's data was weak at best, it was still the best of the regression models and I will use these two year's worth of data in my time series analysis.

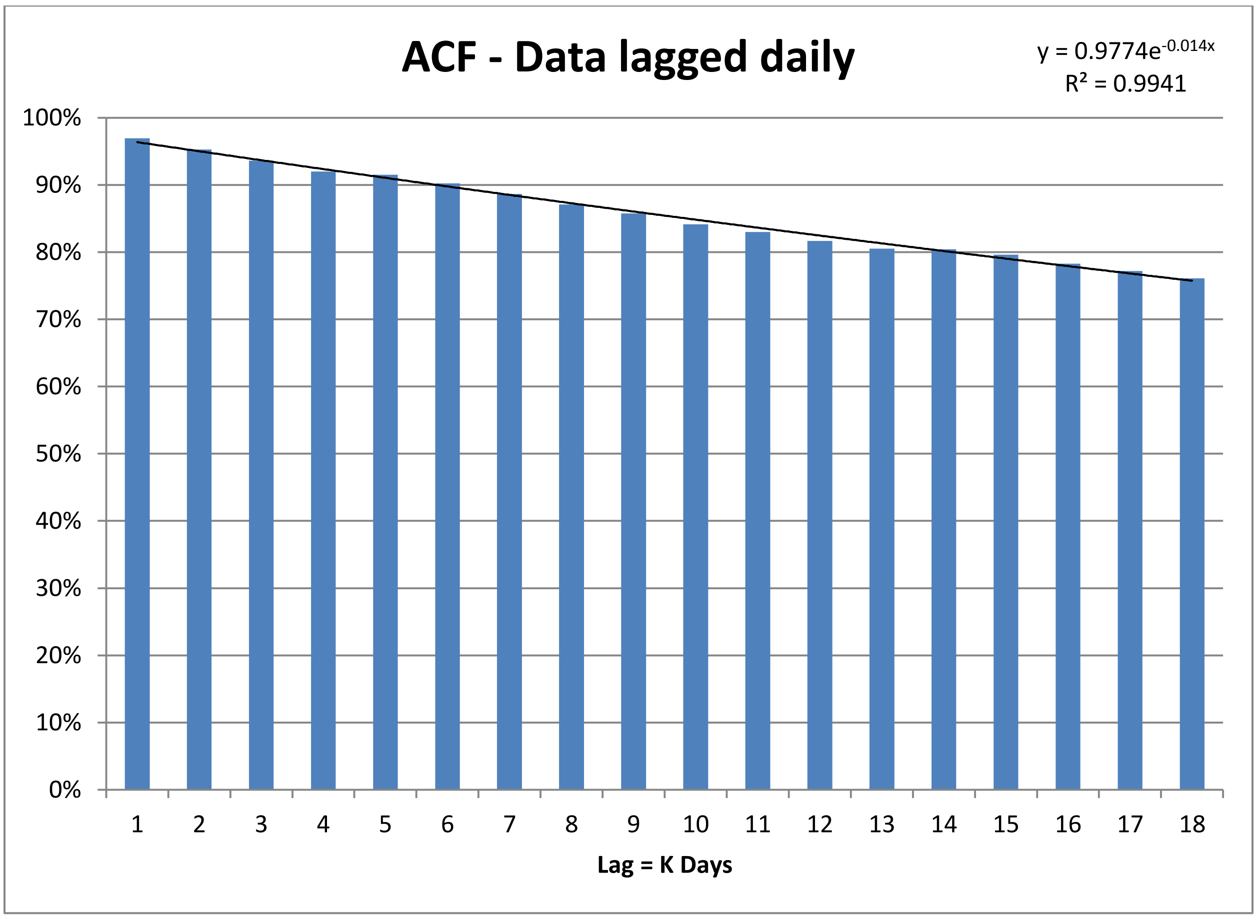
Constructing an ACF using the following equations:

$$ACF(k) = \rho_k = \frac{\gamma_k}{\gamma_0}$$

$$\gamma_k = E\left[(x_t - \mu)(x_{t-k} - \mu_{t-k})\right] = E\left[x_{t-k}x_t\right] - \mu^2$$

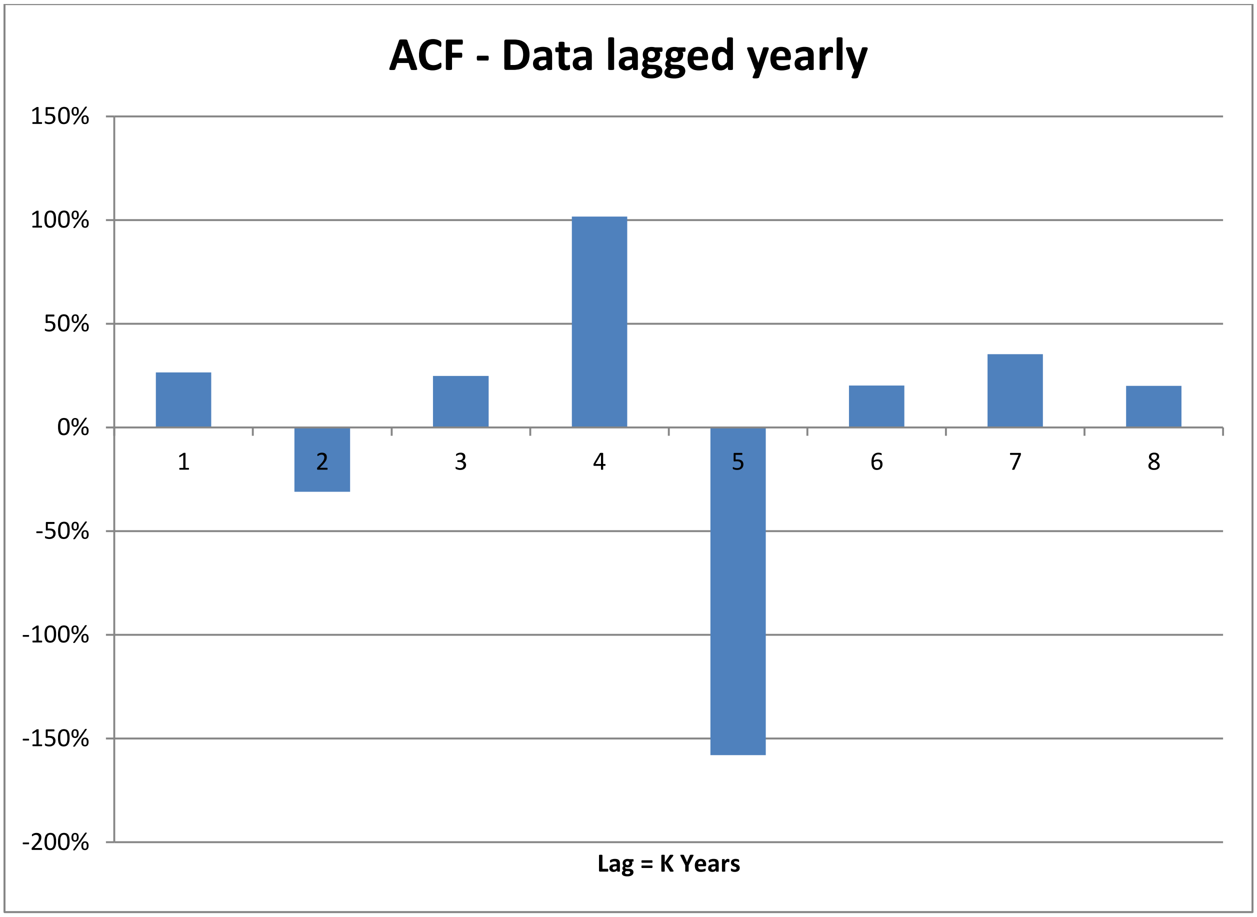
$$\gamma_0 = E\left[(x_t - \mu)^2\right] = E\left[x_t^2\right] - \mu^2$$

Results in the following graph:



The autocorrelations for lag k are mapped very well by the exponential function $y = 0.9774e^{-0.014x}$. This exponential decay suggests that the S&P 500 Index would be best modeled by an Auto Regressive or AR process.

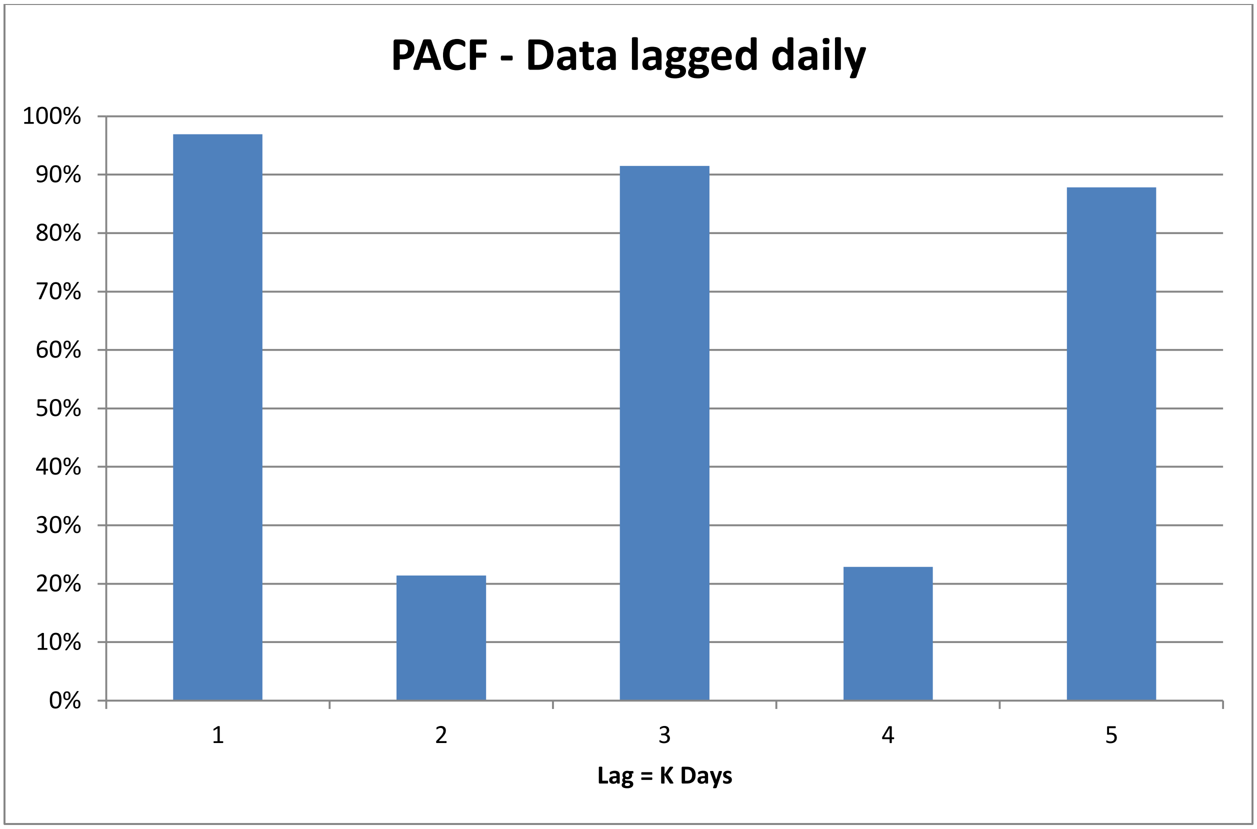
However, if we look at the ACF for lags of one or more years, the ACF is much less consistent:

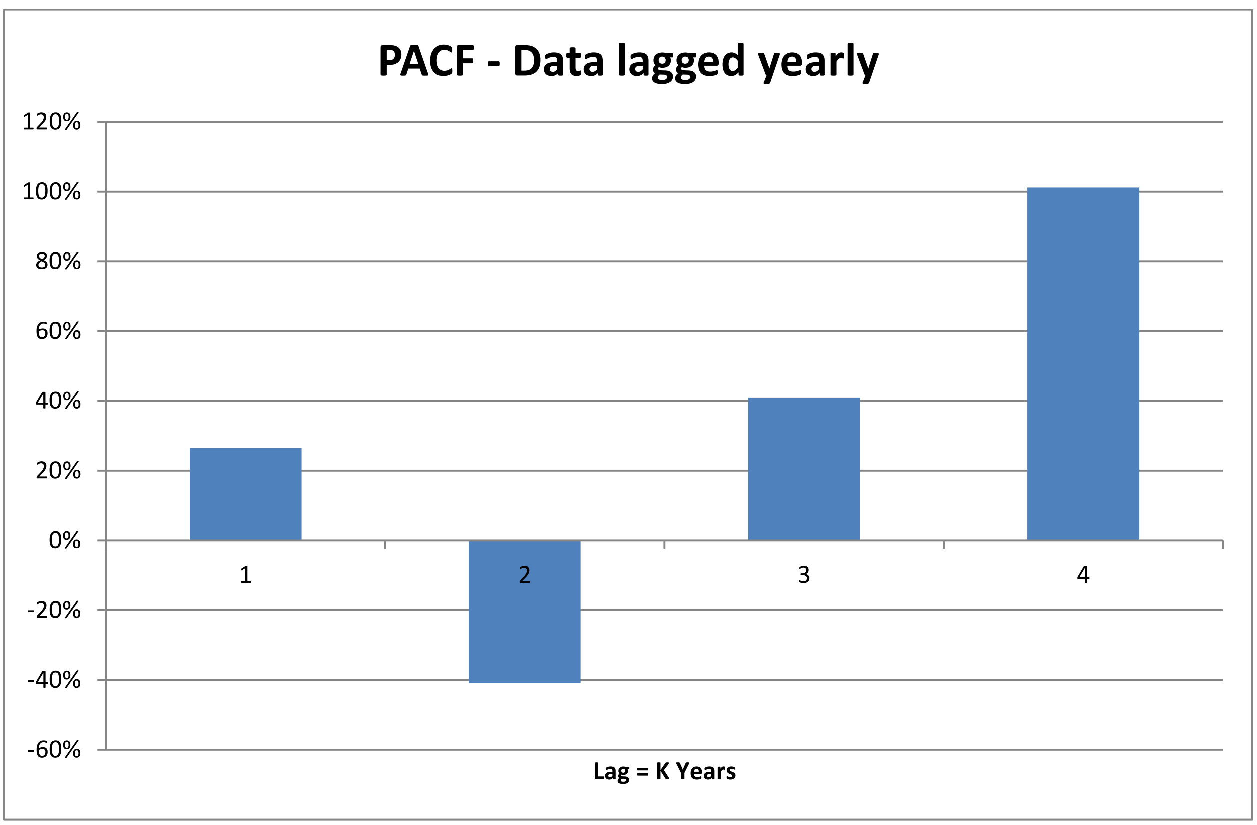


This graph seems to suggest that the S&P 500 Index would be best modeled by a Moving Average or MA process.

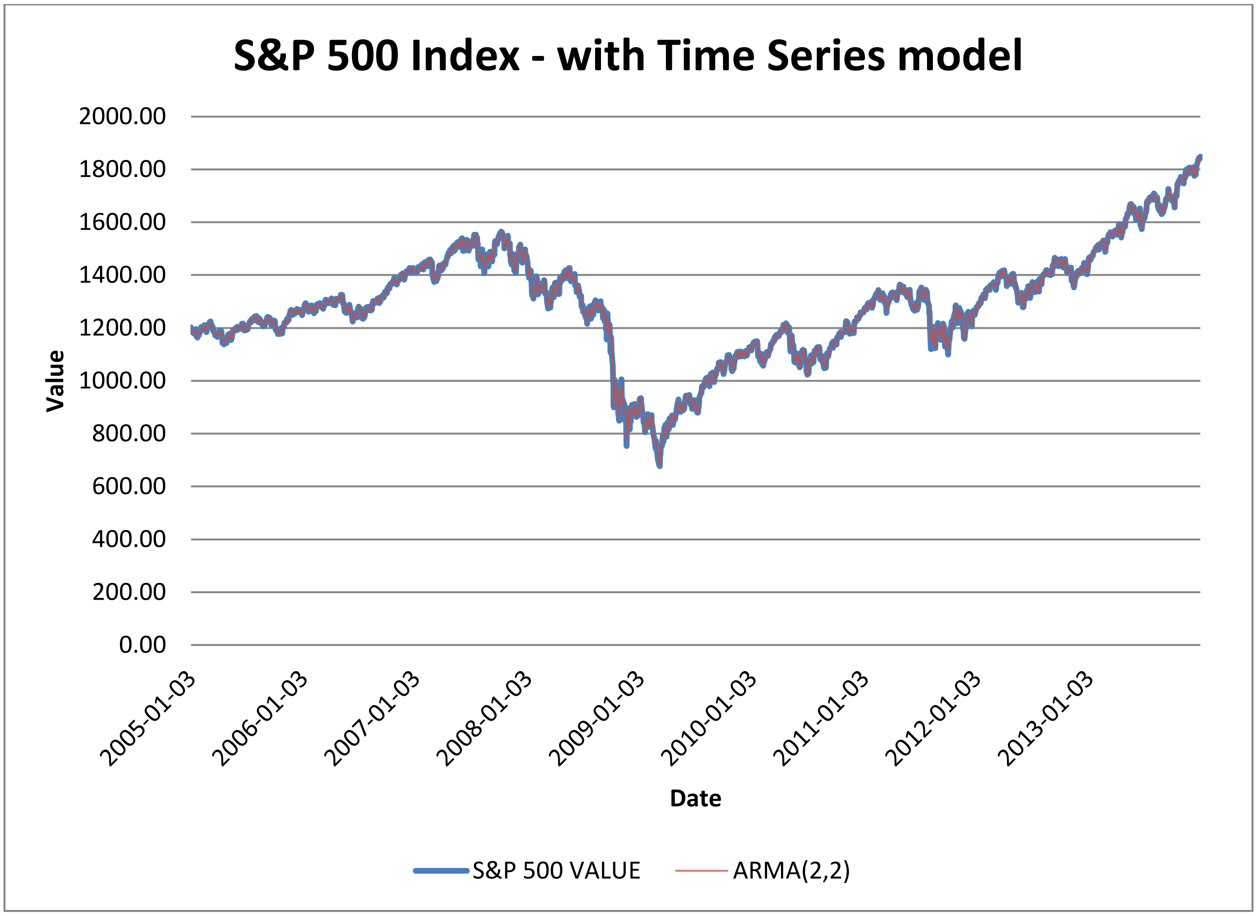
To confirm whether the S&P 500 would be better modeled with a AR or MA process, we'll look at the PACFs for the various lag periods to determine which model would have the better fit.

The PACFs results in the following:


PACF - Data lagged daily
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
2
3
4
5
Lag = K Days


PACF - Data lagged yearly
120%
100%
80%
60%
40%
20%
0%
-20%
-40%
-60%
1
2
3
4
Lag = K Years

The PACFs are inconclusive regarding whether an AR or MA model would be best used to describe the movement of S&P 500 Index. Using a mixed model will benefit from the advantages of both an autoregressive and moving average process. After several iterations, the two period ARMA model had the best fit of the existing data.



Conclusion

The two period autoregressive moving average model does a great job modeling the movements of the S&P 500's value. However, one weakness that this model possesses is in dealing with very large market movements. The autoregressive moving average model does not account for these large changes until after they have occurred. This is apparent in the graph during the financial crisis in the fall of 2008 where the model lags behind the actual data. Overall, this model is a fair estimator of the S&P 500 Index's value.