Stepwise Regression to Predict Market Performance in the Janet Yellen Era

by

Zachary Williams

Mercy College

Abstract

This paper adds to research on the effect of past interest rates on current market performance; specifically, this papers attempts to determine if a model can be developed to predict market performance based on historical interest rates. This research utilizes data from two different time periods to construct two different models, and each model is tested to determine its predictive ability during the time period when Janet Yellen was Chair of the Board of Governors of the Federal Reserve System. This research finds that models with strong predictive ability can be developed within time periods, but due to issues of non-stationarity and overfitting, those models become much weaker when applied to a different time period. Future research is necessary to create more successful models, and that future research needs to involve rigorous analysis of specific time periods to avoid problems with non-stationarity, and that research needs to utilize a number of data analysis techniques to avoid problems with overfitting.

Keywords: Interest Rate Effects, Market Performance, Lagged Interest Rates, Non-Stationarity, Overfitting

Journal of Management and Innovation, 4(2), Fall 2018

Introduction

This research expands on the research of Williams, Credle, & McLain (in press) and the work of Williams (2018) to test the strength of the predictive ability of past interest rates on current market performance, and the end result of this research is an assessment of the strength of two predictive models. The first model was developed based on the time period when Alan Greenspan was Chairman of the Board of Governors of the Federal Reserve System, and the second model was developed based on the time period when Ben Bernanke was Chairman of the Board of Governors of the Federal Reserve System. Both models are applied to the time period when Janet Yellen was Chair of the Board of Governors of the Federal Reserve System.

The work by Williams et al. (in press) demonstrated the predictive power of lagged interest rates. That research demonstrated that interest rates from up to 80 months in the past were able to explain a substantial amount of the variation in current market performance; in fact, past interest rates were able to explain much more of the variation than were current interest rates. The work by Williams (2018) expanded this topic of analysis by analyzing interest rates from up to 495 months in the past; that study found that these interest rates from over 40 years in the past were significantly correlated to current market performance and in some cases had very strong explanatory value. Furthermore, the work by Williams (2018) found that interest rates from the recent past are negatively correlated with current market performance but interest rates from the more distant past have a positive correlation with current market performance. The finding of the negative correlation was consistent with traditional economic theory that indicates that higher interest rates slow economic growth, and the finding of the positive correlation was consistent with some of the non-empirical literature that indicates that higher interest rates are often more prudent and ultimately better for the economy.

With the exception of the time periods when the correlation was switching from negative to positive, the research by Williams (2018) found that all of the correlations between past interest rates and current market performance were statistically significant. And in many cases, the r-square values from interest rates in the distant past were much higher than from current interest rates. This finding indicates that a predictive model for current market performance using past interest rates might be very successful. The fact that some interest rate lags have a negative correlation and that some interest rate lags have a positive correlation suggests that those independent variables have different explanatory power that would not be compromised by multicollinearity. Additionally, the expansive time period between independent variables would also suggest that those variables have different explanatory power. For example, an interest rate from 80 months in the past that has strong negative correlation might be combined with an interest rate form 350 months in the past that has a strong positive correlation; the combination of these two independent variables could result in a model with low multicollinearity because of the significance difference in time period and the difference in positive versus negative correlation.

However, the work by Williams (2018) found that the correlations, r-square values, and p-values were very different by time period. This finding indicates that a predictive model for current market performance using past interest rates may not be useful when applied to different regimes. The research in this paper addresses that issue, and this paper attempts to determine whether or not a model developed in one time period can be successfully used to predict market performance in a different time period.

The process of this research involves developing two models that will predict current market performance. The fist model is based on market performance and past interest rates during the time period when Alan Greenspan was Chairman of the Fed (the Greenspan model), and the second model is based on the market performance and past interest rates during the time period when Ben Bernanke was Chairman of the Fed (the Bernanke model). These models are developed using stepwise regression with current market performance as the dependent variable and current and past interest rates as the independent variables. The stepwise regression was stopped after 20 iterations when r-square values reached approximately 98%. The results of both regressions were then individually applied to the time period when Janet Yellen was Chair of the Fed, and the predictive success of each model was evaluated. Each model was tested by determining the predicted value of the Dow Jones Industrial Average (DJIA) and comparing that to the actual value of the DJIA for each month during the Janet Yellen time period. Lastly, the independent variables from each model were entered as independent variables into a regression analysis for the Janel Yellen time period; this last step was taken to determine if the correlation coefficients and the significance values of each independent variable would remain stable when applied to a different regime.

Literature Review

As mentioned, the work of Williams (2018) found significant differences in each time period with the relationship between past interest rates and current market performance. Because of this concern, the literature review in this paper will pay special attention to non-stationarity problems and overfitting issues.

Non-Stationarity

Non-stationarity occurs when the relationship among variables changes from one time period to another (Hendry & Pretis, 2016). This creates a problem when researchers attempt to construct a model based on one time period and use it to forecast outcomes in a different time period. Hendry & Pretis (2016) specify two broad sources of non-stationarity. The first source is the accumulated effect of past shocks; in other words, things change over time, and future changes accentuate the effect of past changes. The second source of non-stationarity is the sudden occurrence of new shocks; this is the introduction of a new variable that was not quantified in the previous regime.

If indeed non-stationarity is an issue in the analysis of past interest rates and current market performance, then it will be much more difficult to precisely quantify observed trends and to precisely quantify the future effects of past interest rates. However, this does not mean that the trends observed are invalid. In fact, the presence of non-stationarity often makes it easier to identify long-term trends

(Hendry & Pretis, 2016). Hendry & Pretis note that when all of the variables remain constant, it is difficult to identify causal relationships; however, when regime change occurs, the endurance of long-term trends can be more noticeable. Thus, if non-stationarity issues are present in the research of this paper, that may be helpful in determining the long-term trends.

There are numerous researchers who have identified non-stationarity problems in data sets, and only a few will be mentioned here for context. The non-stationary problem has been a common topic in recent machine learning advances (Sugiyama & Kawanabe, 2012). The ability of computer algorithms to find patterns in data is incredibly strong; however, Sugiyama & Kawanabe (2012) explain that the value of the output of these algorithms is often compromised by a computer's inability to detect non-stationarity. These authors suggest solutions to this problem that could lead to autonomous machine learning. In a production context, authors Li, Chan, Chung, & Niu (2015) studied inventory control policies in production processes, and those researchers found that assuming stationarity across periods would deteriorate inventory control performance and that it was better to assume non-stationarity. In a pricing context, author Kelly Semrad (2016) found that adjusting hotel room pricing based on the assumption of stationarity was not likely to yield long-term profits.

This literature demonstrates that the problem of non-stationarity can be observed in many contexts, and it could certainly be relevant in the analysis periods in the paper.

Overfitting

An issue that is different but related to non-stationarity is overfitting. Overfitting occurs when an analyst constructs a model that has strong precision based on random noise in a unique pool of data, but the precision of the model drops significantly when it is applied to a new data set that does not have the same random noise (Hallinan, 2014). In his 2012 book, The Signal and the Noise, Nate Silver explained that quantitative models attempt to find an underlying signal that will properly predict an outcome; if the signal is a true indicator of the outcome, then it will be consistent over time. During certain time periods, there will be noise that appears to affect the outcome. However, that noise could be completely unrelated to the outcome, and the supposed relationship could be completely coincidental. The term signal essentially means a change in the independent variable, and the term noise means error in the model.

Nate Silver (2013) offers three practical ways of guarding against overfitting, and Jennifer Hallinan (2014) suggests two methods of data analysis that might prevent overfitting. Silver (2013) suggests that quantitative analysts should be less confident about the accuracy of their forecasts, avoid relying exclusively on non-human analyses, and use Bayes’ theorem to adjust their own predictions for error. Regarding his point about confidence levels, Silver concludes that forecasters are accurate less often than their conveyed confidence level. For example, if an economist predicts that her forecast will be accurate 90% of the time, Silver suggests that that analyst will be accurate much less than 90%. To make matters worse, the economist who presented that forecast is unlikely to have publicly stated the confidence interval and the confidence level; instead, the economist is more likely to have presented a single point estimate. Silver also postulates the importance of having a human analyst make a reasonable assessment of the predictive factors. Computer algorithms are able to find all sorts of correlations in large pools of data, but those correlations could be due to random noise. Silver suggests that a way to guard against this is to take the simple, practical step of subjecting the stated correlation to a test of human logic. Relationships that don’t make sense intuitively to a person familiar with the topic should be excluded from predictive models. Lastly, Silver suggests that Bayes’ theorem could be used to adjust the accuracy of model predictions. Specifically, Silver suggests using Bayes’ theorem to adjust for overall likelihood of an event occurring. For example, if a model predicts that a recession is 90% likely if the

Federal Reserve takes a certain action, Silver suggests that the 90% prediction should be adjusted based on the overall chance of a recession. The 90% figure should be adjusted down by taking the ratio of the correctly predicted recessions to predicted recessions. If there are a very large number of total predicted recessions, then this 90% could be significantly lower in reality.

The data analysis methods suggested by Jennifer Hallinan (2014) to detect and adjust for non-stationarity are three-set validation and cross validation. Three-set validation involves developing a model on one-third of the data, refining the model based on a different one-third of the data, and testing the model on the last one-third of the data. Refining the model on the second set of data is a way to factor out noise, and utilizing a third set of completely untouched data allows for a true test of the model's accuracy. This is an excellent method if the researchers have enough data, but the disadvantage of the three-set validation method is that it requires an extremely large amount of total data; additionally, dividing the data could cause problems with non-stationarity.

Cross validation is another option; this process is sometimes referred to as leave-one-out validation. In cross validation, the data are divided into equal parts. It is common to divide the data into groups of 10 (Hallinan, 2014). A model would then be developed based on nine of the parts and tested on the tenth part. This process would be repeated ten times and refined based on the level of noise found at each iteration. Like three-set validation, this method allows the model to be tested on a section of the data that did not influence the development of the model, and this method does not require as much data as three-set validation since the majority of the data can be used for each iteration.

It is important to distinguish the difference between non-stationarity and overfitting, so that each issue might be addressed. Non-stationarity recognizes that there might have been a change in the underlying relationship. To use the terms of Hallinan and Silver, non-stationarity means that the actual signal has changed. By contrast, overfitting addresses the question of whether the model is actually reading a true signal or just random noise. Overfitting guards against the danger of relying on coincidental, random relationships, and non-stationarity guards against relying on relationships that have since changed.

Methodology

This study measures the ability to predict current market performance as a dependent variable with current and past interest rates as independent variables. Ultimately, this study constructs predictive models based on one time period and tests the validity of those models in different time periods. The model inputs are based on broad market and interest rate measures, and the model construction and evaluation are based on traditional regression analysis.

Data

The dependent variable of market performance in this study is represented by the monthly performance of the Dow Jones Industrial Average (DJIA). The DJIA data are provided by reports from Macrotrends.com. The DJIA contains 30 public, large, heavily-traded U.S. companies, and most market analysts consider the DJIA to be an accurate indication of the strength of the U.S. stock market (Hom, 2012). The fact that the DJIA is price-weighted does cause some concern because a price-weighted index is biased in favor of stocks with high prices, and there is often a completely arbitrary difference among stocks in the DJIA; however, the composition of the index has been adjusted constantly over its life in order to maintain its place as the most common reflection of stock market performance (Hom, 2012).

The independent variable of interest rates in this study is represented by monthly listings of the Federal Funds rate and lagged iterations of the Federal Funds rate. These data were collected from the Federal Reserve Economic Data (FRED). The Federal Funds rate is generally considered a good indication of U.S. interest rates for two reasons; one, this interest rate is specifically targeted by the Federal Reserve in order to change interest rates in the overall U.S. economy (Bernanke and Blinder 1992), and two, this interest rate is determined supply and demand of capital in the banking industry (Ochoa, 1999).

Time periods

There are three time periods utilized in this analysis – the Greenspan time period, the Bernanke period, and the Yellen period. The Greenspan time period consists of market performance data and interest rate data during the time when Alan Greenspan was Chairman of the Fed from September 1987 to January 2006. The Bernanke time period consists of market performance data and interest rate data during the time when Ben Bernanke was Chairman of the Fed from February 2006 to February 2014. The Yellen time period consists of market performance data and interest rate data during the time when Janet Yellen was Chair of the Fed from March 2014 to February 2018.

Lagged Independent Variables

The first step in this analysis is to determine the correlations, r-square values, and p-values of all of the independent variables when compared to the dependent variable. This is the type of analysis that was conducted by Williams, Credle, & McLain (in press) and extended by Williams in 2018. The research in this paper uses interest rates from up to 495 months in the past as independent variables. The research by Williams (2018) indicated that almost all of these independent variables, including the ones from over 40 years in the past were statistically correlated to current market performance; therefore, these interest rates could be useful if developing predictive models. This study analyzes all of the independent variables within each time period. For example, for the Greenspan time period, current market performance at each month from 1987 to 2006 is compared to the Federal Funds rate from that month and the Federal Funds rate from each of the previous 495 months. The process is then repeated for the Bernanke time and then for the Yellen time period. For each time period, r-square values, correlation coefficients, and p-values are calculated and reported in the Figures in this paper.

Stepwise Regression

The next step in this analysis involved conducting stepwise regression conducted with SPSS to determine the best predictive model from the Greenspan period and the best predictive model from the Bernanke period. The stepwise method added and deleted independent variables one at a time. The variables with the strongest individual p-values were added as long as the p-value was lower than 5%. Once added, variables that were found to have weak p-values after other independent variables were added were removed if the p-value was greater than 10%. Stepwise regression has the advantage of using the strongest predictive variables and also removing variables when multicollinearity becomes a factor. The stepwise method adds the best independent variables one at a time and then removes those variables if the variable’s p-value deteriorates and moves above 10%. Once the stepwise model is completed, the variables added later in the process can be removed in the interest of creating a more simplified, parsimonious model. In the analysis in this paper, only the first 20 iterations of the stepwise regressions were used in the final model because after 20 iterations, the r-square values had reached 98%; thus, by definition any additional variables would only be able to add a total of 2% to the r-square value of the model.

R-Square Metric

The r-square metric indicates the amount of variation in the in the dependent variable that can be explained by any one independent variable or by any combination of independent variables (Freed, Jones, & Bergquist 2013). The increasing magnitude of the r-square values as interest rates from further in the past were analyzed was a key finding of the Williams, Credle, McLain study (in press). This research expands that analysis by attempting to determine if a combination of independent variables will have strong explanatory and predictive power.

Correlation Coefficients

Like r-square values, correlation coefficients also indicate the strength of the relationship between a dependent variable and one or more independent variables (Freed et al. 2013). One advantage of a correlation coefficient is that it indicates whether the relationship is positive or negative (Freed et al. 2013). Once the stepwise regression analysis was complete, this study specifically looked at the correlation coefficients to determine if the positive or negative correlations were consistent with the overall correlations in that time period. For example, if the first 100 interest rate lags during the Greenspan period were negatively correlated, then this would indicate that any independent variables from the first 100 lags in the Greenspan model would have a negative correlation coefficient.

Predictive Model Tests

Once the model from the Greenspan period and the model from the Bernanke period were constructed, they were then applied individually to the Yellen period. First the Greenspan model was used to predict stock market values for each of the 48 months during the Yellen period; those results were then compared to the actually value of the DJIA. The predicted values and actual values were correlated with each other and correlation coefficients, r-square values, adjusted r-square values, standard error of estimates, and model significance levels were measured. The process was then repeated with the Bernanke model.

Additionally, the beta of the predicted value of the DJIA was measured with respect to the actual value of the DJIA. This calculation was the correlation coefficient of the two variables multiplied by the ratio of the standard deviation of the predicted value to the standard deviation of the actual value.

Multiple Regression with Independent Variable Inputs

The last step in this analysis involved using the independent variables found in the Greenspan model and the Bernanke model and then separately using those variables to create a new regression model based on the Yellen period. For example, if the independent variables in the Greenspan model are found to be the 45th lag, the 210th lag, and the 351st lag, then those three variables would be imputed as independent variables for the data during the Yellen period, and SPSS would recalculate the appropriate correlation coefficients and the supporting metrics for that regression. This step creates a new regression model using the same independent variables. The point of this step is not to assess the strength of the new regression model; instead, the point is to observe the stability of the regression coefficients. The two key questions are if the same independent variables remain significant and how much the regression coefficients change.

This is a less traditional method of evaluating the strength of a regression model, but in this case, it is likely to provide insight in two ways. One, if the correlation coefficients change signs or are drastically different once they are inserted into a model that is based on the Yellen period, then this would indicate a particularly unstable relationship between that independent variable and current market performance. For example, suppose the Bernanke model were to specify that a 1% increase in the 200th lag of the Federal Funds rate should result in 350-point decrease in the DJIA; and when the 200th lag is inserted into a model based on Yellen period, the new coefficient indicates that a 1% increase should result in 200-point increase in the DJIA. This example would indicate a red flag for using the 200th lag as an independent variable. The other way this analysis could be useful is if an independent variable is highly significant in the initial model, but the significance level indicates questionable significance once the variable is applied to Yellen time period; this would also cast doubt on using the 200th lag as an independent variable.

Hypothesis Tests

At each necessary step in this analysis, hypothesis tests will be conducted to determine statistical significance. The null hypothesis will always be that there is no correlation between the independent variable(s) and the dependent variable. The alternative hypothesis will be that there is a statistically significant relationship between the independent variable(s) and the dependent variable. The significance level will be 5%.

The first series of hypothesis tests involve the significance of the correlation coefficients in each of the three time periods. Each of the 496 versions of the Federal Funds rate will be tested against current market performance, for a total of 1,488 hypothesis tests. A hypothesis tests will be conducted to determine if the Greenspan model and the Bernanke model as a whole is statistically significant; additionally, a hypothesis test will be conducted to determine the significance for each independent variable that is ultimately included in each model. This part of the analysis will include a total of 34 hypothesis tests. Next, a hypothesis test will be conducted to determine if there is a statistically significant relationship between the predicted values of the DJIA and the actual values of the DJIA for both the Greenspan model and the Bernanke model; thus, there will be 2 hypothesis tests. Finally, when the Greenspan and the Bernanke model are applied to the Yellen period, hypothesis tests are conducted to determine the overall significance of each model and the individual significance of each independent variable, for a total of 34 hypothesis tests. Thus, in all, 1,558 hypothesis tests are conducted in the research.

Results

Correlations, R-Square Values, and P-Values of Independent Variables by Time Period

Greenspan period.

Figure 1, Figure 2, and Figure 3 illustrate the r-square values, correlations, and p-values for each correlation in the Greenspan period. Figure 1 indicates that r-square values increase as time lags increase and then diminish to zero at approximately 175 lags. At that point r-square values increase and reach a new peak in explanatory value of approximately 65% after 370 lags. As was the finding with Williams et al. (in press), interest rates from the past have significant explanatory value. After 175 lags, Figure 2 indicates that the correlation between the independent variables and the dependent variable becomes positive. Figure 3 indicates that around this 175-lag point is the area where hypothesis tests fail at the 5% significant level. This finding is consistent with Williams (2018); that research found that when the correlations are shifting between negative and positive, they fail hypothesis tests and are not statistically significant.

However, Figure 3 indicates that all of the other independent variables prior to the 175-month mark are significant and negatively correlated, and all of the independent variables after this point are significant and positively correlated.

Bernanke period.

Figure 4, Figure 5, and Figure 6 illustrate the r-square values, correlations, and p-values for each correlation in the Bernanke period. Figure 4 indicates fluctuating r-square values, and Figure 5 indicates that this fluctuation is where the correlation coefficients are changing between negative and positive. Additionally, Figure 6 indicates that during this change, the correlations fail hypothesis tests and are not statistically significant. These findings are consistent with Williams, Credle, & McLain (in press) and Williams (2018). However, the number and frequency of the fluctuations was not been observed in those previous two studies.

Yellen period.

Figure 7, Figure 8, and Figure 9 illustrate the r-square values, correlations, and p-values for each correlation in the Yellen period. The results of the Yellen period contain aspects that are consistent with past research and aspects that consistent with new observations in the Bernanke period; additionally, the Yellen period offers two new insights. As is consistent with past research, the r-square values fluctuate and do so as the correlations switch between negative and positive, and at these transition points, the correlations are not statistically significant. As with the new observations seen in the Bernanke period, there are many more fluctuations. The first new insights in the Yellen period is that the current and most recent interest rates have very high r-square values. The other new insight that the fluctuations from further in the past have higher r-square values than observed in other periods, and these high r-square values are observed whether the correlation is positive or negative.

Greenspan Model

The predictive model developed based on the Greenspan time period using stepwise regression seems to be extraordinarily robust. The details of the model are in Table 1 and Table 2. The model is based on 18 independent variables and has an r-square of 98% and an adjusted r-square of 97.7%. The one standard deviation of error is 261.1 points on the DJIA, and the model is strongly significant.

However, there is one particularly concerning issue. Some of the correlation coefficients do not conform to positive or negative correlations observed in Figure 2. Two examples of this are Lag 326 and Lag 491, which can be seen in Table 1. Figure 2 indicates that interest rates from this far in the past are positively correlated with current market performance. However, both of these lags have highly negative correlation coefficients.

Predictive results.

Table 5 shows that when the Greenspan model is used to predict the DJIA during the Yellen period, the results indicate a rather weak model. The r-square drops drastically to 24.6%, and the adjusted r-square is 23%. The standard error of the estimate is 2249.4 points on the DJIA, which is almost 2000 points higher than when the Greenspan model is used in the Greenspan period as observed in Table 2. The model is statistically significant at 0.03%, but this is relatively much higher than the p-value in Table 2 of 0.00000%.

Applied and adjusted to Yellen period.

Table 7 and Table 8 illustrate the result when the independent variables from the Greenspan model are used to calculate a regression model based on the data from the Yellen period. In this scenario, a new regression model is calculated by SPSS, and the point of this scenario is to measure the sensitivity of the independent variables across time periods. Table 8 indicates that the newly constructed model is very strong, but that is not relevant to this analysis; what is relevant is the amount of change from Table 1 to Table 7. If there is a large degree of change, this would indicate that the independent variables are highly sensitive to time period bias.

There are two key results in these tables. One, only 4 of the 18 independent variables are found to be statistically significant in the new regression in Table 7. Two, the correlation coefficients for these 4 independent variables are vastly different then they are in Table 1. These results indicate independent variables that are highly sensitive to time period biases, which will be discussed in the conclusion section.

Bernanke Model

The predictive model based on data from the Bernanke era using stepwise regression is highly predictive and strongly significant. These strong results are similar to model developed based on the Greenspan time period. The details of the Bernanke model are in Table 3 and in Table 4. The stepwise regression model settled on 12 independent variables after adding and removing for 20 iterations; the model has an r-square of 98.5% and an adjusted r-square of only slightly lower at 98.3%. The standard error of the estimate is 258.1 points on the DJIA, and the model is highly significant with a p-value of 0.00000%.

Predictive results.

Table 6 indicates that when the Bernanke model is used to predict values for the DJIA during the Yellen time period, the model loses predictive value. However, the drop in value of the Bernanke model is not nearly as significant as the drop in value of the Greenspan model. The r-square in Table 6 is 60.5%, and the adjusted r-square is only slightly lower at 59.6%. the standard error of the estimate is 1628.7, which is an increase 1370 compared to Bernanke model applied to the Bernanke period in Table 4. The model is statistically significant, and there is no noticeable decrease in significance compared to Table 4.

Applied and adjusted to Yellen period.

Table 9 and Table 10 show the results when the independent variables from the Bernanke model are inputted to calculate a new regression model based on the historical data of the Yellen time period. The results illustrated in these tables are very similar to results when the Greenspan model is applied and adjusted to the Yellen period. Table 10 indicates a robust regression model, but that is not the relevant aspect of this step in the analysis. The relevant aspect is that the change in independent variables observed from Table 3 to Table 9 is significant. This large change in independent variables represents highly sensitive variables that may not be useful predictors across time periods. This issue will be discussed more thoroughly in the conclusion section.

Conclusion

The topic of the lagged interest rate effect is potentially highly applicable to many fields of study, including Monetary Policy, Economic Theory, and Finance. The results of this study illustrate that the same trends observed in earlier work on this topic are present in the time periods of this study; those trends are the increasing r-square values as interest rates from further and further in the past are compared to current market performance, correlation coefficients that change from negative to positive, and distant interest rate lags that are statistically significant. A better understanding of these trends could contribute significantly to the fields mentioned above. For example, as the Federal Reserve Board of Governors seeks to enact the best monetary policy, knowledge of the lagged and lasting effect of their actions could be critical to their decision process; additionally, those decision makers would certainly benefit from an understand of when there is a positive correlation and when there is a negative correlation between interest rates and economic performance. Similar examples could be drawn from Economics and Finance.

But the issue of quantifying these trends precisely seems to be very difficult because of a time period bias. The research of Williams (2018) reached this conclusion by comparing the characteristics of different time periods. The research in this study has gone further by attempting to build a predictive model that can be empirically tested in different time periods. The results of the empirical tests in this paper indicate a lack of success in precisely quantifying any of these trends. It appears that this lack of success is due to issues with non-stationarity and overfitting. Those two issues will be discussed below along with possible ways to overcome those issues in future research.

Non-Stationarity

Figure 2, Figure 5, and Figure 8 indicate vast differences in the effect of past interest rates on current market performance. Figure 2 from the Greenspan period indicates that interest rates initially have a negative effect that changes to a positive effect as rates from further in the past are compared to current market performance. Similarly, Figure 5 from the Bernanke era shows an initial negative correlation that changes to a positive correlation (the very recent positive correlations are not statistically significant). However, as interest rates from further in the past are analyzed, there are multiple and relatively rapid fluctuations between positive and negative correlations in Figure 5. These rapid fluctuations in Figure cannot be explained by the analysis of Figure 2.

The work of Williams (2018) explained Figure 2 with literature that demonstrated an empirical negative correlation between interest rates and market performance as well as non-empirical literature that supported a positive long-term correlation between interest rates and market performance. But there is no literature that would support a rapidly fluctuating correlation as seen in Figure 5. This fluctuation is clearly due to unique characteristics of that time period. The results in Figure 8 illustrate this same conclusion; the rapid and larger fluctuations seen in the Yellen time period must be due to unique factors of the 2014 to 2018 time period.

Since the issue of non-stationarity seems very hard to avoid, future research on this topic should pay close attention to the individual characteristics of each time period. Before attempting to quantify these interest rate trends and before attempting to develop predictive models, researchers need to understand the empirical qualities of the time period, such as the metrics shown in Figures 1-9. Additionally, researchers must seek to understand overall market sentiments and the general perception of the interest rate decision makers. When researchers find differences between time periods, those differences must be accounted for.

Overfitting

The results of this analysis also indicate that overfitting was a significant issue. The first red flag is that the correlation coefficients for the Greenspan model (Table 1) and Bernanke model (Table 3) that do not match the expected direction of the correlation in Figure 2 and in Figure 5. Since the direction of the correlations does not match, these regression models do not make intuitive sense. For example, Table 1 indicates that Lag 108 is statistically significant and strongly positive with a correlation coefficient of 332.0. Thus, for a 1% increase in the Federal Funds rate from 108 months in the past, there is an expected 332.0-point increase in the DJIA. This does not make sense in light of the information in Figure 2. Figure 2 indicates that interest rates from 108 months in the past are statistically significant and strongly negative. Thus, this seems to indicate that the regression model has overfit to some combination of characterizes that are not likely to be relevant in different time periods.

Ultimately, where the predictive model breaks down and where the most conclusive evidence of overfitting is when each model is applied to the Yellen period and the strength of each model drastically decreases. The Greenspan model appears to be strong in Table 2, but the Greenspan model appears to be very weak in Table 5. The r-square drops to 24.6% and the standard error of the estimate increases 1,988 DJIA points. Likewise, the Bernanke model appears to be strong in Table 4, but the Bernanke model appears to be weak in Table 6. The r-square drops to 60.5% and the standard error of the estimate increases 1,370 DJIA points. The Bernanke model does not deteriorate as much as the Greenspan model; that might be due to the fact that the Bernanke period is closer in time to the Yellen period, but that should be addressed in future research.

In addition to the deterioration of the overall model metrics, the results in Table 7 and in Table 9 indicate that most of the independent variables in Greenspan model and in the Bernanke model become insignificant when those variables are imputed into the regression analysis of the Yellen period. The point of using Greenspan and Bernanke variables in the Yellen period was to test the sensitivity of the independent variables. If the independent variables remained statistically significant and if the correlation coefficients remained similar, then that would indicate an enduring, long-term correlation of those independent variables with current market performance. In this step of the analysis, a new regression model is calculated in SPSS. For example, in the Bernanke period, Table 10 indicates that the new regression model is very robust, but that is not an important part of this analysis; the important part of this analysis is the amount of change from Table 3 to Table 9. If there is a large degree of change, that would indicate that the independent variables are highly sensitive to time period bias. In comparing Table 3 to Table 9, only three of the independent variables remain statistically significant in Table 9 – Lag 141, Lag 85, and Lag 322. Also, there is a significant change in all three of these correlation coefficients when these independent variables are applied to different time periods. The most drastic example is Lag 85. When Lag 85 is imputed into the Bernanke model based on the Bernanke data, the correlation coefficient is 1582. This means that a 1% increase in the Federal Funds rate would predict a 1582-point increase in the DJIA. When Lag 85 is calculated based on the data from the Yellen period, the correlation coefficient is -972. This means that a Federal Funds rate increase of 1% would predict a 972-point decrease in the DJIA. Thus, from Table 3 to Table 9 there is a 2,554-point change in the correlation coefficient of this variable. These results indicate that the independent variables are highly sensitive to time period bias, and these results cast doubt on a forecaster's ability to use a model developed in one time period to predict the outcomes in a different time period.

Avoiding Overfitting

Based on the best practices described by Silver and Hallinan, there are a number of measures that could be taken to avoid the overfitting problem in this research. Nate Silver's (2013) first recommendation of avoiding overconfidence is relevant to this paper. Analysts who construct these types of models might feel the temptation to be overconfident because the large pool of data can produce phenomenal r-square values. For example, the r-square values in Table 2 and Table

4 are 98% and 98.5%, respectively. And those values could have been higher if this analysis went beyond 20 iterations. But this overconfidence would be misplaced because of overfitting. However, awareness of this issue could allow analysts to use these models cautiously and for useful purposes. For example, Table 6 indicates that the Bernanke model is reasonably strong with an r-square of 60.5%. As long as the users of this model are not expecting an r-square of 98%, then 60.5% could have some value.

Nate Silver's (2013) second recommendation of ensuring that every predictive model be evaluated by a human being could also be helpful to the analysis in this paper. The idea that past interest rates have a strong effect on current market performance certainly does pass a test of human intuition. There is abundant literature that discusses the relationship between these two variables (Williams et al. in press). Additionally, the idea that interest rates from the relatively recent past have a negative correlation and that interest rates from the more distant past have positive correlation is also supported by literature on the topic (Williams, 2018) and would pass a test of human intuition. The two main issues in this research that would cause a human analyst to have doubt about the models in this research are that the correlation coefficients of the models do not always match the positivity or negativity calculated for that time period and that the positivity or negativity calculated for that the period fluctuates rapidly in the Bernanke and Yellen period. The fact that correlation coefficients do not match expectations is counter to human intuition, and those independent variables should be adjusted or removed. Additionally, the rapid fluctuations of the Bernanke and Yellen periods cannot be supported by the literature and cannot be rationalized by human intuition; therefore, any conclusions based on these fluctuations should be stated cautiously.

Silver’s (2013) third recommendation that probability estimates should be adjusted down with Bayes’ theorem could also be useful in strengthening the analysis in this paper. The application of Bayes’ Theorem that Silver suggests would involve the ratio of correct predictions to all predictions of that nature. For example, if a model suggests an increase in the DJIA, the probability that the model suggests for that increase should be adjusted down if there are a large number of total predictions that the DJIA will increase. This could easily happen with a complex model that makes many predictions, so future research that incorporates Bayes’ theorem could be useful.

The methods suggested by Hallinan (2014) for correcting overfitting problems could be useful for this research. The research in this paper did utilize some steps of a three-set validation model, but this research did not incorporate changes based on the second data set. This step could have been taken; in fact, there might well be enough data to conduct much more than a three-set validation. For example, the Greenspan model could have been tested and adjusted based on the Bernanke time period before being tested on the Yellen period. Moreover, the model could have been tested and adjusted based on many other time periods before being applied to the Yellen period. Future research could use weekly or daily on interest rates and market performance in order to create more points of data. Future researchers might need to adjust for inflation and other non-stationarity across time periods, but there would be no shortage of data.

References

Bernanke, B. & Blinder, A. (1992). The Federal Funds Rate and the Channels of Monetary Transmission. American Economic Review, American Economic Association, vol. 82(4), pages 901-921, September.

Federal Reserve Economic Data (n.d.). Retrieved August 10th, 2018 from https://fred.stlouisfed.org.

Freed, N., Jones, S., & Bergquist, T. (2013). Understanding business statistics. Hoboken, NJ.: John Wiley & Sons.

Hallinan, J. (2014, November 12). Why Aren't My Results As Good As I Thought? You're Probably Overfitting. Retrieved August 20, 2018, from https://machinelearningmastery.com/arent-results-good-thought-youre-probably-overfitting/

Hendry, D.F. & Pretis, F. (2016). All change! The implications of non-stationarity for empirical modelling, forecasting and policy. Oxford Martin Policy Paper

Hom, E. (2012, October 30). What is the Dow Jones Industrial Average? Business News Daily. Retrieved from https://www.businessnewsdaily.com/3342-dow-jones-industrial-average.html

Kirk, J. T. (2011). Reprogramming the Kobayashi Maru test: A tale of an inside job and the genius behind it. Manuscript in preparation.

Li, N., Chan, F. T., Chung, S. H., & Niu, B. (2015). The impact of non-stationary demand and forecasting on a failure-prone manufacturing system. 2015 International Conference on Industrial Engineering and Operations Management (IEOM). doi:10.1109/ieom.2015.7093707

Macrotrends (n.d.). Retrieved August 9th, 2018 from macrotrends.net

Ochoa, M. (1999). Federal funds futures: A barometer of interest rate expectations. Trading Markets. Retrieved from http://tradingmarkets.com/recent/Federal_funds_futures_A_barometer_of_interest_rate_expectations-669412.html

Semrad, K. (2016). The Effects of Non-Stationarity Demand on Hotel Financial Performance. Journal of Hotel & Business Management,05(02). doi:10.4172/2169-0286.1000140

Silver, N. (2013). The signal and the noise: The art and science of prediction. London: Penguin.

Sugiyama, M., & Kawanabe, M. (2012). Machine learning in non-stationary environments: Introduction to covariate shift adaptation. Cambridge, MA: MIT Press.

Williams, Z., Credle, S.H., McLain, M. (in press). The effect of past interest rates on current market performance. Journal of Business and Finance Research.

Williams, Z. (2018). The negative, positive, and increasing correlation of historic interest rates with current market performance. Manuscript in preparation.

Table 1

Greenspan Model

Variables	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t-statistic	Significance
Constant	13198.5	1026.591		12.857	0.0%
Lag 424	143.9	33.069	0.166	4.35	0.0%
Lag 171	-142.6	26.31	-0.21	-5.419	0.0%
Lag 108	332.0	49.463	0.374	6.713	0.0%
Lag 239	30.0	15.153	0.065	1.983	5.0%
Lag 204	-67.9	15.442	-0.132	-4.395	0.0%
Lag 264	43.3	14.611	0.097	2.965	0.4%
Lag 168	-147.2	34.901	-0.188	-4.218	0.0%
Lag 88	-368.3	43.834	-0.408	-8.403	0.0%
Lag 326	-153.8	26.655	-0.189	-5.771	0.0%
Lag 98	-190.5	47.919	-0.214	-3.976	0.0%
Lag 54	-448.5	79.563	-0.27	-5.638	0.0%
Lag 491	-208.7	61.341	-0.098	-3.402	0.1%
Lag 366	94.9	26.766	0.125	3.545	0.1%
Lag 377	174.2	34.857	0.231	4.998	0.0%
Lag 271	80.0	18.584	0.18	4.304	0.0%
Lag 215	53.0	17.139	0.104	3.09	0.3%
Lag 412	-113.2	40.025	-0.123	-2.829	0.6%
Lag 236	41.2	16.403	0.088	2.513	1.3%

Table 2

Greenspan Model - Performance and Inferential Metrics

Correlation Coefficient	R Square	Adjusted R Square	Std. Error of the Estimate	Model Significance
0.990	98.0%	97.7%	261.1	0.00000%

Table 3

Bernanke Model

Variables	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t-statistic	Significance
Constant	10213.6	647.827		15.766	0.0%
Lag 242	-885.2	44.812	-0.927	-19.754	0.0%
Lag 141	-217.8	38.914	-0.125	-5.597	0.0%
Lag 85	1582.4	58.455	1.495	27.07	0.0%
Lag 360	131.3	15.963	0.266	8.224	0.0%
Lag 372	139.9	16.538	0.303	8.461	0.0%
Lag 347	-57.0	13.516	-0.101	-4.217	0.0%
Lag 320	-113.7	16.554	-0.202	-6.867	0.0%
Lag 356	71.7	15.846	0.14	4.525	0.0%
Lag 402	-91.5	21.55	-0.137	-4.248	0.0%
Lag 322	62.8	16.595	0.11	3.783	0.0%
Lag 43	979.3	60.981	0.902	16.06	0.0%
Lag 385	-74.2	14.722	-0.156	-5.043	0.0%

Table 4

Bernanke Model - Performance and Inferential Statistics

Correlation Coefficient	R Square	Adjusted R Square	Std. Error of the Estimate	Model Significance
0.993	98.5%	98.3%	258.1	0.00000%

Table 5

Greenspan Model Predicted Values Correlated with Actual DJIA in Yellen Period

Correlation Coefficient	R Square	Adjusted R Square	Std. Error of the Estimate	Model Significance	Beta of Predicted Dow
0.496	24.6%	23.0%	2249.4	0.03362%	0.44

Table 6

Bernanke Model Predicted Values Correlated with Actual DJIA in Yellen Period

Correlation Coefficient	R Square	Adjusted R Square	Std. Error of the Estimate	Model Significance	Beta of Predicted Dow
0.778	60.5%	59.6%	1628.7	0.00000%	0.51

Table 7

Greenspan Model Applied and Adjusted to Yellen Period

Variables	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t-statistic	Significance
Constant	21133.8	9516.225		2.221	3.4%
Lag 424	36.0	47.989	0.043	0.751	45.9%
Lag 171	1307.5	524.757	1.096	2.492	1.9%
Lag 108	-1382.8	268.247	-0.865	-5.155	0.0%
Lag 239	221.4	466.989	0.045	0.474	63.9%
Lag 204	-453.3	559.392	-0.096	-0.81	42.4%
Lag 264	19.7	658.232	0.009	0.03	97.6%
Lag 168	-1452.4	336.837	-1.212	-4.312	0.0%
Lag 88	42.8	424.512	0.036	0.101	92.0%
Lag 326	366.7	270.832	0.151	1.354	18.6%
Lag 98	-109.2	333.5	-0.092	-0.328	74.6%
Lag 54	-667.5	4252.483	-0.01	-0.157	87.6%
Lag 491	-21.1	161.872	-0.021	-0.13	89.7%
Lag 366	-6.4	226.366	-0.004	-0.028	97.8%
Lag 377	-487.1	191.578	-0.246	-2.543	1.7%
Lag 271	-205.7	452.001	-0.088	-0.455	65.2%
Lag 215	862.4	647.295	0.1	1.332	19.3%
Lag 412	-47.6	53.559	-0.064	-0.889	38.1%
Lag 236	474.0	612.796	0.066	0.774	44.5%

Table 8

Greenspan Model Applied to Yellen Period - Performance and Inferential Statistics

Correlation Coefficient	R Square	Adjusted R Square	Std. Error of the Estimate	Model Significance
0.994	98.8%	98.1%	354.5	0.00000%

Table 9

Bernanke Model Applied and Adjusted to Yellen Period

Variables	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t-statistic	Significance
Constant	21293.1	5107.453		4.169	0.0%
Lag 242	-250.0	447.749	-0.071	-0.558	58.0%
Lag 141	874.7	291.511	0.426	3.001	0.5%
Lag 85	-972.2	310.264	-0.768	-3.133	0.3%
Lag 360	272.3	179.012	0.165	1.521	13.7%
Lag 372	124.7	195.187	0.073	0.639	52.7%
Lag 347	514.7	263.519	0.188	1.953	5.9%
Lag 320	-246.9	312.205	-0.107	-0.791	43.4%
Lag 356	24.3	146.691	0.013	0.166	86.9%
Lag 402	49.1	59.695	0.065	0.823	41.6%
Lag 322	-595.1	320.055	-0.245	-1.86	7.1%
Lag 43	-3665.9	6657.722	-0.057	-0.551	58.5%
Lag 385	-194.4	134.125	-0.15	-1.449	15.6%

Table 10

Bernanke Model Applied to Yellen Period - Performance and Inferential Statistics

Correlation Coefficient	R Square	Adjusted R Square	Std. Error of the Estimate	Model Significance
0.988	97.6%	96.8%	461.1	0.00000%

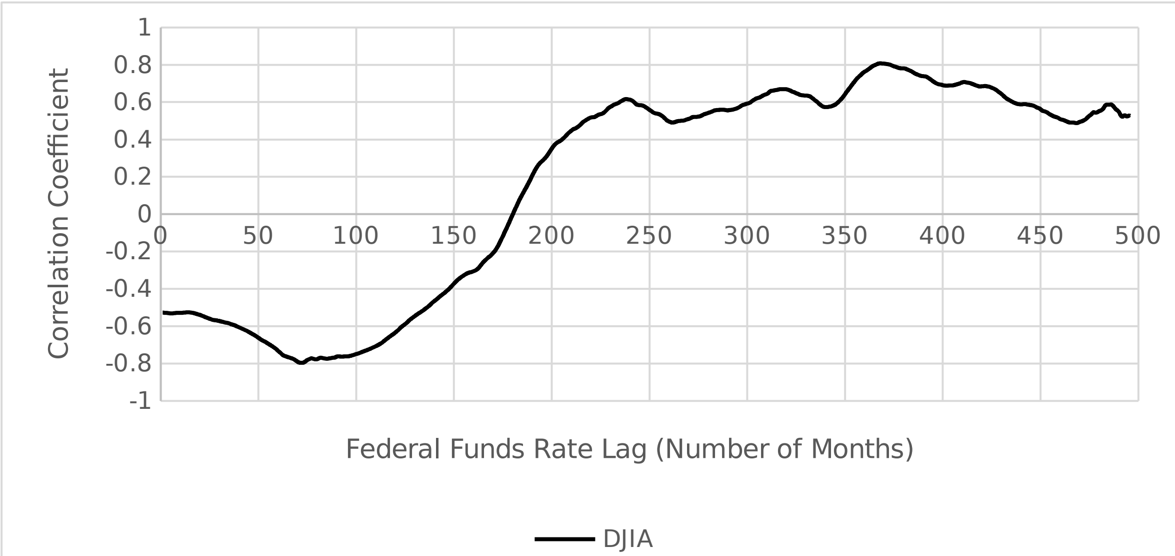

Figure 1. The Greenspan time period September 1987 to January 2006. The r-squared values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

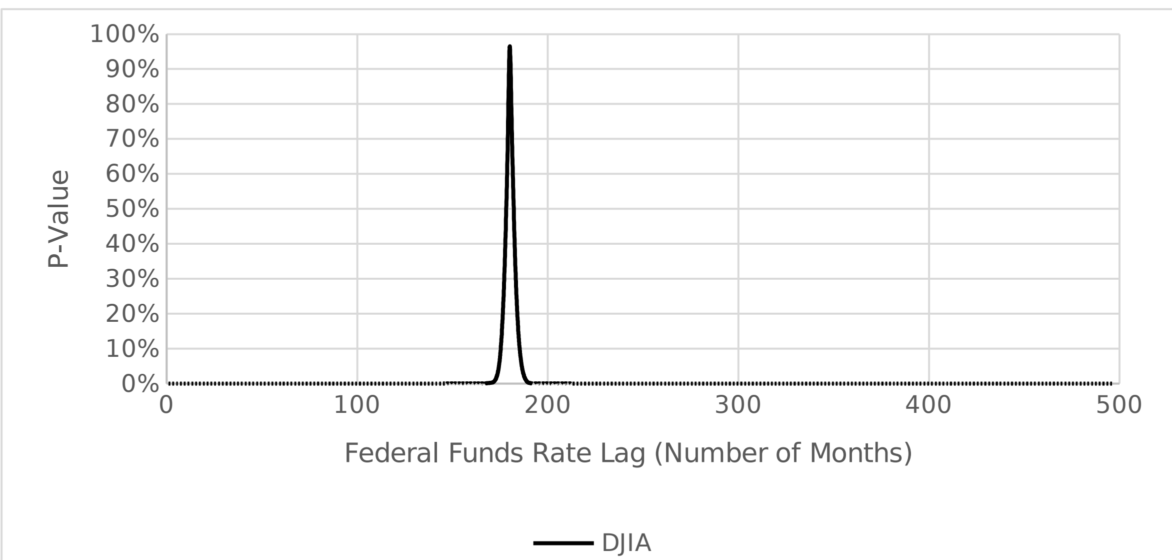

Figure 2. The Greenspan time period September 1987 to January 2006. The correlation coefficient values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

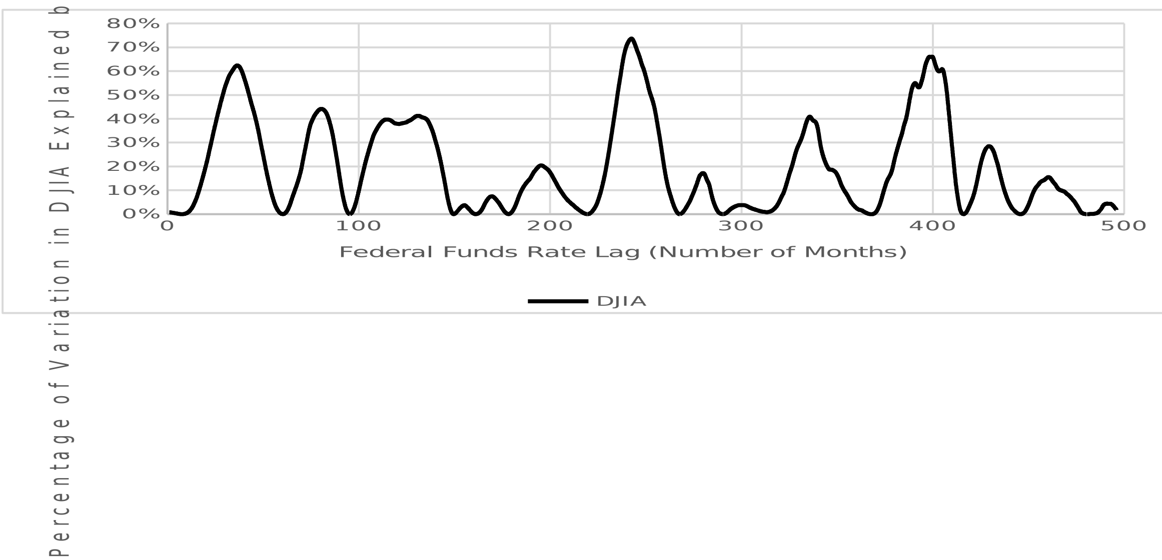

Figure 3. The Greenspan time period September 1987 to January 2006. The p-values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

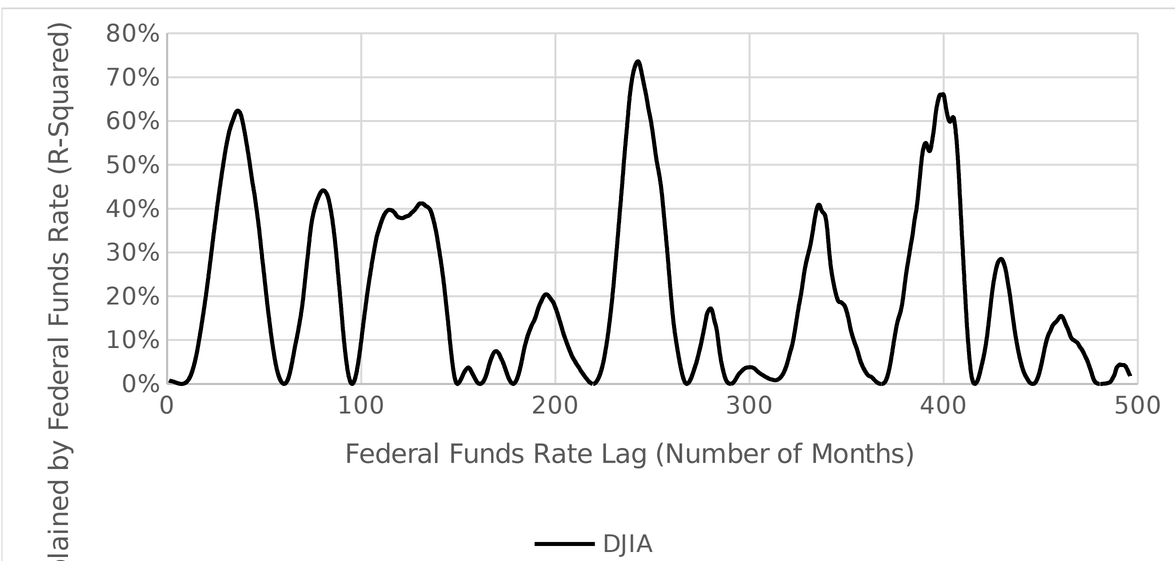

Figure 4. The Bernanke time period February 2006 to February 2014. The r-squared values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

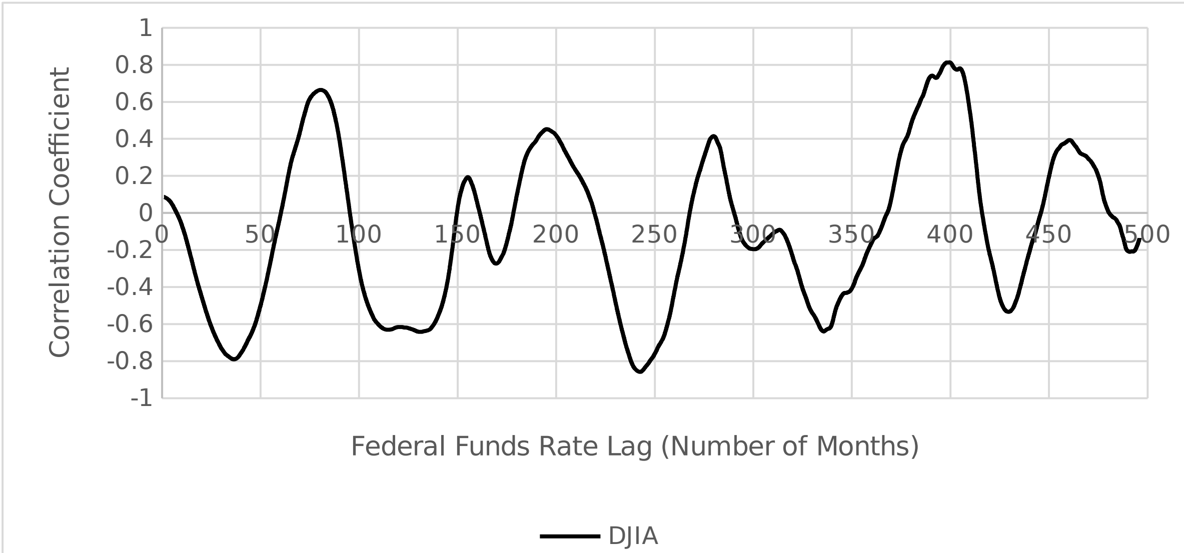

Figure 5. The Bernanke time period February 2006 to February 2014. The correlation coefficient values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

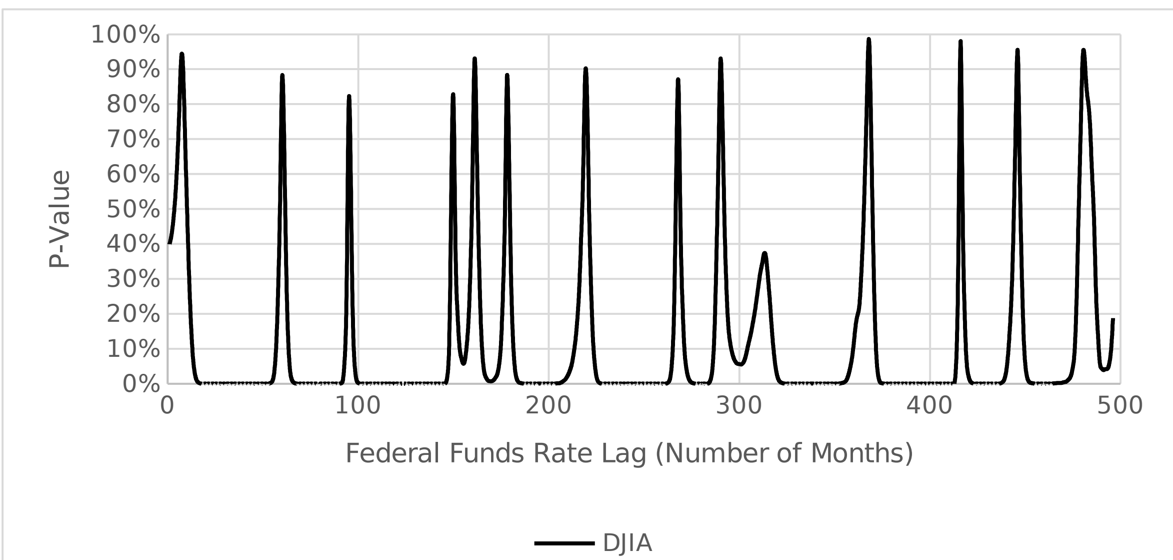

Figure 6. The Bernanke time period February 2006 to February 2014. The p-values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

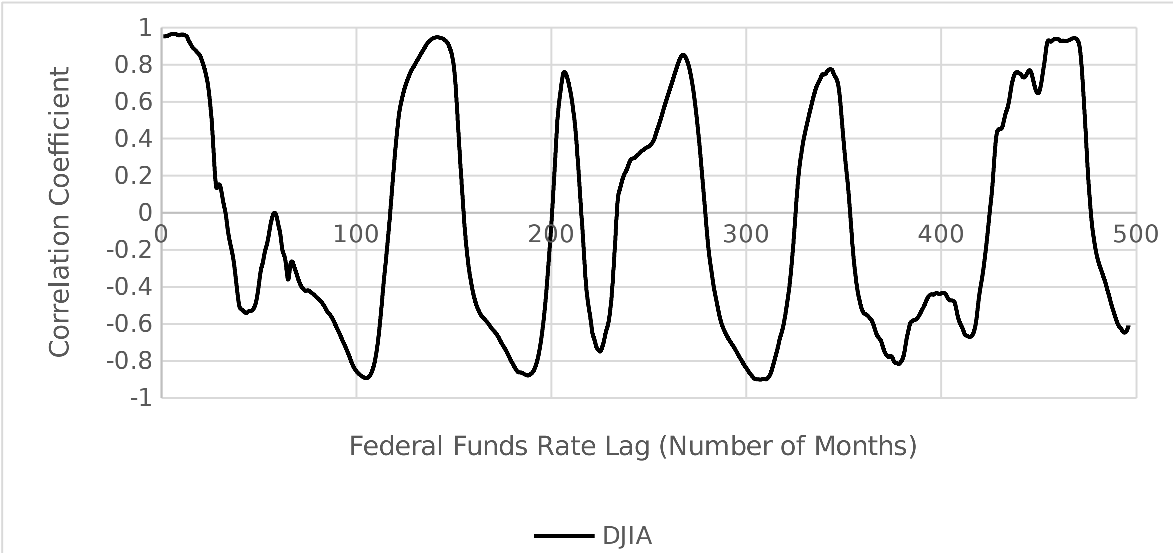

Figure 7. The Yellen time period March 2014 to February 2018. The r-squared values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

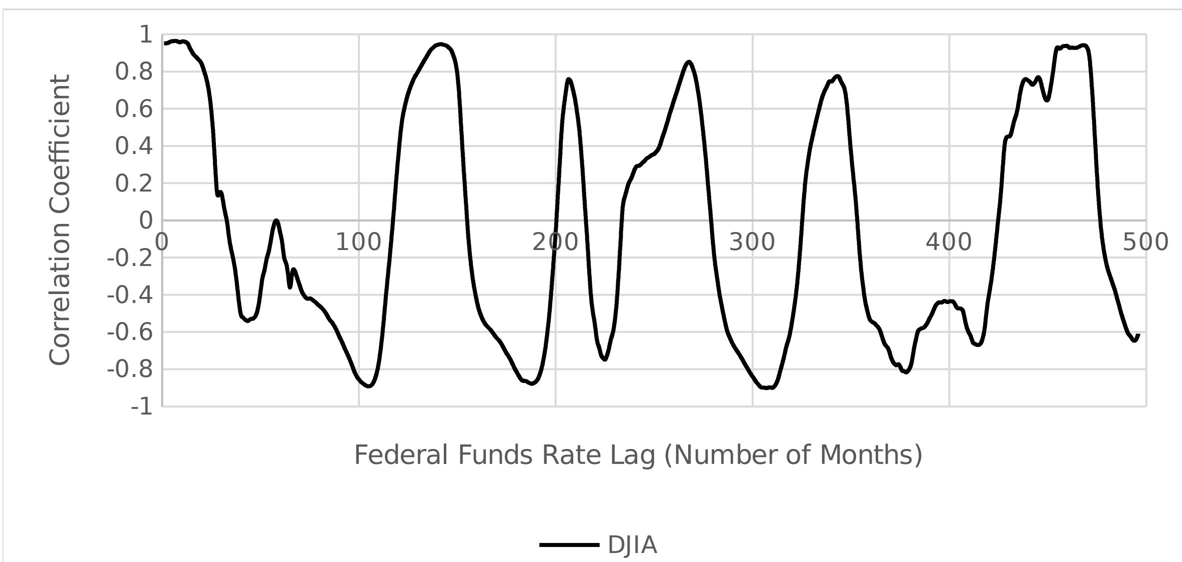

Figure 8. The Yellen time period March 2014 to February 2018. The correlation coefficient values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.

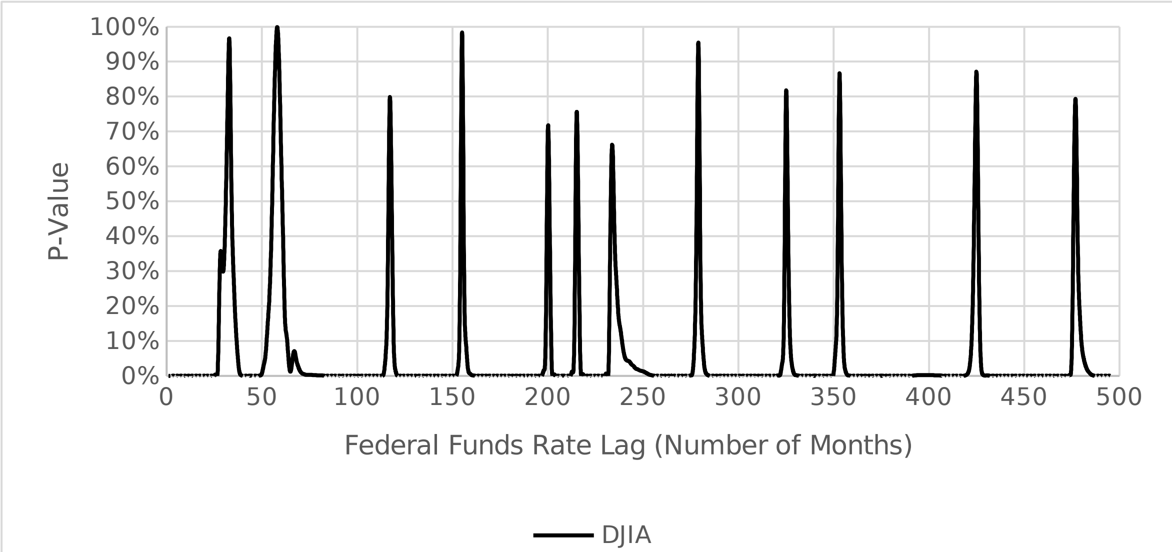

Figure 9. The Yellen time period March 2014 to February 2018. The p-values are on the vertical axis, and the number of months of lag in the interest rate is on the horizontal axis.