# MAT 303 Module Three Problem Set Report

Second Order Models

Nicholas Kreuziger

Nicholas.kreuziger@snhu.edu

Southern New Hampshire University

## **1. Introduction**

The dataset being explored is a historical dataset about wage growth in the labor force in addition to other related data points. This dataset can be analyzed to determine wage growth patterns. Governments and other interested parties can use these analyses to help steer political and corporate agendas in different directions. The table below summarizes the dataset’s variables and the meaning of those variables.

This data will be utilized with a Correlation Analysis and complete Second Order Regression Model. There will be two Regression Models with varying Qualitative and Quantitative datapoints.

## **2. Data Preparation**

This dataset has 5 columns of data and 99 rows. The columns are wage growth, inflation, unemployment, indication of recession and the GDP growth rate.

| **Variable** | **What does it represent?** |
| --- | --- |
| wage\_growth | Wage growth rate |
| inflation | Rate of inflation |
| unemployment | Unemployment rate |
| economy | Economy in recession or not in recession |
| gdp | GDP growth rate |

## **3. Quadratic (Second Order) Model with One Quantitative Variable**

### Correlation Analysis

**Chart, scatter chart

Description automatically generated**

Above is a scatterplot of wage growth and inflation. The observe trend is not linear, there appears to be a peak around an inflation of 9. At 10 a downtrend emerges. Since a non-linear relationship exists a Quadratic second order model would be more appropriate, as a Quadratic second order model has a non-linear curved shape. A Quadratic second order model incorporates the original predictor variable while also treating the squared value of that predictor variable as an additional predictor variables with a respective coefficient. i.e β1X+ β2X2

### Reporting Results

We will create a second-order regression model with the quantitative variable inflation to as the predictor variable. The response variable will be wage growth. This will be model 1. The general form of our regression model wherein y = wage growth and X = inflation is:

The resulting model computations for model 1 using statistical software are below:

Text

Description automatically generated with low confidence

From these computations the resulting R-Squared *R-Squared* and adjusted R-Squared *Adjusted R-Squared* values are,

86.14% and 85.85%. R-Squared is the variation in the Dependent Variable (Wage Growth) explained by the independent variables. These numbers indicate you can predict 86.14% of the variance in wage growth by inflation and inflation squared. Adjusted R-Squared is the variation explained by only the independent variables that significantly help in explaining the dependent variable, Adjusted R-Squared penalizes adding independent variables that do not help in predicting the dependent variable. The Adjusted R-Squared value only differing by 0.29% (86.14%-85.85%) indicates that the variables currently residing in the model are not significantly penalizing variance accountability.

The beta estimates indicate that the general form of the model becomes:

The beta estimates beta and beta squared are:

β1 : 1.8108

β2 : -0.0813

Beta 1 has no significant meaning in the presence of a squared term, Beta 2 value indicates upwards or downwards concavity. Since Beta 2 is a negative value, it indicates downwards concavity.

### Evaluating Model Significance

Evaluating this model’s significance at a 5% level, we will be using the ANOVA (Analysis of Variance) F-Test. An F-test applied to a multiple regression model is used to determine if the overall model and its variables are collectively influencing the response variable in a statistically significant manner. After determining the results of the F-Test, if a relationship exists you can evaluate individual relationships in the model with a multiple regression individual t-test. This is useful when you want to troubleshoot or assess the value of certain predictor variables in a model.

The resulting model computations for model 1 using statistical software are below:

Text

Description automatically generated with low confidence

The Overall F-Test is testing if the response variable has a relationship with at least one of the predictor variables. The Significance Level of our test is 5%, mathematically α = 0.05.

The Null hypothesis, H0 is that the predictor variables have no relationship with the response variable. Mathematically stated, the beta coefficient of predictor variables inflation or inflation squared are equal to 0.

H0 : β1 = β2 = 0

The Alternative hypothesis, Ha is that at least one predictor variable has a relationship with the response variable wage growth. Mathematically stated, the beta coefficient of predictor variables inflation and inflation ^2 are not equal to 0.

Ha : βi ≠ 0 for *i* = 1, 2

The P-Value of the Overall F-Test is 2.2 e-16. Since P-Value < Significance Level the null hypothesis is rejected in favor of the alternative hypothesis. At least one predictor variable has a relationship to the response variable Wage Growth.

Individual t-tests can be used once the F-Test has determined the model has a predictor variable with a relationship to the response variable. Every individual t-test has a similar Null and Alternative Hypothesis. We will be testing with a significance level of 5%, mathematically α = 0.05.

The Null Hypothesis, H0, is that the predictor variable coefficient βi (i = 1, 2 for each individual test) = 0. The Alternative Hypothesis, Ha , is that the predictor variable coefficient βi  (i = 1, 2 for each individual test) ≠ 0.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | p-value | Significance level | Result |
| Inflation | 2e-16 | 0.05 | Reject Null Hypothesis |
| Inflation ^2 | 8.81e-09 | 0.05 | Reject Null Hypothesis |

The t-test results indicates both inflation and inflation squared are significant.

### Making Predictions Using Model

Using this model, we can make predictions about wage growth given the inflation value. Let’s say that inflation is 7.41.

The predicted wage growth with an inflation of 7.41 is 8.55916 with this model.

A prediction interval is used to predict what range a future individual observation may fall in all outcomes. A confidence interval is used to predict the range the average of the response variable will fall. In short, a prediction interval tries to account for all possible values with the model for a single observation while a confidence interval tries to predict where the average of the probability distribution will fall given specific values for the predictor variables. Confidence intervals will always be narrower than Prediction intervals since a Confidence interval is the estimation of a range of an average while the predictor interval is the range of values for a single instance.

Using this model, we can construct the confidence and prediction intervals for wage growth given a inflation value of 7.41.

The 95% prediction interval for wage growth in these economic conditions is,

Graphical user interface, text

Description automatically generated

The prediction interval lower and upper ranges are 6.4568 and 10.6616. If we were to wage growth in an economy experiencing inflation at 7.41 we would experience wage growth within this lower and upper range 95% of the time.

Table

Description automatically generated with medium confidence

The confidence interval lower and upper ranges are 8.2473 and 8.871. If we were to observe wage growth in an economy with an inflation of 7.41, we would arrive at an average Wage Growth value between 8.2473 and 8.871. This range is predicted to be a part of the true population 95% of the time.

## **4. Complete Second Order Model with Two Quantitative Variables**

### Reporting Results

We will create a second-order regression model with the quantitative predictor variables inflation and GDP growth. The response variable will be wage growth. This will be model 2. The general form of our regression model wherein y = wage growth and X1 = inflation and X2 = GDP growth is:

The resulting model computations for model 2 using statistical software are below: Table

Description automatically generated

From these computations the resulting R-Squared *R-Squared* and adjusted R-Squared *Adjusted R-Squared* values are,

91.13% and 90.65%. R-Squared is the variation in the Dependent Variable (Wage Growth) explained by the independent variables. These numbers indicate you can predict 91.13% of the variance in wage growth by inflation and GDP growth in a second-order model. Adjusted R-Squared is the variation explained by only the independent variables that significantly help in explaining the dependent variable, Adjusted R-Squared penalizes adding independent variables that do not help in predicting the dependent variable. The Adjusted R-Squared value only differing by 0.48% (91.13%-90.65%) indicates that the variables currently residing in the model are not significantly penalizing variance accountability.

The beta estimates indicate that the general form of the model becomes:

The beta estimates for the squared coefficients of inflation and GDP growth are:

β4: -0.02737

β5 : -0.00318

Squared Term Coefficient values indicates upwards or downwards concavity. Since both Terms Squared have a negative Coefficient, the concavity is downwards.

### Evaluating Model Significance

Evaluating this model’s significance at a 5% level, we will be using the ANOVA (Analysis of Variance) F-Test. An F-test applied to a multiple regression model is used to determine if the overall model and its variables are collectively influencing the response variable in a statistically significant manner. After determining the results of the F-Test, if a linear relationship exists you can evaluate individual relationships in the model with a multiple regression individual t-test. This is useful when you want to troubleshoot or assess the value of certain predictor variables in a model.

The resulting model computations for model 2 using statistical software are below:

Table

Description automatically generated

The Overall F-Test is testing if the response variable has a relationship with at least one of the predictor variables. The Significance Level of our test is 5%, mathematically α = 0.05.

The Null hypothesis, H0 is that the predictor variables have no relationship with the response variable. Mathematically stated, the beta coefficient of predictor variables inflation or inflation squared are equal to 0.

H0: β1 = β2 = β3 = β4 = β5 = 0

The Alternative hypothesis, Ha is that at least one predictor variable has a relationship with the response variable wage growth. Mathematically stated, the beta coefficient of predictor variables inflation, gdp, inflation ^2 and inflation:gdp are not equal to 0.

Ha: βi ≠ 0 for *i* = 1, 2, 3, 4, 5

The P-Value of the Overall F-Test is 2.2 e-16. Since P-Value < Significance Level the null hypothesis is rejected in favor of the alternative hypothesis. At least one predictor variable has a relationship to the response variable Wage Growth.

Individual t-tests can be used once the F-Test has determined the model has a predictor variable with a relationship to the response variable. Every individual t-test has a similar Null and Alternative Hypothesis. We will be testing with a significance level of 5%, mathematically α = 0.05.

The Null Hypothesis, H0, is that the predictor variable coefficient βi (i = 1, 2, 3, 4, 5 for each individual test) = 0. The Alternative Hypothesis, Ha, is that the predictor variable coefficient βi  (i = 1, 2, 3, 4, 5 for each individual test) ≠ 0.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | p-value | Significance level | Result |
| Inflation | 4.68e-07 | 0.05 | Reject Null Hypothesis |
| GDP Growth | 3.42e-06 | 0.05 | Reject Null Hypothesis |
| Inflation ^ 2 | 0.1706 | 0.05 | Fail Reject Null Hypothesis |
| GDP Growth ^2 | 0.5837 | 0.05 | Fail Reject Null Hypothesis |
| Inflation : GDP Growth | 0.0563 | 0.05 | Fail Reject Null Hypothesis |

The t-test results indicates both inflation and inflation squared are significant. For this model the terms Inflation and GDP Growth raised to a power, in addition to the Inflation and GDP Growth Interaction Term are not statistically significant.

### Making Predictions Using Model

Using this model, we can make predictions about wage growth given the inflation and gdp value. Let’s say that Inflation (X1) is 7.41 and GDP Growth (X2) is 9.59.

The predicted wage growth with an inflation of 7.41 and GDP growth of 9.59 is 8.45789 with this model.

A prediction interval is used to predict what range a future individual observation may fall in all outcomes. A confidence interval is used to predict the range the average of the response variable will fall. In short, a prediction interval tries to account for all possible values with the model for a single observation while a confidence interval tries to predict where the average of the probability distribution will fall given specific values for the predictor variables. Confidence intervals will always be narrower than Prediction intervals since a Confidence interval is the estimation of a range of an average while the predictor interval is the range of values for a single instance.

Using this model, we can construct the confidence and prediction intervals for wage growth given an inflation value of 7.41 and GDP growth of 9.59.

The 95% prediction interval for wage growth in these economic conditions is,

Graphical user interface, text, application

Description automatically generated

The prediction interval lower and upper ranges are 6.744 and 10.1718. If we were to observe wage growth in an economy experiencing inflation at 7.41 and GDP growth of 9.59 we would observe wage growth within this lower and upper range 95% of the time.

Table

Description automatically generated

The confidence interval lower and upper ranges are 8.1751 and 8.7407. If we were to observe wage growth in an economy with an inflation of 7.41 and GDP growth of 9.59, we would arrive at an average Wage Growth value between 8.1751 and 8.7407. This range is predicted to be a part of the true population 95% of the time.

## **5. Complete Second Order Model with One Quantitative and One Qualitative Variable**

### Reporting Results

We will create a second-order regression model with the quantitative predictor variable inflation and the qualitative predictor variable economy. A qualitative variable is not squared in a second order model because a qualitative variable is only 1 or 0, True or False. The response variable will be wage growth. This will be model 3. The general form of our regression model wherein y = wage growth and X1 = inflation and X2 = economy is:

The resulting model computations for model 3 using statistical software are below:

Text, table

Description automatically generated with medium confidence

From these computations the resulting R-Squared *R-Squared* and adjusted R-Squared *Adjusted R-Squared* values are,

87.38% and 86.7%. R-Squared is the variation in the Dependent Variable (Wage Growth) explained by the independent variables. These numbers indicate you can predict 87.38% of the variance in wage growth by inflation and GDP growth in a second-order model. Adjusted R-Squared is the variation explained by only the independent variables that significantly help in explaining the dependent variable, Adjusted R-Squared penalizes adding independent variables that do not help in predicting the dependent variable. The Adjusted R-Squared value only differing by 0.68% (87.38%-86.7%) indicates that the variables currently residing in the model are not significantly penalizing variance accountability.

The beta estimates indicate that the general form of the model becomes:

The beta estimates for the squared coefficients of inflation and recession are:

β4: -0.05175

β5: -0.16300

Squared Term Coefficient values indicates upwards or downwards concavity. Since both Terms Squared have a negative Coefficient, the concavity is downwards.

### Evaluating Model Significance

Evaluating this model’s significance at a 5% level, we will be using the ANOVA (Analysis of Variance) F-Test. An F-test applied to a multiple regression model is used to determine if the overall model and its variables are collectively influencing the response variable in a statistically significant manner. After determining the results of the F-Test, if a linear relationship exists you can evaluate individual relationships in the model with a multiple regression individual t-test. This is useful when you want to troubleshoot or assess the value of certain predictor variables in a model.

The resulting model computations for model 3 using statistical software are below:

Text, table

Description automatically generated with medium confidence

The Overall F-Test is testing if the response variable has a relationship with at least one of the predictor variables. The Significance Level of our test is 5%, mathematically α = 0.05.

The Null hypothesis, H0 is that the predictor variables have no relationship with the response variable. Mathematically stated, the beta coefficient of predictor variables inflation or inflation squared are equal to 0.

H0 : β1 = β2 = β3 = β4 = β5 = 0

The Alternative hypothesis, Ha is that at least one predictor variable has a relationship with the response variable wage growth. Mathematically stated, the beta coefficient of predictor variables inflation, gdp, inflation ^2 and inflation:gdp are not equal to 0.

Ha : βi ≠ 0 for *i* = 1, 2, 3, 4, 5

The P-Value of the Overall F-Test is 2.2 e-16. Since P-Value < Significance Level the null hypothesis is rejected in favor of the alternative hypothesis. At least one predictor variable has a relationship to the response variable Wage Growth.

Individual t-tests can be used once the F-Test has determined the model has a predictor variable with a relationship to the response variable. Every individual t-test has a similar Null and Alternative Hypothesis. We will be testing with a significance level of 5%, mathematically α = 0.05.

The Null Hypothesis, H0, is that the predictor variable coefficient βi (i = 1, 2, 3, 4, 5 for each individual test) = 0. The Alternative Hypothesis, Ha , is that the predictor variable coefficient βi  (i = 1, 2, 3, 4, 5 for each individual test) ≠ 0.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | p-value | Significance level | Result |
| Inflation | 5.45e-08 | 0.05 | Reject Null Hypothesis |
| Economy Recession | 0.10773 | 0.05 | Fail Reject Null Hypothesis |
| Inflation ^ 2 | 0.00167 | 0.05 | Reject Null Hypothesis |
| Inflation:Economy Recession | 0.39118 | 0.05 | Fail Reject Null Hypothesis |
| Inflation ^2 : GDP Growth | 0.59197 | 0.05 | Fail Reject Null Hypothesis |

The t-test results indicates both inflation and inflation squared are significant. For this model the terms Economy Recession, Inflation:Economy Recession and Inflation ^ 2:Economy Recession are not statistically significant.

### Making Predictions Using Model

Using this model, we can make predictions about wage growth given the Inflation and Recession State. Let’s say that Inflation (X1) is 7.41 and economy (X2) is 0, indicating No recession.

The predicted wage growth with an inflation of 7.41 and no recession is 8.437871 with this model.

A prediction interval is used to predict what range a future individual observation may fall in all outcomes. A confidence interval is used to predict the range the average of the response variable will fall. In short, a prediction interval tries to account for all possible values with the model for a single observation while a confidence interval tries to predict where the average of the probability distribution will fall given specific values for the predictor variables. Confidence intervals will always be narrower than Prediction intervals since a Confidence interval is the estimation of a range of an average while the predictor interval is the range of values for a single instance.

Using this model, we can construct the confidence and prediction intervals for wage growth given an inflation value of 7.41 and no Recession.

The 95% prediction interval for wage growth in these economic conditions is,

Table

Description automatically generated with medium confidence

The prediction interval lower and upper ranges are 6.3961 and 10.4796. If we were to observe wage growth in an economy experiencing inflation at 7.41 and no recession, we would observe wage growth within this lower and upper range 95% of the time.

Text

Description automatically generated with low confidence

The confidence interval lower and upper ranges are 8.1175 and 8.7582. If we were to observe wage growth in an economy with an inflation of 7.41 and no recession, we would arrive at an average Wage Growth value between 8.1175 and 8.7582. This range is predicted to be a part of the true population 95% of the time.

## **6. Conclusion**

In these analyses we have reviewed 3 separate models combining the inflation, GDP growth and economy (indication of recession TRUE or FALSE) in various ways. Recall that Inflation and GDP Growth are Quantitative while Recession is Qualitative.

The table below summarizes the use of variables in each model and the model’s associated R Squared and Adjusted R-Squared Values. When assessing variable use in order of precedence, always assume left to right labelling of x-variables. I.E for Model 2 you can assume X1 is inflation and X2 is GDP. For model 3 X1 is inflation and X2 is economy.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Inflation | GDP | economy | R-Squared | Adj R-Squared |
| Model 1 | X |  |  | 86.14% | 85.85% |
| Model 2 | X | X |  | 91.13% | 90.65% |
| Model 3 | X |  | X | 87.38% | 86.7% |

The model’s equations are as follows:

|  |  |
| --- | --- |
| Model | Equation |
| Model 1 |  |
| Model 2 |  |
| Model 3 |  |

Comparing the Confidence Intervals can help to further guide the conclusion. Since the Confidence interval is predicting the range of an average for the response variables over many iterations it can help to conclude which model is fitting the smallest range of averages. The model’s confidence intervals for an interval of inflation 7.41, GDP 9.59 and no recession where applicable. If a model does not have the variable, it is not applied. The table below summarizes the confidence interval results of all 3 models:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Fit | Low | High | Δ High - Low |
| Model 1 | 8.5592 | 8.2473 | 8.871 | 0.6237 |
| Model 2 | 8.4579 | 8.1751 | 8.7407 | 0.5656 |
| Model 3 | 8.4379 | 8.1175 | 8.7582 | 0.6407 |

I would recommend these models based upon the above tables with an emphasis on using Model 2. 91.13% accountability for variance of wage growth with the predictor variables inflation and GDP is excellent. Furthermore, the delta of high and low confidence intervals only being 0.5656 indicates the model will not sway erroneously in repeated wage growth applications. It is the smallest delta in confidence interval high and low ranges of the 3 models, albeit a minimal delta exists for all 3 and is likely not noteworthy in the larger scheme.