# MAT 303 Module Three Problem Set Report

Second Order Models

Kayla Sacks

Kayla.sacks@snhu.edu

Southern New Hampshire University

## **1. Introduction**

The statistical analysis I will be doing consists of using three different second order regression models using a historical economic data set to find how different variables affect wage growth. The first model will focus on the effect that unemployment has on wage growth. The second model will be a complete second order model with unemployment and GDP growth as the predictor variables. The third model with be another complete second order but will use economy, which is a qualitative variable, instead of GDP growth. I will perform both the overall F-test and individual T-test for each model to test if they each meet a 5% level of significance. The results from this analysis can be used by economists to predict how different situations could have an effect on the economy.

## **2. Data Preparation**

The original data set that is being used has 99 rows and 6 columns, or variables. For this analysis I will only be using 4 of the variables between three different models. The important variables that I will be using are:

* wage\_growth – The wage growth rate. This will be the dependent variable for all three models.
* unemployment – The unemployment rate. This will be a predictor variable for all three models.
* gdp – The GDP growth rate. This is the second predictor variable that will be used in the second model.
* economy – If the economy is in recession or not in recession. This is a qualitative predictor variable that will be used for the final model.

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

### Correlation Analysis

To start the analysis, the scatterplot below was created to show the relationship between wage growth and unemployment.

A graph with red dots

Description automatically generated

The scatterplot above clearly shows that as wage growth decreases, unemployment increases. We will use second order regression models for the analysis as a curve is clearly visible in the plot and a first order model will not provide the most accurate estimates as it does not account for the curvature as a second order model will.

### Reporting Results

The general form for this regression model is:

The prediction regression model is:

Next, we will create the second order regression model for wage growth with unemployment as the independent variable. The summary of the regression output is below.

A screenshot of a computer program

Description automatically generated

Using the output from the regression model, we will write the prediction model equation.

The (R-squared) value for this model is 0.9436 and the (Adjusted R-squared) for this model is 0.9424. The value tells us that the regression model explains 94% of the variation in wage growth using the predictor variable of unemployment.

The bata estimated for unemployment is -1.7432. This coefficient does not have an interpretation with any meaning since a squared term of the variable exists in the model. The bata estimate for is 0.0674. The coefficient is positive which indicates a curve with an upward concavity.

### Evaluating Model Significance

To evaluate the model’s significance, we will perform the overall F-test first to see if the model is significant at a 5% level of significance. For this test we will start by testing if the model rejects the null hypothesis () in favor of the alternative hypothesis (). The equations for these are:

From the regression model, we can see that the p-value is < 2.2e-16 which is less than our significance level of .05 which means that we will reject the null hypothesis in favor of the alternate hypothesis. The rejection of the null hypothesis for the F-test means that there is a significant relationship between the dependent variable wage growth and the predictor variables .

We will now perform the individual T-test to see if the p-values for predictor variables will reject the null hypothesis in favor of the alternate hypothesis by checking if they meet a 5% level of significance. The equation for each hypothesis is below for *i* = 1, 2:

Unemployment has a p-value < 2e-16 and has a p-value of 6.07e-15. Both p-values are below the 0.05 level of significance so they will reject the null hypothesis. From this we can say that both unemployment and have a statistically significant relationship to wage growth.

### Making Predictions Using Model

Make predictions using the regression model. Address the following questions in your analysis:

Using the prediction regression model, we can predict the wage growth if unemployment is 2.54.

Our calculated value is close to the predicted wage growth from the 95% prediction interval and confidence interval is 8.2414 and is within the upper and lower bounds for both intervals.

The 95% prediction interval for wage growth is (6.9071, 9.5758). This means that we can be 95% certain that the individual data point for wage growth will fall between the interval if unemployment is 2.54.

The 95% confidence interval for wage growth is (8.0936, 8.3893). This means that we can be 95% certain that the average wage growth will be between the interval if unemployment is 2.54.

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

### Reporting Results

The general form for the complete second order regression model which includes two predictor variables is:

The prediction equation for this regression model is:

Next, the second order regression model for wage growth with unemployment as the independent variable. The summary of the regression output is below.

A screenshot of a computer screen

Description automatically generated

Using the output from the regression model, we will write the prediction model equation.

The (R-squared) value for this model is 0.9587 and the (Adjusted R-squared) for this model is 0.9565. The value tells us that the complete second order regression model explains 96% of the variation in wage growth using the predictor variables of unemployment and GDP growth.

The beta estimate for and are 0.0377 and-0.0066 respectively. The estimate is positive which indicates that the curved relationship has a positive concavity whereas the is negative which indicates a negative concavity.

### Evaluating Model Significance

To find if the second model meets a 5% level of significance, we will carry out another F-test starting with the null hypothesis and alternate hypothesis. The equation for each hypothesis is below.

The overall p-value for the complete second order model is < 2.2e-16 which is less than the 0.05 level of significance so we will reject the null hypothesis in favor of the alternate hypothesis. The conclusion of this test tells us that at least one predictor variable is statistically significant to our dependent variable, wage growth. With the conclusion of the F-test, we will move onto the individual T-test to determine the variables that meet the 5% level of significance. We will once again use the null hypothesis and alternate hypothesis to test each variable for .

The values that reject the null hypothesis in the T-test are unemployment with a p-value of 8.26e-6, GDP with a p-value of 0.0468, and with a p-value of 0.0049. The interaction term between unemployment and GDP as well as both fail to reject the hull hypothesis which means they are not statistically significant to the dependent variable.

### Making Predictions Using Model

Using the prediction regression model, we can predict the wage growth if unemployment is 2.50 and GDP growth is 6.50.

Our calculated value is the same as the predicted wage growth from the 95% prediction interval and confidence interval of 7.806.

The 95% prediction interval for the wage growth is (6.6315, 8.9805). This means that we can be 95% certain that the individual data point for wage growth will fall between the interval if unemployment is 2.50 and GDP growth is 6.5.

The 95% confidence interval for the wage growth is (7.583, 8.0289). This means that we can be 95% certain that the average wage growth will be between the interval if unemployment is 2.50 and GDP growth is 6.5.

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

### Reporting Results

The general form for the complete second order regression model which includes two predictor variables with one being qualitative is:

The prediction equation for this regression model is:

The second order regression model for wage growth with unemployment as the quantitative predictor variable and economy as the qualitative predictor variable was then created and the summary of the regression output is below.

*A screenshot of a computer

Description automatically generated*

Using the output from the regression model, we will write the prediction model equation.

The (R-squared) value for this model is 0.9475 and the (Adjusted R-squared) for this model is 0.9446. The value tells us that the complete second order regression model explains 95% of the variation in wage growth using the predictor variables.

### Evaluating Model Significance

We will now carry out the final overall F-test for this analysis using the null hypothesis and alternate hypothesis to check if the model meets a 5% level of significance. The same equations will be used as with the previous F-test and is below.

The overall p-value for this model is < 2.2e-16 which is less than the 0.05 level of significance so we will reject the null hypothesis in favor of the alternate hypothesis. Next will be to conduct the individual T-test for each dependent variable. We will find the dependent variables that will meet the 5% level of significance with .

For this T-test, the only variable that failed to reject the null hypothesis is the interaction term between economy and with a p-value of 0.0512. The conclusion of this tells us that unemployment, economy, and the squared of each are statistically significant to the dependent variable.

### Making Predictions Using Model

Using the prediction regression model, we will calculate the predicted wage growth if unemployment () is 2.50 and the economy () is not in recession. We will start with the prediction equation with the outputs from the third regression model.

The calculated wage growth from the prediction equation is close to the value from the model prediction which is 8.3132.

The 95% prediction interval for the wage growth is (7.003, 9.6235). This means that we can be 95% certain that the individual data point for wage growth will fall between the interval if unemployment is 2.50 and the economy is not in recession.

The 95% confidence interval for the wage growth is (8.1573,8.4692). This means that we can be 95% certain that the average wage growth will be between the interval if unemployment is 2.50 and the economy is not in recession.

The prediction interval is wider than the confidence interval because the prediction interval also considers the uncertainty in the estimation.

## **6. Conclusion**

Based on the analysis and assuming that the sample size is a sufficient size, I would recommend using this model. Using all three models, we can clearly see a connection between our dependent variable, wage growth, and our predictor variables of unemployment, GDP growth, and whether the economy is in recession or not.

The correlation between the variables can be seen from the values from each regression model, with the values being 0.9436 for the first model, 0.9587 for the second model, and 0.9475 for the final model. The next reason that I would recommend this model is the results of the overall F-tests and individual T-tests. Each model rejected the null hypothesis in favor of the alternate for the F-tests which indicates they all have variables that are statistically significant to wage growth. For the individual T-tests, the first model did not have any variable fail to reject the null hypothesis, the second model had two variables, the interaction term between unemployment and GDP and , that failed to reject the null hypothesis, and the third only had one variable, the interaction term between economy and , that failed to reject the null hypothesis. This leads to the conclusion that unemployment, GDP growth, and economy are each statistically significant to wage growth.

The importance of this analysis is that economists can use the models to study and evaluate the effect that different variables, such as unemployment, GDP growth, or the economy, have on wage growth. They can then use the variables that are statistically significant to more accurately predict wage growth in different situations.