# MAT 303 Module Three Problem Set Report

Second Order Models

Ashley De Venuto

ashley.devenuto@snhu.edu

Southern New Hampshire University

## **1. Introduction**

The data set that is being explored for this assignment is regarding economic wage growth over the years. The results from the data can give insight on how inflation, unemployment, economy, and gdp can impact wage growth across the country. It can also help find solutions on how to increase wage growth. I will be creating a scatter plot the shows wage growth and unemployment. Next, I will conduct a Quadratic (Second Order) Model with One Quantitative Variable, testing its significance at 5% significance level, and conduct a prediction test using the model by testing the for wage growth if unemployment is at 2.54%, and testing the prediction and confidence intervals at 95%. Then, I will conduct a Second Order Model with Two Quantitative Variables, running the same test for significance at 5%, change the prediction test to wage growth if unemployment is 2.5 and gdp is 6.5, and testing the prediction and confidence intervals at 95%. Lastly, I will conduct a Second Order Model with One Quantitative and One Qualitative Variable, running the same test for significance at 5%, change the prediction test to test for wage growth when unemployment is 2.50 and economy is no recession, and testing the significance at 5% significance level.

## **2. Data Preparation**

There are six main variables within this data set. The variables are wage growth, inflation, unemployment, economy, education, and gdp. This means that there are six columns that represent the six variables. There are six records within the data set. This means there are six rows the represent the six records. For the analysis that I’m conducting, I will be focused on four of the six variables. Those variables being wage growth, unemployment, economy, and gdp.

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

### Correlation Analysis

A graph with red dots

Description automatically generated

The scatterplot shows a strong negative correlation between wage growth and unemployment. It’s negative because the direction of the points is going downward when unemployment goes up and wage growth goes down. The reason why the negative correlation is strong is due to how close the points are on the scatterplot. If I were to draw a line through the points, most of those points would be on that line. I believe that the second order regression order is more suitable in this case. This is due to the predictor variable(unemployment) may or may not have a non-linear relationship with the response variable (wage growth). The predictor variable is not only dependent on a response variable and can increase or decrease due to other factors.

### Reporting Results

The general form of the second-order regression model would look like:

The prediction equation of the second-order regression model would look like this:

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><mi>y</mi><mo>&#x2227;</mo></mover><mo>=</mo><msub><mover><mi>&#x3B2;</mi><mo>&#x2227;</mo></mover><mn>0</mn></msub><mo>+</mo><msub><mover><mi>&#x3B2;</mi><mo>&#x2227;</mo></mover><mn>1</mn></msub><mo>&#xA0;</mo><mi>x</mi><mo>+</mo><msub><mover><mi>&#x3B2;</mi><mo>&#x2227;</mo></mover><mn>2</mn></msub><mo>&#xA0;</mo><msup><mi>x</mi><mn>2</mn></msup></mstyle></math>","origin":"MathType for Microsoft Add-in"}

The second-order regression model will look this:

Rounded up to 4places after decimal.

The is 0.9436 and the is 0.9424. As you can see, both values are close together because there is only one input variable. The shows that there is 94% of the wage growth variant can be explained using this model while using unemployment as the predictor variable. I believe that the data fits this model. The beta estimate is -1.7432 for unemployment and 0.0674 for . Since *2* is a positive number, that means that there is a curved relationship with an upward concavity.

### Evaluating Model Significance

Null Hypothesis:

Alternative Hypothesis:

The p-value is 2.2e-16 for the overall F-test. This is lower than the 5% significance level. We will reject the null hypothesis and accept the alternative hypothesis. I can conclude from the results of the F-test that there is a statistically significant relationship between wage growth and unemployment.

To determine which term (s) have a significant relationship with the response variable, I ran another t-test. The null hypothesis looks like this:

The alternative hypothesis looks like this:

Unemployment has a p-value of 2e-16 and has a p-value of 6.07e-15. Both p-values are less than the 5% level of significance. Both variables have a statistically significant relationship with wage growth. This means that I don’t reject the null hypothesis.

### Making Predictions Using Model

When testing for the predicted wage growth if unemployment is 2.54 is 8.2414. The prediction interval is between 6.9071 and 9.5758. This means that there can be a 95% certainty that a data point representing wage growth can fall between that range when taking regression error into account. The confidence interval for wage growth when unemployment is 2.54 is 8.2414. The interval falls between 8.0936 and 8.3893 at 95% certainty for wage growth.

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

### Reporting Results

The general form and prediction equation of the second order regression model for wage growth as the response variable and unemployment and GDP growth as predictor variables are as follows:

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><mi>y</mi><mo>&#x2227;</mo></mover><mo>=</mo><mover><msub><mi>&#x3B2;</mi><mn>0</mn></msub><mo>&#x2227;</mo></mover><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>1</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>3</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>4</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msup><msub><mi>x</mi><mn>1</mn></msub><mn>2</mn></msup><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>5</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msup><msub><mi>x</mi><mn>2</mn></msub><mn>2</mn></msup></mstyle></math>","origin":"MathType for Microsoft Add-in"}

Here is the prediction model equation using the outputs obtained from the R script:

All numbers are rounded to 4 decimal places.

The value for the model is 0.9587 and the for the model is 0.9565. The tells me that there is about 96% (rounded up) of the variance in wage growth that can be explained this model where unemployment and gdp are used as predictor variables. The beta estimates for (rounded to 4 decimal places) and the beta estimate for  (rounded to 4 decimal places). There is an upward concavity when it comes to since it’s positive and there’s a downward concavity for since it’s negative.

### Evaluating Model Significance

To find out if this model is significant at a 5% level of significance, I had to conduct an overall F-test. The first part of that is by identifying the null and alternative hypothesis. The null hypothesis looks like this:

The alternative hypothesis looks like this:

The p-value of this model is 2.2e-16. This is less than 0.05 or 5% therefore I will reject the null hypothesis and accept the alternative hypothesis. This also means that at least one of the predictor variables has a significant relationship with wage growth. To determine which of the predictor variables are significant at a 5% level of significance, I must conduct individual T-tests. The null hypothesis looks like this:

The alternative hypothesis looks like this:

The unemployment p-value is 8.26e-06. This is lower than the 0.5 or 5% significance level. This means that the null hypothesis is rejected, and the alternative hypothesis is accepted. The GDP p-value is 0.0468 (rounded to 4 decimal places). This is equal (0.0468 rounded up) to the 0.5 or 5% significance level. This means that I accept the null hypothesis. The p-value is 0.0049 (rounded to 4 decimal places). This is less than the 0.05 or 5% significance level. This means that I reject the null hypothesis and accept the alternative hypothesis. The p-value is 0.1282 (rounded up 4 decimal places). This is higher than the 0.05 or 5% significance level. The unemployment:gdp p-value is 0.7668 (rounded up 4 decimal places). With this information from the results, I can conclude that unemployment and have a significant relationship with wage growth at a 5% significance level. I can also conclude that GDP, , and unemployment:gdp don’t have a significant relationship with wage growth at a 5% significance level.

### Making Predictions Using Model

The predicted wage growth when unemployment is 2.50 and GDP growth is 6.50 is 7.806. The wage growth at a 95% predicted interval is between 6.6315 and 8.9805. This means that I can be 95% certain when keeping in mind any regression error, that a point representing wage growth will fall somewhere between that range of numbers. The wage growth at a 95% confidence interval is 7.583 and 8.0289. This means that I can be 95% certain that data points representing wage growth will fall between the range of numbers.

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

### Reporting Results

The general form of the second order regression model looks like this:

The prediction equation looks like this:

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><mi>y</mi><mo>&#x2227;</mo></mover><mo>=</mo><mover><msub><mi>&#x3B2;</mi><mn>0</mn></msub><mo>&#x2227;</mo></mover><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>1</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>3</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msub><mi>x</mi><mn>1</mn></msub><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>4</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msubsup><mi>x</mi><mn>1</mn><mn>2</mn></msubsup><mo>+</mo><mover><msub><mi>&#x3B2;</mi><mn>5</mn></msub><mo>&#x2227;</mo></mover><mo>&#xA0;</mo><msubsup><mi>x</mi><mn>1</mn><mn>2</mn></msubsup><mo>&#xA0;</mo><msub><mi>x</mi><mn>2</mn></msub></mstyle></math>","origin":"MathType for Microsoft Add-in"}

The second order regression model for wage growth using unemployment and economy as predictors looks like this:

All numbers were rounded to four decimal places.

The for the model is 0.9475 and the for the model is 0.9446. The tells me that approximately 95% of the wage growth variance can be explained with a model that uses unemployment and economy as the predictor variables.

### Evaluating Model Significance

The null hypothesis looks like this:

The alternative hypothesis looks like this:

The p-value of the model is 2.2e-16. This is less than the significance level of 0.05 or 5%. This means that the null hypothesis is rejected, and the alternative hypothesis is accepted. This also means that at least one of the variables has a significant relationship with wage growth. Therefore, this model is significant at a 5% level of significance. To figure out which variables are significant to wage growth, I had to conduct individual T-tests to check if they were significant at a 5% level of significance. The null hypothesis looks like this:

The alternative hypothesis looks like this:

The individual T-test shows the following results:

The p-value for unemployment is 2e-16 which is less than the 0.05 or 5% level of significance. The p-value for economyrecession is 0.0142 which is less than the 0.05 or 5% level of significance. The p-value for I(unemployment^2) is 1.24e-06 which is less than the 0.05 or 5% level of significance. The p-value for unemployment:economyrecession is 0.0272 which is less than the 0.05 or 5% level of significance. The p-value for economyrecession:I(unemployment^2) is 0.0512 which is equal to the 0.05 or 5% level of significance. This means that all the variables except for economyrecession:I(unemployment^2) has a significant relationship with wage growth at a 5% level of significance.

### Making Predictions Using Model

The predicted wage growth if unemployment is 2.50 and the economy is not in a recession is 8.3132. The prediction interval at 95% is between 7.003 and 9.6235. This means that there is a 95% certainty that a data point for wage growth will be within that range of numbers. The confidence interval at 95% falls between 8.1573 and 8.4692 when it comes to wage growth. The prediction interval is wider than the confidence interval because of the uncertainty around a single value while calculating the sample for the prediction interval.

## **6. Conclusion**

Based on the analysis that I have conducted and assuming the sample size is sufficiently large, I would recommend using this model. The data used while conducting the analysis fit well with the model. Each model had at least one variable that was significant with wage growth. This model could be used to determine what other factors may play a role in wage growth. This model can also help with improving wage growth by pinpointing which variables are lowering wage growth. By determining this, the government can begin to fix those areas to help raise wage growth.