# MAT 303 Module Three Problem Set Report

Second Order Models

Justin Farquhar

Justin.farquhar@snhu.edu

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## **1. Introduction**

The data set that we are exploring is one containing various historical data to study wage growth of the labor force. The results of this analysis could be used to observe trends and try to figure out what caused outliers or see what has the largest effect on the economy. For this problem set we will be using multiple second order regressions utilizing varying amounts of quantitative and qualitative values.

## **2. Data Preparation**

This data set contains fewer variables and in such, I would argue that all of them are important in this case. Wage growth is the main variable that we will determining how the other variables affect it. We do not analyze inflation in our own analysis, but it was done on the example code in the problem set. Even though we are not looking through inflation, it would be considered important as it adjusts buying power of people, making wage growth much more meaningful and necessary. Unemployment can have a substantial effect on wage growth as if someone does not have a job they would not have growth, but a company may be more likely to give raises to their employees after consolidation. Economy shows if the economy was in recession or not in recession, which could also be substantial, as being in a recession can have massive impacts on the economy, let alone individual people. GDP should correlate just about perfectly with wage growth in a perfect world, but this is usually not the case as just because the GDP is doing well, individual companies may not be. This data set contains 6 columns and 99 rows.

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

### Correlation Analysis

A graph with red dots

Description automatically generated

There is a negative relationship between wage growth and unemployment to which wage growth decreases as unemployment goes up. A first order relationship could get fairly close to working in this data set, but I believe a second order model was the right choice to give a much more accurate picture given the slight curve in the data points. If you were to use a first order model in this case, a lot of the data would be fairly far off of the linear line, leading to errors and misinterpreted results.

### Reporting Results

*Report the results of the regression model. Address the following questions in your analysis:*

General Form

Prediction Equation

The R squared value for this model is 0.9436, meaning that these variables account for 94.36% of the variance of wage growth. Adjusted R squared is 0.9424, which means that the two values are close together, indicating that the predictors in the model are appropriate for explaining the variance and the model is not overfitted. The beta values for unemployment and unemployment^2 are on the smaller side, but it makes sense as the lead beta value is only 12.2342 and a negative wage growth would be very unlikely to happen. Unemployment^2 is going to be a much higher number, which makes the much smaller beta number being appropriate to not make the wage growth quickly grow out of control.

### Evaluating Model Significance

Looking at the p-value of the model, it returned 2.2e-16, which indicates that this model is extremely significant. Looking at the other variables, unemployment is 2e-16, indicating that it is significant. Then for unemployment^2 it returned 6.07e-15 which also indicates that it is significant.

### Making Predictions Using Model

A screenshot of a graph

Description automatically generated

Looking at this screenshot, the predicted wage growth if the unemployment were 2.54 is 8.2414. The 95% prediction interval is 6.9071 to 9.5758. Meaning that there is 95% confidence that the actual wage growth will fall between these numbers. The confidence interval is 9.0936 to 8.3893. Meaning that there is a 95% confidence that the true mean value will fall between these two values.

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

### Reporting Results

General Form

Prediction Equation

The R squared value for this model is 0.9587, meaning that these variables account for 95.87 % of the variance of wage growth. Adjusted R squared is 0.9565, which means that the two values are close together, indicating that the predictors in the model are appropriate for explaining the variance and the model is not overfitted. For this model, most of the beta estimates are lower, however with having multiple predictors in the model, each can only account for so much of the overall wage growth, especially when the initial beta value is only 8.9894.

### Evaluating Model Significance

The model returned a p-value of 2.2e-16, indicating that it is significant. Unemployment has a p-value of 8.26e-6, showing that it is significant. GDP has a p-value of 0.4682, indicating that it is significant at the 5% level of significance. Unemployment^2 has a p-value of 0.00489, indicating that it is significant at the 5% level of significance. GDP^2 has a p-value of 0.12815, indicating that it is not significant at the 5% level of significance. Unemployment with GDP has a p-value of 0.76678, indicating that it is not significant at the 5% level of significance.

### Making Predictions Using Model

A screenshot of a cell phone

Description automatically generated

Looking at this screenshot, the predicted wage growth if the unemployment were 2.50 and GDP growth were 6.50 is 7.806. The 95% prediction interval is 6.6315 to 8.9805. Meaning that there is 95% confidence that the actual wage growth will fall between these numbers. The confidence interval is 7.583 to 8.0289. Meaning that there is a 95% confidence that the true mean value will fall between these two values.

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

### Reporting Results

General Form

+

Prediction Equation

The R squared value for this model is 0.9475, meaning that these variables account for 94.75 % of the variance of wage growth. Adjusted R squared is 0.9446, which means that the two values are close together, indicating that the predictors in the model are appropriate for explaining the variance and the model is not overfitted. Looking at the beta values already shows that the predictors are somewhat working individually against each other to adjust wage growth which could lead to less significance in the end.

### Evaluating Model Significance

The p-value for this model is 2.2e-16, indicating that it is significant at the 5% level of significance. The p-value for unemployment was 2e-16, indicating that it is significant at the 5% level of significance. The p-value of economyrecession is 0.0142, indicating that it is significant at the 5% level of significance. The p-value of unemployment^2 is 1.24e-6, indicating that it is significant at the 5% level of significance. The p-value of unemployment and economyrecession is 0.0272, indicating that it is significant at the 5% level of significance. The p-value for economyrecession and unemployment^2 is 0.0512, indicating that it is not significant at the 5% level of significance.

### Making Predictions Using Model

A screenshot of a cell phone

Description automatically generated

Looking at this screenshot, the predicted wage growth if the unemployment were 2.50 and the economy not being in recession is 8.3132. The 95% prediction interval is 7.003 to 9.63235. Meaning that there is 95% confidence that the actual wage growth will fall between these numbers. The confidence interval is 8.1573 to 8.4692. Meaning that there is a 95% confidence that the true mean value will fall between these two values.

## **6. Conclusion**

Assuming that the sample size is sufficiently large, I would argue that all three of these models are appropriate to be used depending on the circumstance and depending on which predictors that you wanted to utilize to determine wage growth. Each of the models were very significant, each just contained one or two predictors that were not significant at the 5% level of significance, but this should only marginally affect the result. These results can easily be implanted to predict the outcome of wage growth based off of the predictors selected. This will give a rough idea of what wage growth should be, which allows for preparing for whichever outcome.