# MAT 303 Module Three Problem Set Report

Second Order Models

Zachary Tankersley

Zachary.tankersley@snhu.edu

Southern New Hampshire University

## **1. Introduction**

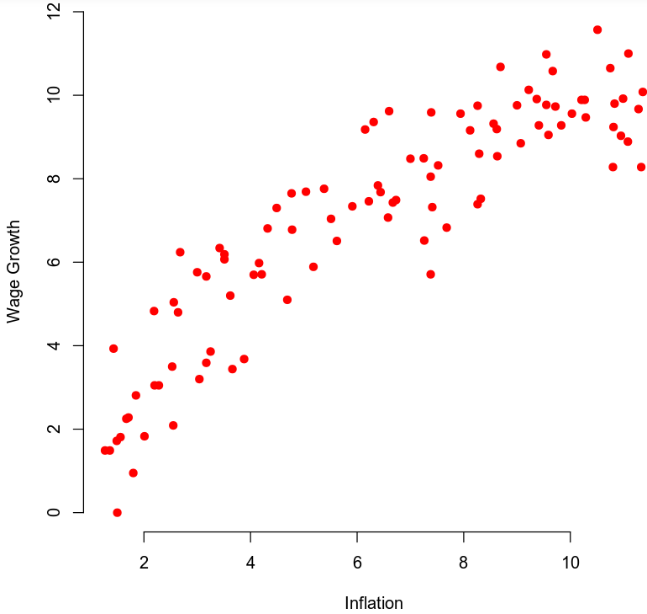
This data set is a historical data set containing information about wage growth in the labor force and can/will possibly be used to make gain insight into how various factors such as inflation and unemployment influence wage growth. To analyze this data and problem set, a second-order model using quantitative and/or qualitative variables will be employed.

## **2. Data Preparation**

The response variable in each analysis and model will be wage growth while using inflation, GDP, and the economy as the various predictor variables. There are 100 rows, and 6 columns present in the source data set, with the top row containing header/title values. In practice, we will be using 99 (all) of the rows and all 6 columns.

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

### Correlation Analysis

****

Based on the scatterplot of wage growth against inflation a *slight* curvature can be seen towards the upper limits, however, at this stage, it is unclear whether or not more benefit would be seen in a first-order model. Based on this initial inquiry, a second-order model is recommended.

### Reporting Results

The general form of the regression model that will be used for wage growth vs. inflation will be:

The true model after using summary analysis function results will be:

From the summary analysis function, we can also see that we have an R-Squared value of 0.8614 and an Adjusted R-Squared value of 0.8585. These values indicate that our model works *decently well*, however, I would recommend trying to add in and modify which predictors are used in the final model to hopefully increase those values. The beta estimates for inflation and inflation2 are 1.81077 and -0.088129 respectively. This would indicate that inflation has a strong positive relationship with wage growth, while inflation2 has a negative relationship with wage growth.

### Evaluating Model Significance

To carry out an overall F-test and determine if the model is significant at 5%, we must first determine the null and alternative hypotheses for the model which are displayed as:

From this we can gather that based on a P-value of 2.2e-16 which is well under the 0.05 or 5% significance level, we may reject the null hypothesis in favor of the alternative hypothesis by saying that a statistically significant relationship does exist between wage growth and inflation. To determine if the variables are significant, we must run an individual T-test on each variable after identifying the null hypothesis which is:

The P-values shown for inflation and inflation2 are 2e-16 and 8.81e-09 respectively, which are both well under the 5% significance threshold. Going back to our hypothesis, we can reject the null in both cases in favor of the alternative which means both variables *are* significant to our model and pass the individual T-tests.

### Making Predictions Using Model

When using the model to predict wage growth for an inflation rate of 7.41%, the model returns a value of 8.5591%. the 95% prediction interval for wage growth is [6.456768, 10.66156] meaning that 95% of predicted values should fall within that range. The 95% confidence interval is [8.247337, 8.870991] meaning that the average of all values predicted by the model should fall in that range.

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

### Reporting Results

The general form of a second-order regression model with wage growth as the response variable and inflation *and* GDP as predictor variables is:

The true form of the model, using figures received from a summary analysis function is:

The R-squared and Adjusted R-Squared that we receive from the summary analysis function are 0.9113 and 0.9065 respectively. Judging from these values we can conclude that these two variables work well within this model, producing a fairly accurate outcome. This model's R values are higher than the last meaning it is likely more suited for this task, explaining ~90% of the variance in wage growth. The beta estimates for GDP2 and inflation2 are -0.003176 and -0.027371 respectively, hinting at a negative relationship between both of the variables and the response variable causing a downward curvature.

### Evaluating Model Significance

To determine the model and term significance we must again come up with null and alternative hypotheses to compare against, which would look identical to the hypotheses from the last model. Conducting an overall F-test at a significance rate of 5%, we can view our P-value of 2.2e-16 which is very much under our threshold, causing us to favor the alternative hypothesis and say that a statistically significant relationship does exist between wage growth and least one variable. When conducting individual T-tests on the variables we see P-values for inflation are 4.68e-07, for GDP are 3.42e-06, for inflation2 are 0.1706, for GDP2 are 0.5837 and finally for inflation vs GDP are 0.0563. While some of the variables are above the 5% threshold for significance, both inflation and GDP *are* significant.

### Making Predictions Using Model

When we make a prediction using this model and assuming an inflation rate of 7.41% and GDP growth of 9.59%, we are returned with a wage growth value of 8.4579%. the 95% prediction interval is [6.743987, 10.17179] meaning that 95% of all predicted values should fall within that range. The 95% confidence interval is [8.175066, 8.740712] meaning that the average of all predicted values should fall within that range.

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

### Reporting Results

To model a second-order regression model using wage growth as the response variable, inflation as a quantitative predictor variable, and economy as a qualitative predictor variable we can write the general form as follows:

Using the results of the summary analysis function we can input the beta values and come up with the model in its true form:

The summary analysis also yields the R-Squared and Adjusted R-Squared values of 0.8738 and 0.867, respectively. This would indicate that this model does “fit” and predict our data fairly well as it explains about 87% of the data.

### Evaluating Model Significance

To determine the significance of this model and its individual variables, we must conduct an overall F-test and individual T-test(s) at a 5% significance level while also determining P-values. To do this we must first determine the null and alternative hypotheses. Luckily, they remain the same in this scenario as they did in the last two. For the model hypothesis:

For the individual term significance:

The P-vale for the model is 2.2e-16 which is *much* lower than our significance threshold of 0.05 which would indicate that we should reject the null hypothesis in favor of the alternative and can conclude that at least one predictor variable is significantly related to the wage growth variable. For the individual T-tests we are returned with the values 5.45e-08 for inflation, 0.10773 for economy, 0.00167 for inflation2 , 0.39118 for inflation:economy, and finally 0.59197 for inflation2 :economy. When compared to our significance threshold of 0.05 we can see that no variable(s) save for inflation and inflation2 meet the standard for inclusion.

### Making Predictions Using Model

When using the previous model to predict wage growth if inflation is at 7.41% and the economy is *not* in a recession, we are returned with a wage growth value of 8.4379%. The 95% prediction interval is [6.39611, 10.47963] meaning that 95% of predicted values should fall within that range. The 95% confidence interval is [8.117521, 8.758221] meaning that the average of all values predicted by this model should fall within that range. The prediction interval is wider than the confidence interval because the prediction interval is predicting a range of values, while the confidence interval is predicting the mean of that range.

## **6. Conclusion**

Assuming the sample size is sufficient (~n=100), in their current forms, I would likely recommend the second model over the first two for the simple fact that the R-Squared and Adjusted R-Squared values are higher, which means that the variance in the response variable (wage growth) is more accurately explained by that model, slightly more so than the other two models. With that said, if “dead accuracy” is not necessarily the final goal, and some other nuance may be explained by the first and third models, then that could also influence the decision on which model to go with. When considering the significance of certain variables in the different models, several did not meet the critical significance thresholds (often missing by a large margin, especially regarding the third model), indicating that some of the *other* variables included in this data set which were *not* analyzed may explain the “curvature” in the scatterplot more accurately than the variables that *were* analyzed. This model could be used for many reasons, likely pertaining to what impacts these specific variables impact wage growth. For instance, if the United States government wanted to gain insight into how rising or falling inflation, and GDP can influence the rate of wage growth, or whether or not the economy is in a recession influences wage growth.