# MAT 303 Module Three Problem Set Report

Second Order Models

Catherine Taft

Catherine.taft@snhu.edu

Southern New Hampshire University

## **1. Introduction**

Apparently I’ve been fired from the car company after they realized I don’t understand or care about cars. I think “rear axle ratio” was what did me in. I’m now working for the government analyzing economic and labor statistics, which is much more in my vein of interest. In this analysis we’ll be looking at wage growth patterns alongside a variety of economic factors. This is important for understanding the relationship between wage growth and these conditions, both in a descriptive and predictive sense.

We’ll begin by plotting some basic relationships between these variables, and then doing some quadratic multiple regression models with different combinations of qualitative and quantitative variables.

## **2. Data Preparation**

The variables we’ll be analyzing for this problem set are wage growth (the response variable), GDP, inflation rate, and economy (in recession or out). Most of the variables in the data set are pretty reasonable to examine in relation to wage growth but those are good ones to start with.

The data set has 99 rows and 6 columns. I’m not sure how it’s organized (i.e. the last data set was organized by car model) because there doesn’t seem to be a row for country or year or anything. So I’m not sure if GDP refers to a country’s GDP at the time the data was collected or the US’s GDP in a given year.

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

### Correlation Analysis

We’re going to start with a scatterplot between wage growth and inflation:

Chart, scatter chart

Description automatically generated

At first glance this looks like it could be linear and applicable to a first order model, but at second glance, if you consider where you’d place a regression line, you can see that it curves up and then back down slightly, or even sort of flattens out at the end, so we’d probably want to use a second order model on this. In general this makes sense; the connection between wage growth and GDP is fairly obvious, but it makes sense to see a slowing of that growth eventually.

### Reporting Results

The equation for the second order model we’re using is:

We ran the model in R and received the following output:

Table

Description automatically generated

So based on this, we can create the model for wage growth using inflation as the predictor variable:

The R-squared and Adjusted R-squared values are 0.8614 and 0.8585 respectively, which means that roughly 86% of the variation in wage growth is explainable by this specific model. That is a good number and indicates a reliable model.

The beta estimates for inflation and are 1.81077 and -0.08129 respectively. Since we’re using an term in this equation, there the coefficient for inflation doesn’t represent the slope and doesn’t have a meaning per se. The coefficient being negative indicates downward concavity, which we can see in the scatterplot, albeit somewhat slightly.

### Evaluating Model Significance

Now we’re going to do an overall F-test followed by individual T-tests on this model to see if the model is meaningful and that the variables have a statistically significant relationship. The null hypothesis is that the variables are not statistically significant (the coefficients are 0), and the alternative hypothesis is that at least one of the variables is statistically significant (coefficient not 0):

Looking at the R output above, and evaluating at a 0.05 significance level, the p-value for the model which is far below 0.05, so we will accept the alternative hypothesis that at least one of the variables is statistically significant.

Then to do the individual T-tests at the same significance level, we can see that the p-values for inflation and are and respectively, which are also definitely below 0.05. So both of these terms are statistically relevant to the model.

### Making Predictions Using Model

Based on the model above, to find the predicted wage growth if the inflation rate is 7.41, we can just substitute in that value:

This evaluates to roughly 8.559.

Next we found the prediction and confidence intervals for this model at 95%:

Table

Description automatically generated

The prediction interval is [6.4568, 10.6616] which means that 95% of the time, if inflation is at 7.41%, our data point for wage growth will fall within that interval. The confidence interval is [8.2473, 8.871], which means that 95% of the time, the average wage growth for a group of data with an inflation rate of 7.1 will fall within this interval.

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

### Reporting Results

The general form of the second order regression model with wage growth as the response variable and GDP and inflation as the predictor variables would be:

Here we can have be inflation and can be GDP. Because this is a second order model, it contains both response variables, the interaction term, and the second order terms (the squared values).

To create the equation for our scenario we’ll run the model first:

Table

Description automatically generated

So with our specific information for this model based on the R output above, our model would be:

The R-squared and Adjusted R-squared values for this model are 0.9113 and 0.9065 respectively, which means that around 91% of the variation in wage growth in a model with these specific parameters can be explained by these parameters. That’s a pretty high R-squared value, and I’ve read that sometimes an R-squared value this high can indicate overfitting, but it may just be a very good model.

The beta estimates for GDP2and inflation2­ are 0.003176 and 0.027371 respectively. They’re both negative, so that means the plot will have a downward concavity, meaning the relationship will have a positive correlation, it will peak, and flatten out or begin to fall after that.

### Evaluating Model Significance

To evaluate the model significance we’ll do an overall F-test. The null hypothesis will be that the variables are not statistically significant in the model (coefficients of 0) and the alternative hypothesis is that they are statistically significant (coefficients not 0).

From the output above we can see that the P-value is (same as the previous model, interestingly) which is going to be way below the 0.05 significance level, so we will accept the alternative hypothesis stating that at least one of the variables is relevant to the model.

In order to determine which specific terms are relevant, we’ll do individual T-tests at the same significance level. The null hypothesis is that the coefficient for the term is 0, thereby making it irrelevant to the model. The alternative hypothesis is that the coefficient is not 0, making it influential to the model.

The p-values are as follows: Inflation: , GDP: , Inflation­2: 0.1706,

GDP2: 0.5837, interaction term: 0.0563. At a 5% significance level, only Inflation and GDP are significant to the model. The squared terms and the interaction term are not.

### Making Predictions Using Model

To find the predicted wage growth if inflation is at 7.41% and GDP growth is at 9.59%, we can substitute those values into our equation, where is inflation and is GDP:

That just wrapped around to a second line because it’s so long but that is the single equation. That evaluates to about 8.458.

Next we’ll run the prediction and confidence intervals for this model:

Table

Description automatically generated

Here the 95% prediction interval is [6.744, 10.1718]. That means that 95% of the time, the value for wage growth will be within that interval if the model has these exact parameters. The confidence interval at 95% is [1.1751, 8.7407] which means that 95% of the time, the average wage growth for a group of data with these specific parameters will fall within that interval.

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

### Reporting Results

Now we’re going to do a model for wage growth with inflation and economy as predictor variables. The general form for that would be:

We’ll say that is inflation and is economy. Because this is a second order equation, it needs to have both predictor variables, the interaction terms, and the squared terms.

In order to write the specific equation we need to run the model to get the coefficients:

Table

Description automatically generated

Now we can construct the model for our scenario:

The R-squared and Adjusted R-squared values are 0.8738 and 0.867 respectively, which means that about 87% of the variation in wage growth can be explained by a model with these specific parameters.

### Evaluating Model Significance

To evaluate model significance we’ll do an overall F-test. As before, the null hypothesis will be that the coefficients for all variables are 0 and that no variables are statistically relevant to the model. The alternative hypothesis is that at least one has a non-0 coefficient and is relevant.

We can see the p-value in the above output, which is , which is well below the significance level of 0.05, so we can accept the alternative hypothesis that at least one of these variables is statistically significant to the model.

In order to determine which ones are significant, we can do individual T-tests. The p-values are as follows: inflation: 5.45-8, economy: 0.10773, Inflation2: 0.00167, inflation:economy interaction term : 0.39118, and economy:inflation2 : 0.59197. Based on this, the only terms with a p-value below the significance level of 0.05 are Inflation and Inflation2, so these are the only variables that are statistically significant in the model.

### Making Predictions Using Model

Next we’re going to find the predicted wage growth if inflation is at 7.41 and the economy is not in recession. Because economy is a categorial variable it will use a dummy variable. It will be 0 for not in recession and 1 for in recession.

This evaluates to roughly 8.438. So that would be the expected level of wage growth if the inflation rate is 7.41% and the economy is not in recession.

Then we’ll look at the 95% prediction and confidence intervals for this model:

A screenshot of a conversation

Description automatically generated with medium confidence

So the prediction interval is [6.3961, 10.4796] which means that 95% of the time, the value for wage growth if the model has these specific parameters will fall within that interval. The confidence interval is [8.1175, 88.7582], which means that 95% of the time the average wage growth for a group of data with these model parameters will fall within that interval. The prediction interval is wider than the confidence interval because it takes into account sampling uncertainty and potential variation in the individual data point.

## **6. Conclusion**

I would recommend this model for situations where you’re trying to analyze variables with a non-linear relationship to one another. I’m sure there are more nuanced models than this, but as far as what we’ve seen up to this point, obviously this makes the most sense for non-linear relationships. We had some irrelevant terms in the model but you can always adjust the model. I’m also honestly not sure if the fact that a couple of the terms were not statistically meaningful to the model really matters if having a higher order model means that you need to have all those terms (the interaction terms and squared terms) present in the equation regardless.

This model showed us that we have a really strong correlation between wage growth and the predictor variables of GDP, inflation rate, and recession status (although I think you could argue that recession status is probably more of a sliding scale than just yes or no). The very high R-squared values in these models show this strong correlation, and the individual predictor variables all were statistically relevant to the model upon doing individual T-tests.

The practical importance of this is obviously that economies are complicated, with many interacting factors that would influence something like wage growth. Also clearly these aren’t all going to be linear relationships. So having an appropriate model type to account for the non-linear nature of some of these relationships while also allowing for the interaction between various terms is going to allow economists to model scenarios with a lot more specificity, which is critical all the time but especially during periods of economic uncertainty, because as James Carville said, “It’s the economy, stupid.”