Analysis of industry tax rate, compensation and dividend

David Wang, wangs10@rpi.edu

[1]Rensselaer Polytechnic Institute, Troy, NY, United States

Purpose

This poster attempts to identify and model relationships between tax paid per industry, compensation per industry, dividend paid per industry and Undistributed Corporate Profits per Industry. This data-set is provided by the Bureau of economic research(BEA), QR code to the data-set is available at the bottom of the poster.

1) The first relationship we wish to examine and model is between tax paid per industry and employee compensation.

2) With the second model we want too see if we can predict dividend paid, using relationships between tax paid per industry, compensation per industry, and Undistributed Corporate Profits per Industry.

Data Description

The data-set provided by BEA spans across 60 unique industries and provides data from 1929 to the present. Before the data-set can be used we require some data wrangling.

As we move from decade to decade new industries pop up and and old industry gets merged with others. For example from 1929 to 1948 the coal mining industry was split into two separate categories Anthracite mining and Bituminous / soft coal mining, after 1948 the two industry was merged into one, thus we first need to standardize the industries across the 89 years available to us.

Second the fiscal data provided does not account for inflation, thus if we want to evaluate the fiscal data on the same scale, we would need to manually adjust for inflation. To adjust for inflation we can use the Consumer Price Index (CPI) provided in the data-set, CPI is a measurement that examines the weighted average of prices of a basket of consumer goods and services (e.g transportation, food, medical care). We can use the following formula to calculate the inflation multiplier ((B – A)/A) + 1, where A is the starting CPI and B is the ending CPI. Using the above formula we can construct a matrix of inflation multiplier. Multiplying our data with the inflation matrix, we now have a inflation adjusted data-set.

Methods

1) With the first model we can employ a simple linear model to examine the relationship between tax paid per industry and employee compensation

2) The second model we can utilize a regression tree with 10 fold cross validation. A regression tree allows us to easily visualize the relationship between all the variables, and identify the importance of each variable.

Conclusions

- The results suggest that there is a strong inverse relationship between industry tax rate and average industry wages.
- There exists a small set of outlier industries where this relationship breaks down
- The regression tree suggest that out of the three variables used, undistributed profits played the biggest role in the predicting dividends paid out, this result was expected.
- The poor performance of the regression tree suggest that there are missing variables not captured by the model.

Results

Tax rate data

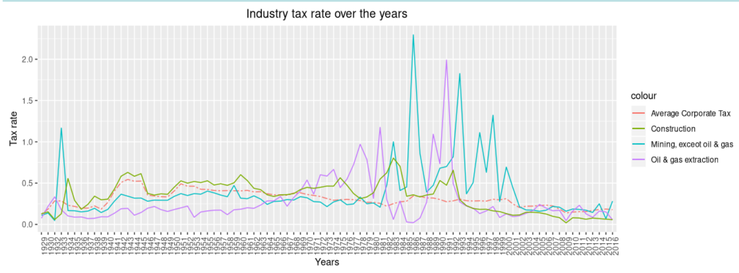

Figure 1. Tax rate of various industry from 1929 to 2016

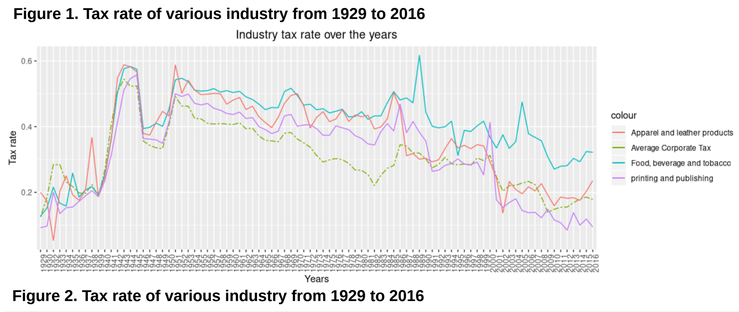

Figure 2. Tax rate of various industry from 1929 to 2016

The graphs on the left shows the tax rate for a selected group of industries, the dotted line indicates the average tax rate across all industries.

Something to note is that since the turn of the century the highest level tax rates has remained consistently low at around 35%.

The spikes in tax rates can be explain by a couple things, for example during the 1930 depression era, top level taxes saw a drastic increase from 25% to 63%. There was also the Revenue Act of 1936, which imposed an additional taxes of up to 15% on undistributed profits.

In the 1970's and 80's we saw the energy crisis and oil glut, where prices of oil saw a drastic rise and dip.

Employee wage data

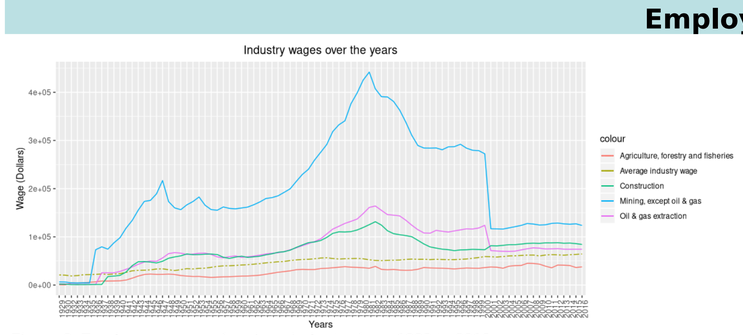

Figure 3. Employee wages of various industry from 1929 to 2016

Figure 4. Employee wages of various industry from 1929 to 2016

On the left we show the inflation adjusted employee wages for a selected set of industries, the dotted line again indicates the average tax rate across all industries.

From the end of WW2 we see a steady increase of wages through out the 60's, 70's, 80's and peaking around the late 90's, with the turn of the century we see a drastic decline in purchasing power.

At first this seems like an error in calculation, but after further research this does appear to be the case, In fact, hourly wages peaked during the late 60's and early 70's. In 1973 the recorded $4.03 per hour average wage has the same purchasing power of $23.68 today.

The QR code on the right takes you, you to a pew research article that further describes the trend

Linear regression of tax rate and Employee wages

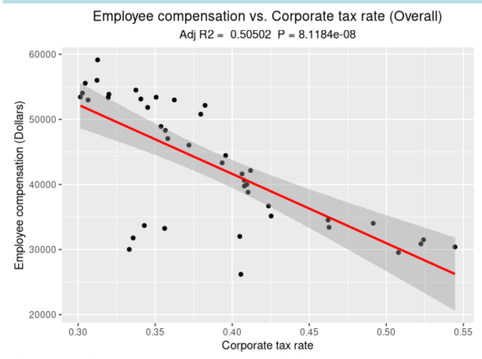

Figure 5. Linear regression for overall industry tax rate and wages

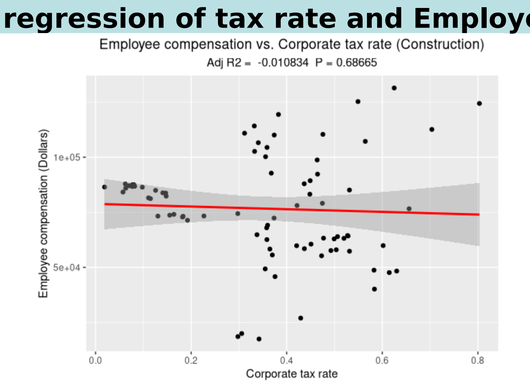

Figure 6. Linear regression for construction industry tax rate and wages

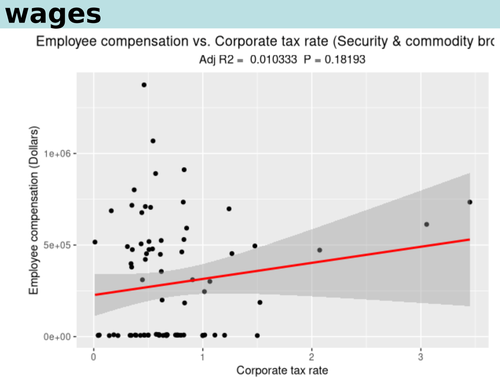

Figure 7. Linear regression for Security & commodity industry tax rate and wages

The leftmost graph shows a scatter plot of the average tax rates and their corresponding wages, besides a couple outliers, there does appear to be a strong linear relationship between the two factors, this linear regression model has a small p-value of 8.11e-8 suggesting strong evidence to reject the null hypothesis.

The strength of the linear relationship described above vary across the industries, but there are some outlier industries where this linear relationship breaks down. For example the middle graph and right graph shows the construction industry and security & commodity industry, for these two industries it would seem that wages are not affect as tax rates varies. The average wage for the construction industry hovers around the 70 to 75k mark as tax rates increase and decrease. The security and commodity industry displays consistent low tax rates and high variance in wages, this can possible be explained by being a commission based industry and tax advantages such as carried-interest, where profits can be taxed as capital gains instead.

Regression Tree

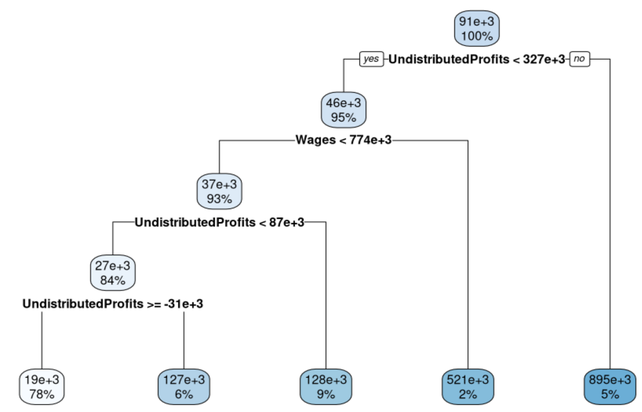

Figure 8. Regression tree to predict industry dividends

Using tax paid per industry, compensation per industry, and Undistributed Corporate Profits per Industry as decision variables we construct a regression tree with 10-fold validation with the objective to predict dividend paid per industry.

From the generated tree to the right, we see that the biggest decision factor is undistributed corporate profits, with tax paid per industry playing no role in the decision tree.

With the data split into 70/30 training and testing data, the model returns a 30.82% accuracy level, which would suggest that the factor used are not good predictors for dividend paid per industry

Glossary:

BEA – Bureau of economic research

Regression Trees -When the decision tree has a continuous target variable.

Bituminous coal - Bituminous coal or black coal is a relatively soft coal containing a tar like substance called bitumen or asphalt.

Anthracite coal - often referred to as hard coal, has a submetallic luster. Contains the highest carbon content, and fewest impurities, and the highest energy density of all types of coal.

Resources:

Regression trees: https://rpubs.com/cyobero/regression-tree

Linear regression plot: https://community.rstudio.com/t/insert-regression-model-into-ggplot2/2439/9

Data-set QR code: