Spring 2020

The Effect of Changes in Taxation on Earned Income

Kyle John Gourley
University of New Hampshire

Paul College Thesis Proposal

The Effect of Changes in Taxation on Earned Income

Kyle Gourley

Professor John Hasseldine

University of New Hampshire

Table of Contents

Introduction

For my thesis I will be focusing on the effects of changes in marginal tax rates of corporations and individuals on earned income. I aim to find relations between which changes to the tax structure drive increases in earned income as well as which changes reduce it. As different politicians propose new plans to the taxation people are often curious about how these changes may affect them, especially if it will affect their income. I would like to use historic taxation data to try to find how in the years of change earned income compared to what we expected earned income to be.

This question really struck me when reading about argument for universal basic income. If taxation doesn't affect earned income, then it may make sense to argue for a universal basic income paid for by taxing businesses since we already know they are becoming significantly more productive, yet wages aren't seeing the same rates of growth. Alternatively, if taxation does affect earned income increasing taxation of businesses could increase issues with earned income.

As we can automate more jobs there may come a time where the government may want to use taxation to help those that may have had their jobs displaced. If this is something that happens it'll be very important to ensure that they aren't harming income more than they are helping. This is just one reason to question how tax rates affect earned income, because earned income is the primary source of income for Americans there are countless reasons to try and answer this question ("Wages and Salaries", 2016).

Review of Literature

To better understand a data analysis on the effects of changes in tax rates it'll be useful to understand why changes may be occurring at certain points. The only major tax reform in recent memory occurred in 2017 and won't start effecting returns until the 2018 tax year. Since the data that will be used for this analysis isn't that current, this review of literature will focus on reforms that occurred further in the past. With a better historical knowledge, looking around the years in which change occurred and the effects can serve to hypothesize the effects.

Initially the US didn't even have an income tax and instead relied on excise taxes for income. In the 1790s a war with France led to the creation of a property tax which was imperfect. The war of 1812 was funded by higher excise taxes rather than property taxes. It wasn't until the civil war that income taxes began, the revenue tax of 1861 was levied on incomes greater than $800 and wasn't rescinded until 1872 (Beattie, 2010).

In 1913 the 16th amendment was passed which removed the proportion to population clause allowing for an income tax. Shortly after this amendment there was an income tax on those with an annual income over $3,000 which touched less than 1% of people. World war one experienced revenue acts that brought up tax rates and lowered exemption levels leading to a 5% increase in US citizens paying tax (Beattie, 2010).

Roosevelt's new deal and world war two is where my data that will be used for this analysis starts and is a time marked by extremely high tax rates. Due to a lack of previous data and taxes being much lower before it doesn't add value to the ability to draw meaningful conclusions. Additionally, being at war would make it difficult to separate correlation from causation. However, when Nixon became the president, they started to peel back taxes making this a good point to try and use data analytics to find what the changes may have caused. Also,

President Reagan brought more tax cutbacks which can be evaluated. Finally, in 1993 we see the first tax increase that I could analyze for causes and effects since all previous increases lack the data for comparison (Beattie, 2010).

One of the first changes to the tax system of interest to this analysis is the Revenue Act of 1935, this was an act that would raise taxes, especially on the wealthy. The change in income tax for lower income individuals making less than $50,000 and the first bracket above $50,000 went from 30% to 31%. Previously the highest tax bracket was $1,000,000 and the rate was 63% however after the new highest bracket was $10,000,000 and the rate was 75%. There were also changes to Inheritance, Estate, and Gift Taxes which are of less concern to me as income tax for individuals and businesses are more of focus as the question is the effect of tax rates rather that taxation in general. Previously the business tax was 13.25% on the first $15,000 in net income and 14.25% on the excess. Roosevelt wanted to make a progressive tax which would slightly decrease the taxes on businesses earning less than $40,000 and slightly increase taxes on higher earning businesses. The brackets ranged from a rate of 10% on business making $2,000 or less and went up to 17.5% on businesses making a million or more. It also may be worth going back and investigating the revenue acts of 1932 and 1934 however for focus on smaller changes in specific years will be more of a focus when the data analysis begins, for now the focus is on looking at larger changes to try to better understand the history of taxation. (Blakely, 1935)

The Revenue Act of 1935 raised taxes especially on the wealthy and followed previous increases in taxes on the wealthy. This occurred during the great depression, so tax revenue had become so low that the government needed money just to fulfill its basic function, so this change was more out of need to create cash flow than to promote any sort of political agenda. In the year following this change we see total tax liability sharply increase despite the total income being

minimal as well as salaries. The question we're moving towards is how reforms in tax effect the US economy. At this time, there were so many forces effecting the economy that it would be nearly impossible to make any valid connections to tax rates.

The Tax Reform of 1969 was the first decrease that occurred and conditions appeared to be more normal than the Revenue Act of 1935 making it more interesting to prepare for my analysis. This tax reform lowered the top marginal bracket for earned income from 70% to 50% so the reduced tax liability in the next year isn't surprising. Additionally, growth of salaries, wages and business income slowed in the couple of years following, it would be interesting to find evidence that suggests that this is related to the reform. It's also important to ensure there isn't evidence that suggest something else caused the change in growth rates, so I don't draw inaccurate conclusions (Weiss, 1970).

The tax reform act of 1986 was another reform to lower taxes this time by Reagan. This act lowered the top marginal rate from 70% to 50% and raised the bottom tax rate from 11% to 15%. This act also got rid of the distinction between long term capital gains rate and ordinary income raising the maximum rate from 20% to 28%. There was also an expansion of the alternative minimum tax and increased the mortgage interest deduction. For businesses, the tax rate was reduced from 50% to 35% but reduced allowances on business expenses. This change to business tax will likely be a major part of what I analyze as such a large change should create impacts large enough to make conclusions on effects of reduction to business taxes (Kagan, 2109).

In 1986 Ronald Reagan signed the Tax Reform Act of 1986 which slashed the top marginal bracket again down to 28%, however he also raised the lowest rate from 11% to 15% so some individuals would be paying less in tax and others more. As would be expected we can see

a much smaller increase in tax liability that previous years. Additionally, salaries and wages as well as total income had a strong year of growth. This result seems different than the tax reduction in taxes in 1969, however it is notable that business corporate tax rate was also reduced from 50% to 35% (Kagan, 2019).

The Tax Reform Act of 1993 by the Clinton administration was one of that increased taxes. There was the addition of the 36% tax bracket, increases to gasoline taxes and an additional tax of 10% on couples filing jointly with income above $250,000. There was also an increase to social security benefits and the removal of the tax cap on Medicare. This also is a reform that raised business tax rates and reduced deductions. This was also the first tax bill to retroactively raise taxes which makes it another valuable one to investigate the effects a tax increase has. (Kagan, 2019)

In 1993 Bill Clinton implemented a change in the top marginal bracket raising it from 31% to 39.6%. In this time, we see little change in growth rates of both tax liability and income which doesn't necessarily help my findings, things would be easiest when a change in the data occurred at the same time as a change in tax (Beattie, 2010).

As I'm getting into the data and comparing what was happening with tax reform and what income was looking like it's becoming apparent that it would be best to try to find one specific variable representative of earned income to be a dependent variable and use independent variables that represent every change that occurred from 1929-2017 to rates of taxation. Initially the thought was to analyze multiple variables from the big tax reforms however the more practical strategy seems to be to try to take all years within the available data and predict the effect of changing the independent variable on the dependent variable.

Research Question and Predictions

My research question is what the effect of changing tax rates of businesses and individuals is on earned income. My prediction is that as businesses pay more taxes in order to compensate for that lost income, they will be hiring fewer people and paying less. Additionally, if individuals are paying more in taxes, they are spending less which will hurt businesses and in turn lead to them paying out wages.

The goal of my research will be to create a model that can predict the effect of changing tax rates on earned income. Additionally, the prediction I made is very broad and refers to general increases or decreases to businesses and individuals not specific marginal rates. I would like to have more specific conclusions about which levels of marginal rates should be changed. I will be attempting to make connections between earned income and taxes rates for both corporations and individuals that could be used in any time period. If successful one would be able to take the variables that are included in the final model and predict what salaries and wages would be given the values of the variables.

Methodology

The data for this thesis will be secondary data collected through the IRS's website as well as some third parties such as the Tax Policy Center Urban Institute and Brookings Institute. The question I am trying to answer is regarding earned income. While earned income can come from many places such as salaries and wages, self-employed income, tips, and other sources for the purpose of this investigation I will be using salaries and wages. Salaries and Wages are a common type of earned income and has easily accessible data.

To analyze the data, it will be organized in excel. The data is coming from multiple sources so it will be important to have it organized in one place. Portions of the data can be imported as CSV files using a delimiter to ensure it imported into each cell separately. Unfortunately, some of the data is not available in an organized fashion already so basic data entry will be used to bring this data to excel. The benefit of putting this time in up front is that once organized into a single table in excel it will be much quicker to create different models. Using the analytical solver platform add on in excel the data table will be used to create regression models.

To predict how tax rates effect earned income data on historical tax rates as well as the ranges of income those rates apply to will be used. Linear regression is the strategy that will be used to try to understand the impact of these changing rates and ranges of income on salaries and wages. Linear regression takes a set of predictor variables which in this case will be the rates and ranges as well and examines if those predictors do a good job of predicting a dependent variable which in this case will be salaries and wages. Linear regression also assesses which variables have a significant impact on the outcome variable and the magnitude of impact on the dependent variable.

The aspect of significance will be important for the process of variable selection as my model will start with many variables and not all of them will necessarily be significant. Significance indicates whether a change in the independent variable has a correlation to changes in the dependent variable, Insignificant variables will be removed until the model is made up only of variables that are significant predictors of salaries and wages.

In order to use linear regression as a tool there is a need to create variables that remain consistent throughout every year. Because the number of rates and corresponding ranges of

income have changed multiple times over the years using all rates and ranges is not feasible for a linear regression model. Instead from each year the rates as well as ranges will be broken up into highest, lowest, median, mean, upper quartile and lower quartile for both corporate income rates as well as individual's income rates.

Because the rates are directly related to their corresponding income ranges for each of categories a new variable will be created by multiplying the rate by its corresponding range (i.e. highest rate x highest range, lowest rate x lowest range). Also, to address the issue that a range is not a single number the amount of income at which you start paying a new rate will be used. In order to condense the variable names, the level of income in which a rate will begin to apply will be referred to as the "tax base".

To choose what values remain in the model the p-value will be used to eliminate variables from the model. The "the p-value is the probability of obtaining results as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct." (Beers, 2020).

In this case the null hypothesis is that the predictor variable does not have an impact salaries and wages. We will reject this null hypothesis if the p-value is less than .05 in favor of the alternative hypothesis being that the predictor variable does have an impact on salaries and wages. A smaller p-value indicates stronger evidence in favor of the alternative hypothesis and .05 is a level commonly used in statistics as a point in which you can reject the null hypothesis. It is also important to note that variables will be removed one at a time as on each different model the p-values will change, and a variable may become significant (p-value below .05) when a different insignificant value is removed.

Once all variable are significant the regression report will be analyzed trying to find if the model is actually predicting how these changes to taxation will affect salaries and wages or if it is merely the equation of a line that comes closest to the data points. This will be done by partitioning the data into a training set and test set and seeing if the error rates significantly increase on the test set. If the error rates do significantly increase on the test set more variables will be removed to see if this I because of overfit.

Exploratory Analysis

The exploratory analysis will be an opportunity to visualize different parts of the dataset to become more familiar with the data. It is not meant to draw conclusions but to raise questions and understand what is happening with the data over time.

Figure 1:

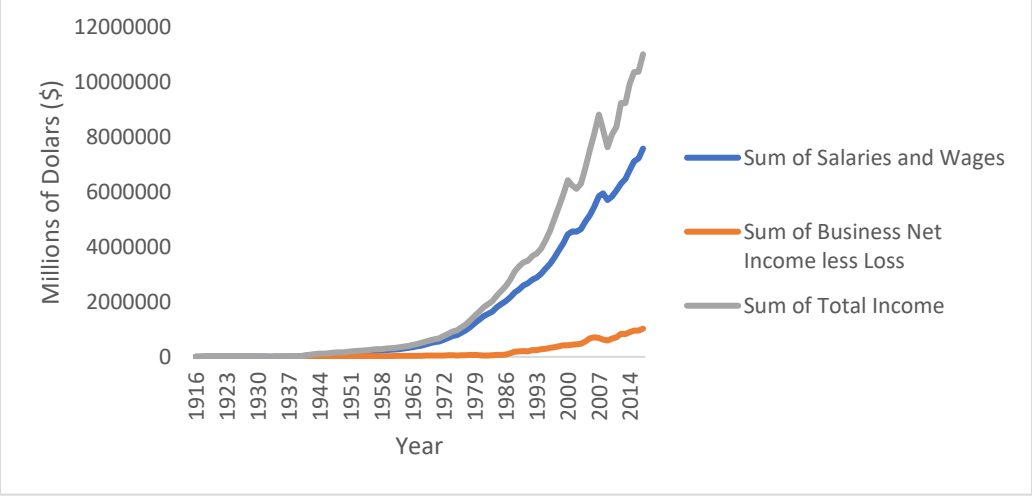


Figure 1 is a time series line graph shows the total of income from 1916-2014 including series of business income/loss as well as total income. This shows us the amount of changes businesses experience compared to salaries as wages when total income changes. The changes appear to be

proportional when looked at in the short term however the rate of growth of business income/loss is much more moderate than salaries and wages as well as total income.

Figure 2:

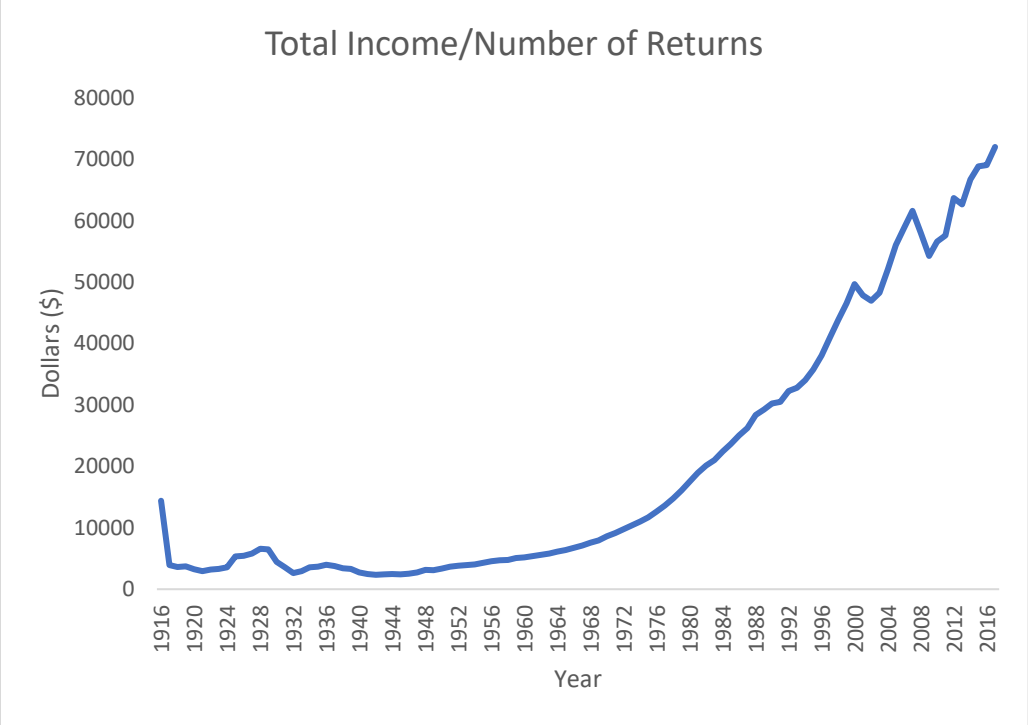


Figure 2 shows the total income divided by the number of returns. This shows the growth of income without being affected by the increase in number of returns. When compared to the figure 1 you see the same distinct spikes around 2000 and 2008 from total income which we now know was due to changes in income not number of returns.

Figure 3:

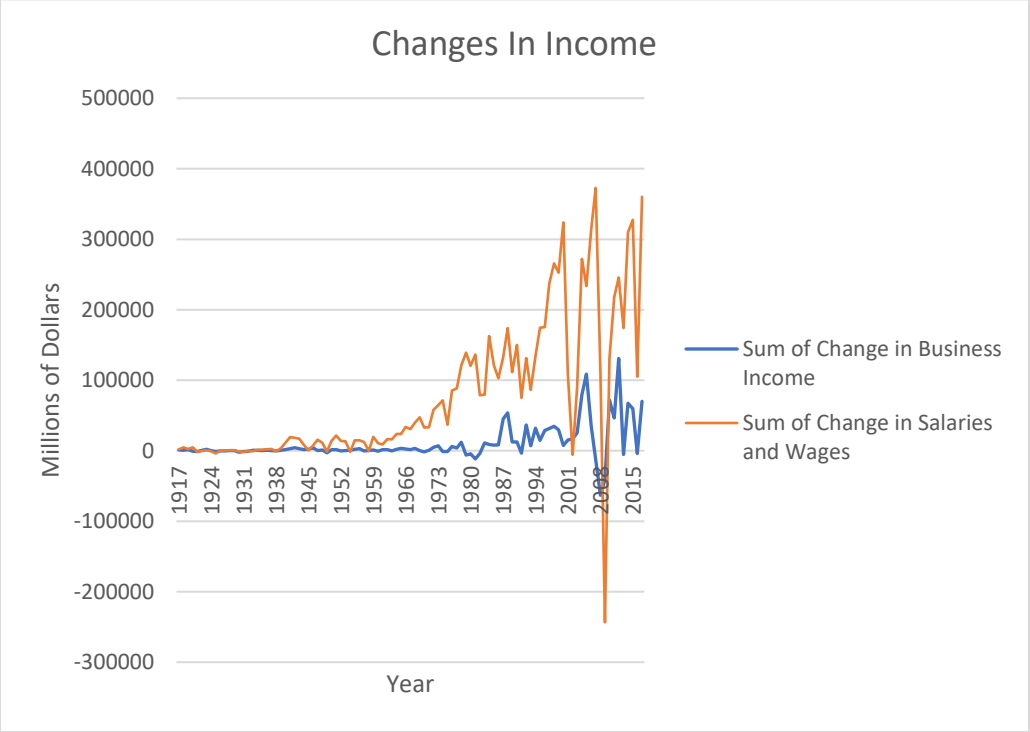


Figure 3 is similar to the first one except this only considers the change and does not include total income. Also, because it is looking at the change in income rather than the total amount of income it's easier to where changes occur. There is little new information however this gives a clearer depiction of what was concluded from the first two graphs.

Figure 4:

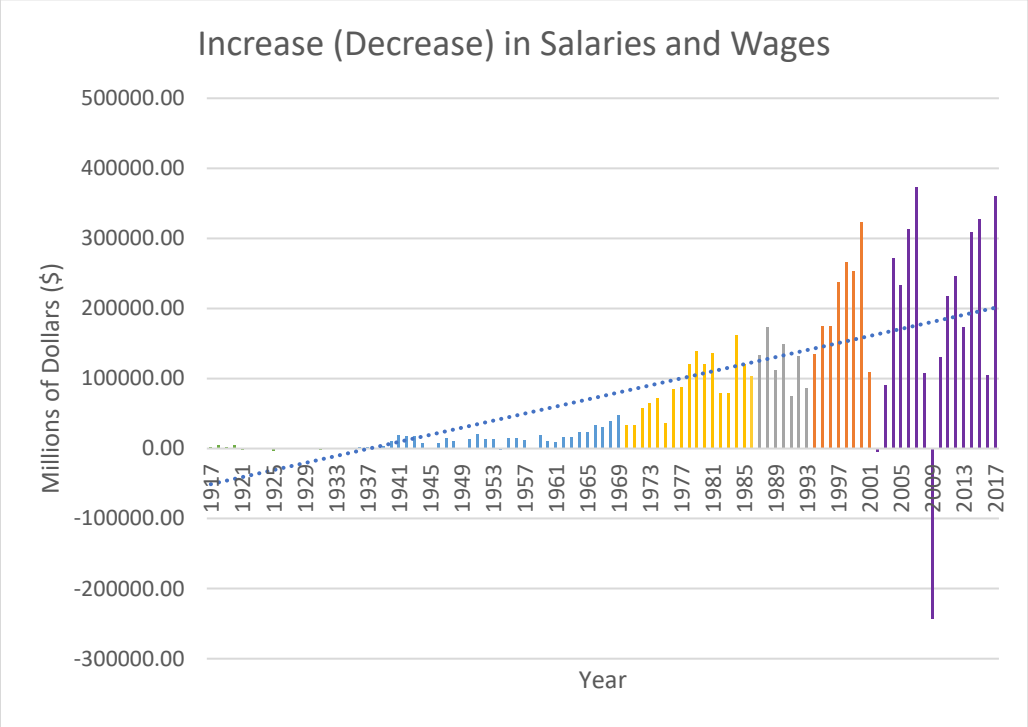


Figure 4 shows the increases and decreases of salaries and wages with colors changing in years of large changes to taxes. There do appear to be some sharp changes in rates on some of the years of change but not in a specific direction, so it is difficult to draw a conclusion.

Figure 5:

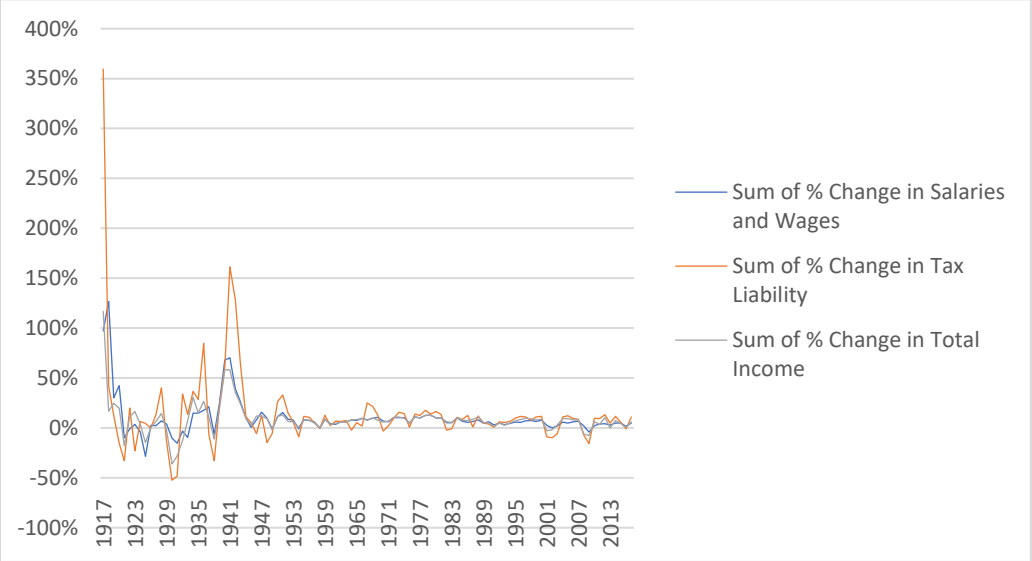


Figure 5 shows the total income as well as salaries and wages percent change compared to the tax liability changes. The tax liability experiences higher percent changes than salaries or total income in the same years.

Figure 6:

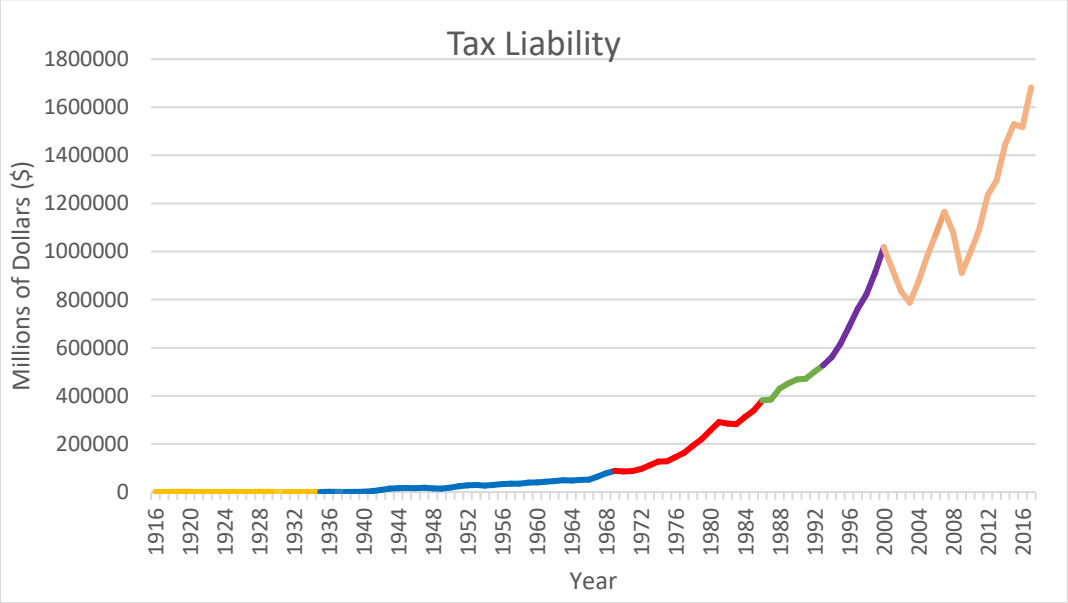


Figure 6 shows the tax liability with changes in colors in the years of change. It does not appear that most years of change had significant changes to the tax liability. 2001 is the only noticeable large change and that may have been from the dot com bubble more so than the tax change.

Figure 7:

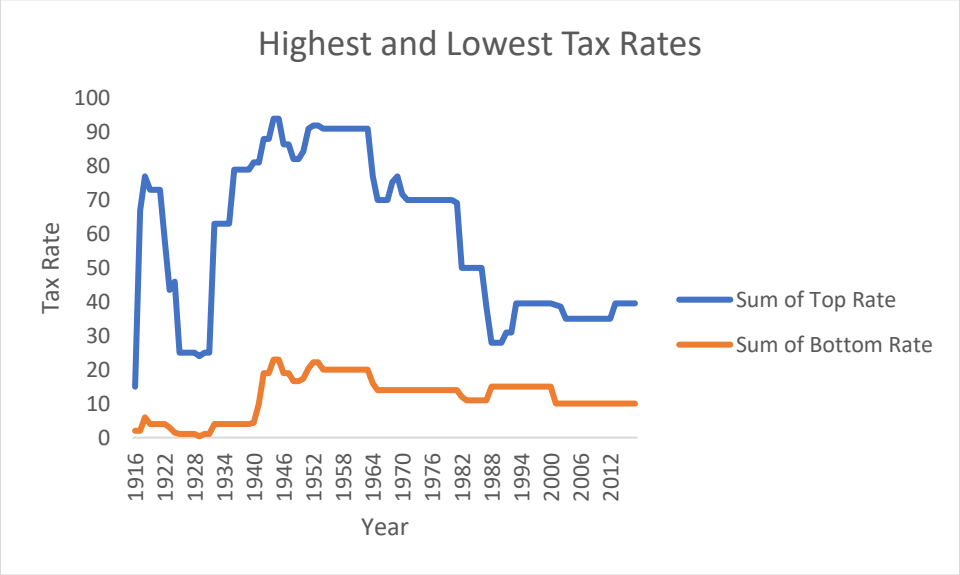


Figure 7 shows the lowest individual rate of taxation as well as the highest individual rate of taxation on income over time. There are visually noticeable similarities between the timing of increases and decreases of the rates especially in earlier years however the highest rate has a much larger range. In later years there is less consistency of timing of increases and decreases as the highest tax rate goes from extreme highs to rates much closer to the lowest rate.

Figure 8:

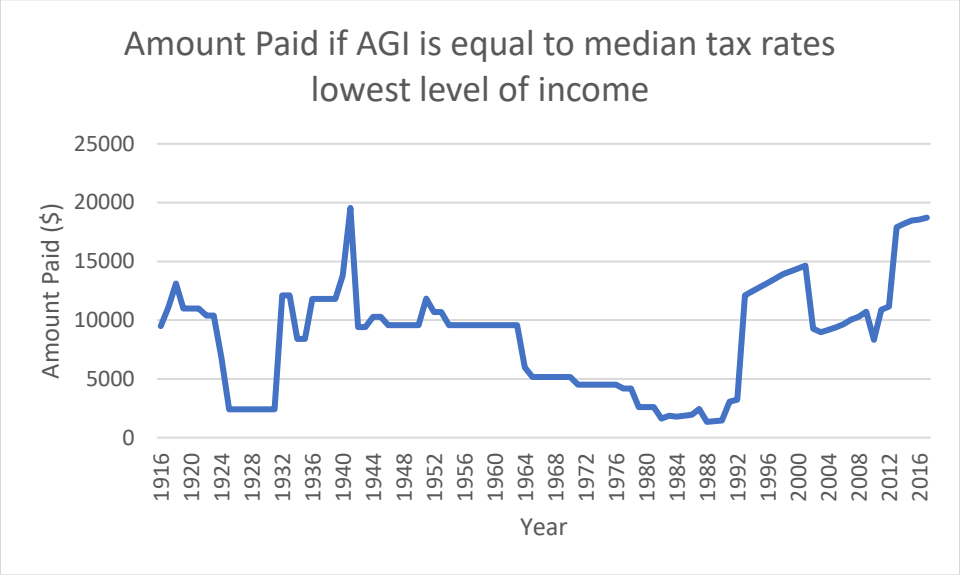


Figure 8 depicts the amount a taxpayer would pay if they reached the level of income in which the next additional dollar would be taxed at the median rate. For much of time this amount would be around $10,000 however there is one spike at 1941 in which this amount reached just shy of $20,000. There are also longer sustained periods of time where this amount was below $5,000

Figure 9:

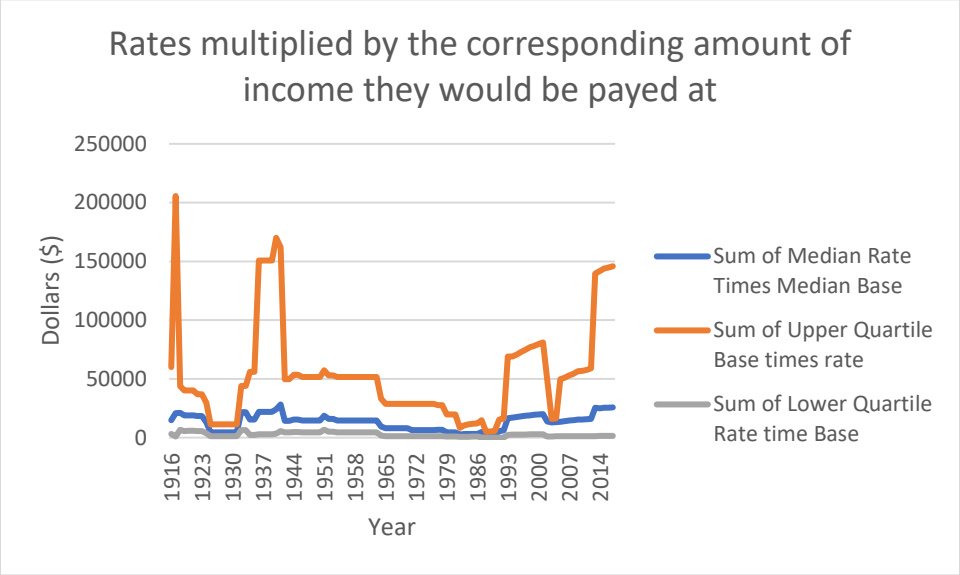


Figure 9 shows the median rate times median base and the upper and lower quartile rates multiplied by the upper and lower quartile bases. In this we can see there is much larger variations in the upper quartile than the median or lower quartile which is likely because in different years the marginal rates enter significantly higher levels of income.

Figure 10:

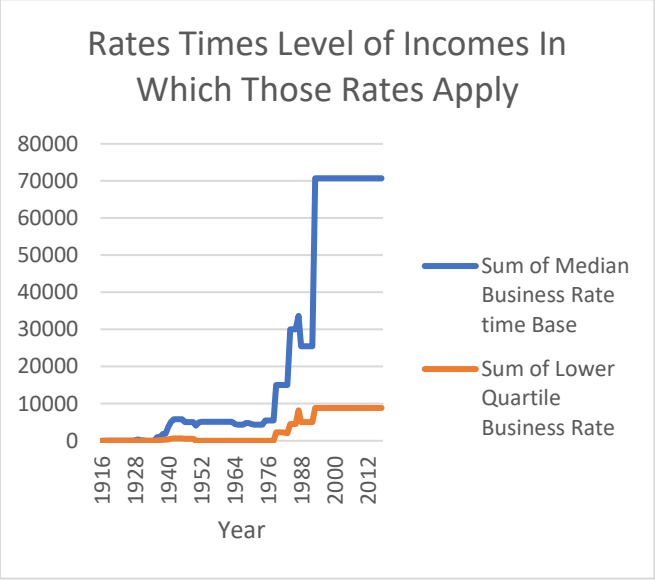
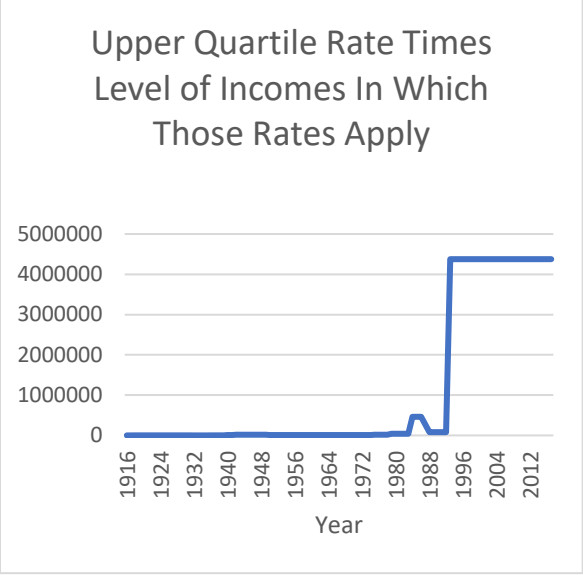


Figure 11:



In figure 10 and figure 11 we can see the median, upper quartile, and lower quartile tax rates multiplied by the median, upper quartile, and lower quartile lower ends of income ranges which we can see all experience sharp increases at the end of the 70s indicating that the amount of taxes being paid are likely increasing.

Regression

The first step in creating the regression model to try to predict what effects salaries and wages is to find out which variables are significant impact on the model. The model will start with 23 variables for the first regression, it's expected that not all variables will be significant so variables will be removed one at a time until all are significant to the model. The variable being removed each time will be the variable with the highest p-value suggesting that it is the one with

the least impact on significance. The reason only one variable will be removed at a time is that new regression model has a whole new set of p-values.

Only variable with a p-value below .05 will be considered significant enough to reject the null hypothesis that the predictor variable does not have a significant impact on our output variable, salaries and wages. The alternative hypothesis is that the predictor variable does have a significant impact on salaries and wages.

The variables being used in the initial regression will be made up of business and individual tax rates and the lowest level of income that those rates apply, for the sake of simplicity those levels of income will be referred to as the tax bases. From the individual's taxes the variable will be highest, lowest, mean, median, upper quartile, lower quartile rate, highest rate times highest base, median rate times median base, mean rate times mean tax base, upper quartile rate times base, lower quartile rate time base, amount paid if adjusted gross income is equal to the median tax base, and amount paid if adjusted gross income was equal to the highest tax base. For Corporate taxes the variables are highest corporate rate, lowest corporate rate, business median rate, business mean rate, business upper quartile rate, business lower quartile rate, median rate times median base, mean rate times mean base, lower quartile rate times lower quartile base, and upper quartile rate times upper quartile base. The reason all the bases were only included with their corresponding bases is because they are there to provide additional context to the rates as each rate only applies to specific levels of income.

Because the research question is trying to find out how changes to taxation affect earned income it is important to not address the variables that aren't affecting earned income. The first variable to be removed was the corporate upper quartile rate, notably the variable corporate upper quartile rate times upper quartile base is one of the more significant variables in this model

suggesting that the rate matters when given the context of what income range it applies to. The next variable removed is the individual lower quartile tax rate times lower quartile base. After this variable, the individual lower quartile rate was removed which would be expected considering this number times the base was insignificant to this model. The corporate median tax rate is the next variable deemed insignificant however just like the upper quartile rate, when multiplied by its corresponding base it becomes very significant, in this case the most significant in the model. The next variable being removed is the median individual tax rate times the median base which continues to suggest that taxation of individuals, specifically those with lower to middle levels of income. The next model only has only one insignificant variable which was the bottom rate which for the same reasons as the bottom rate times bottom base makes sense as an insignificant variable.

Now that all p-values above .05 have been removed it's important that we ensure that the model is not overfit. Overfitting a model is when the model begins to describe random error in the model rather than actual relationships between variables.

Figure 12:

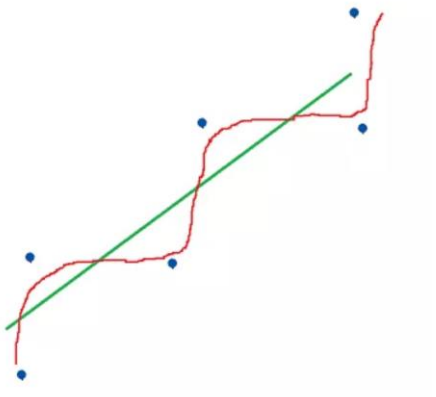


Figure 12 can show how overfit can happen, the green line is the actual relationship to the variables and the red line is that produced from a regression. The regression line is being pulled

away from the actual relationship on account of random error rather (Frost, 2004). This can occur because the model is to complex which in this case would mean to many variables. In order to avoid this the data will be partitioned so that 80% of the data is used to train the model and the other 20% will serve as a test set to see how well the model predicts data that wasn't included in training the model. To see if overfit occurred I will look to see how the Root Mean Squared Error (RMSE) changes between the training set and test set. If there isn't an issue with overfit, they should remain similar and if there is overfit, we will see an increase in RMSE in the test set.

When the data is partitioned, and we use only the variables that had a p-value below .05 there end up being a 36.36% increase in the RMSE from the training dataset compared to the test set indicating that there is some overfit occurring. Previously we had been only removing variables that were insignificant however now that all variables are significant, I will remove variables that are represented twice to try and reduce overfit. The individual upper quartile rate, mean rate, highest rate and the business mean rate and lower quartile rate will all be removed as those rates times their bases are also in the model. For that reason, they provide little value for drawing further conclusions and their removal should help with overfit.

After removing the variables that were redundant, the RMSE increased 26.98% between the training set which was an improvement but suggests there is still overfit in the model. Since related variables were removed, the regression was run again without partitioning the data and some variable's p-values had become above .05 so the next variable being removed will be the least significant. In this case was the amount and individual taxpayer would pay if their income was equal to the highest tax base. In this regression model there were still insignificant variables remaining so now the highest corporate rate will also be removed. Continuing with the removal

of insignificant variables from the model the individual upper quartile rate times the corresponding base will be removed. This model still has one insignificant variable remaining which is the amount an individual would need to pay if their AGI was equal to the median tax base.

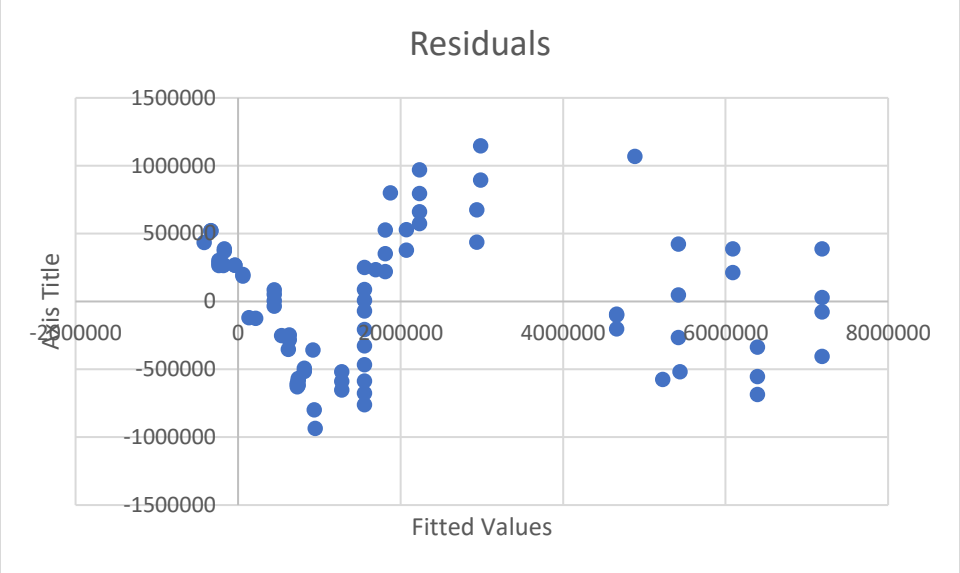
In the latest model if the data is partitioned the increase in RMSE from the training set to the test set it 23.2%. While this number is improving there is still room for improvement and this model uses the variable for the lowest corporate rate which has a p-value of .049594. While that number is within the threshold that had been set it is so close that it would make sense to remove this variable as well. The removal of this variable led to the RMSE only increasing 17.16% Which is a very large improvement from 36.36% which was the increase on the model we initially checked for overfit. While removing variables to reduce overfit some variable's p-value increased to above .05 so highest individual rate times the corresponding base as well as mean individual rate times its corresponding base were removed. The removal of these variables further reduced overfit. When the data was partitioned the RMSE only increased by 12.94%.

This model was made up of the variables median individual tax rate, mean corporate rate times its corresponding base, median corporate tax rate times its corresponding base, upper quartile corporate tax rate times base, and lower quartile corporate tax rate times base. These will serve as the variables we will use to try to use to interpret the meaning of the regression however they may need to change if the model does not appear to be an effective predictor of salaries and wages.

With the variables for the model chosen the next step is to evaluate the model. Before interpreting the equation and what it means the quality of the model will be evaluated. The first metric we will assess is the R squared value which is how close the data is fitted to the regression

line. This is a value between 0% and 100% and in the case of this model the value is 95.39% which is a very high value. A high value is not always an indication that the model is a good fit though, we also must check a plot of the residuals to see if they are random. If the residuals are not random it suggests that variables are not actually predicting salaries and wages.

Figure 13:



While many of the residuals in figure 13 appear to be random its notable that the lower predicted values have a distinct downward trend which bring the validity of this model into question. Looking at the raw data alone its notable that many of the variables being used have 0 as value in early years which is a possible explanation for this pattern in the residuals. For that reason, the data for regression was shortened to go from 1957 until 2017 where the variables will have fewer 0 values.

Now because we are not using the same range of data the significance of variable will change causing us to need to make changes to the variables we are using. Using the same strategy of eliminating the highest P-values and then removing redundant variables we were able to narrow the variables to lower quartile rate times base, mean rate times mean base, median rate times

median base, highest rate times highest base and amount paid if AGI is equal to the maximum tax base for individual rates. For corporate rates it was narrowed to median rate times median base and highest rate times highest base.

Figure 14:

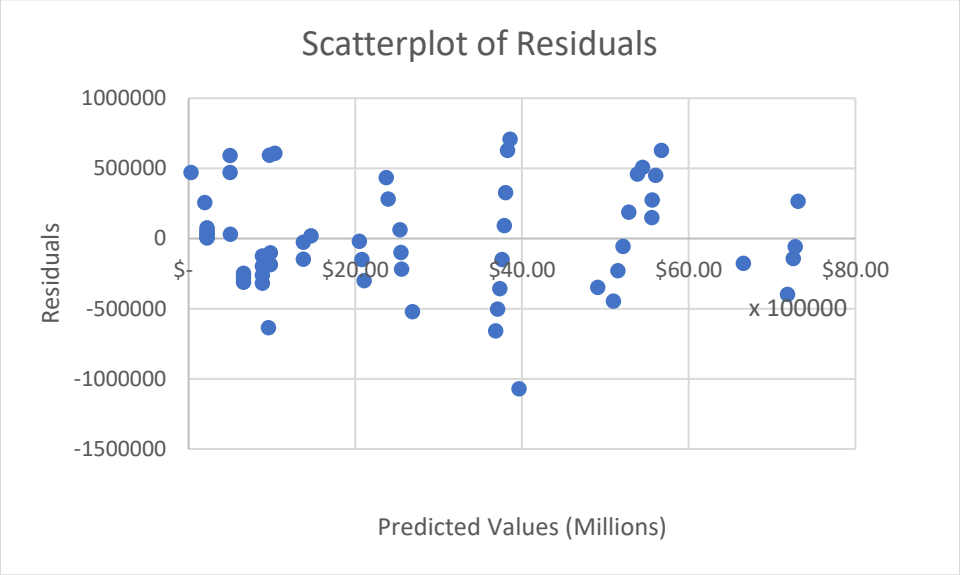


Figure 14 is a scatterplot of the residuals at each predicted value for salaries and wages with the new variables and range of data used which we can see is much more random than the previous scatterplot created using the original variables and dataset. The R squared value of this model was 97.4% which is even higher than in the original model. Because this R squared value is higher, and the residuals appear to be random this seems to be a much stronger model. Also, when the data was split into 80% being used to train the data and 20% to test the data the RMSE only went up by 1.32% which suggest there is no issue of overfit in this data.

The next metric we will explore is the Root Mean Squared Error (RMSE), which in this model is 364,352.7. RMSE is the standard deviation of the residuals, which essentially tells you how far the residuals are spread from the line of best fit (Ducasse, 2019). While 364,352.7 seems like a very large number it is important to consider our dependent variable of salaries and wages.

From the data used to create this model the mean of salaries and wages was $3,791,832.46 and reached a peak of $11,009,900 in 2017 and the trend of this data suggests that this number will only continue to increase (These numbers are all in millions of current dollars). While 364,352.7 is a large number when you consider the size of the dependent variable it isn't am excessive error. Additionally, a limitation of this model is that taxes are not the only thing that may affect salaries and wages so it is not expected that this would be a perfectly accurate model. The goal is to make the model accurate enough to try and draw conclusions about how taxes effect salaries and wages (Moody, 2019).

Now that it has be established that the model is reasonably related to salaries and wages the focus now turns to what the regression model tells us. A linear regression model creates an equation that when values of variable are inputted it predicts the dependent variable. The equation that was created from this model is as follows:

Salaries and Wages = 602359.257705512 + (-931.163597071875 * Lower Quartile Rate time Base) + (583.646949949997 * Mean Individual Rate Times Base) + (-564.047586743709 * Median Individual Rate Times Base) + (85.9666252381008 * Amount Paid if AGI is Equal to Highest Tax Base) + (-95.5987934413537 * Highest Individual Rate Times Highest Base) + (87.7057010925098 * Median Corporate Rate Times Median Base) + (-0.632294694193492 * Highest Corporate Rate Times Highest Base)

Figure 15:

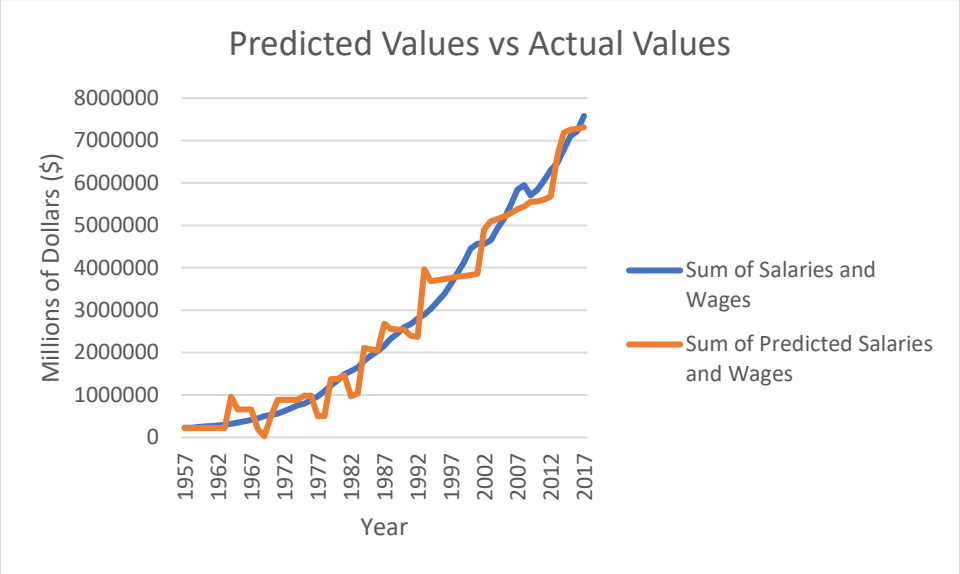


Figure 15 is a time series that shows the line that would be created when the variables are inputted for each year compared to the actual values for salaries and wages.

Analysis

Using the equation created by the regression the impact on salaries and wages of changing each independent variable by one can be predicted. Also, a reminder that all values for salaries and wages are millions of dollars and the effect each variable has will be in millions of dollars.

First in this equation we see a number with no variable connected to it. This is our intercept, the value we would expect salaries and wages would be if all variables are zero, which in this case is $602,359.26. If the lower quartile individual rate times the lower quartile base is increased by one salaries and wages will decrease by $931.16. If the mean individual rate times mean base is increased by one salaries and wages will increase by $583.65. However, increasing the median individual rate times median base by one causes a decrease of $564.05. For each 1$

increase to the amount paid if adjusted gross income is equal to the highest tax base salaries and wages will increase by $85.97. For the highest individual rate times the highest base each one increase causes a $95.50 decrease to salaries and wages. This comprises all the effects of changing individual rates in this model. For corporate rates each increase by one to the median rate times the median base leads to an increase of $87.71 to salaries and wages and each one increase to the maximum rate times the maximum base reduces salaries and wages by $0.63.

While the equation suggests we can predict and level of salaries and wages by inputting each of these variables there are aspects that limit accuracy. One limitation to this regression model is that it is only considering tax rates and some income levels that rates apply to. Even within the variable that we are using there are limitations such as the fact that deductions, credits, and what is considered taxable income is not considered. Additionally, you can say with near certainty that taxes are not the only things that effect salaries and wages, there are countless other economic and social impacts that effect earned income, specifically salaries and wages.

Reducing variables with negative coefficients to zero and maximizing any variables with positive coefficients will result in extreme increases to predictions of salaries and wages which I don't believe would occur. Rather than use the model to predict massive increases it would be more useful to compare the effect of different small changes. This model suggests what variables should be increased or decreased in order to try and have more realistic impacts on salaries and wages.

The first two variables to focus on are mean and median individual tax rates multiplied by their corresponding bases. The impact of increasing these variables by one is close except that raising the mean increases salaries and wages while raising the median decreases salaries and wages. This suggests that to increase salaries and wages the higher half of marginal rates and the

levels of income which they apply to should be increased as this will increase the mean without effecting the median. While raising the mean helps increase salaries and wages it's important to consider that the highest rate times highest base decreases salaries and wages. Based on this it would be most beneficial to salaries and wages to increase rates above the median but not necessarily the highest rates. This same idea is reinforced by the impact of the lower quartile rate times lower quartile base, which is part of the lower half of marginal tax rates and bases. When the lower quartile rate times base is increased, decreases salaries and wages occur which is why the lower half of variables don't seem like they should be increased to raise the mean. The last variable for individual rates is the amount a taxpayer would pay if their AGI was equal to the highest tax base. While this variable spans across both the top and bottom half of tax rates it is meant to represent the effect of changing tax rates for higher income taxpayers whose income reached the top half. For that reason, it fits the theme of all other variables based on the taxation of individuals that raising the higher half of marginal rates and bases helps increase salaries and wages.

In this model, two corporate rate variables were considered, Median Rate times Median Base and Highest Rate Times Highest Base. The Median Rate Times Median Base variable indicates that each one increase to this variable leads to a $87.71 increase in salaries and wages. This is interesting because it suggests that businesses affected by this rate do not try to offset a higher tax expense by lowering the salaries and wages expense. It is also notable that a business does not have to have very large profits to reach the median tax base as from 1957-2017 the mean of all the median corporate tax bases is $104,713.

The other variable relating to corporate taxes was highest rate times highest base, which when increased leads to a decrease in salaries and wages of $0.63. While this suggests that

higher taxes leads lower salaries and wages it is the variable with the lowest coefficient meaning there needs to be very large changes to this variable for it to have a strong negative or positive effect on salaries and wages.

Conclusions

The goal of this data analysis has been to determine what, if any, the effect of tax rates was on earned income. Because of the data available the dependent became narrowed to salaries and wages. While this does exclude bonuses, tips, self-employed income, or any other types of income for work done that doesn't fit the previous categories it made the most sense as the metric to look at for earned income (Kagan, 2020). In 2014 Salaries and Wages made up 77.6 percent of U.S. household income before taxes which includes earned and unearned income so salaries and wages are the largest source of earned income ("Wages and Salaries", 2019)

Using data from the Tax Policy Center on salaries and wages in conjunction with information on marginal tax rates from the internal revenue service a multiple linear regression model was created. This showed that there was a significant relationship between marginal tax rates and salaries and wages.

While the model did show there was a relationship there were limitations. This model only considered marginal tax rates when attempting to predict salaries and wages. This is useful for comparing the different variables relating to marginal tax rates, however it does not necessarily show the actual dollar amount that salaries and wages would change given a change to our variables. If marginal tax rates were a perfect predictor of salaries and wages, we would know what salaries and wages would be given any set of marginal tax rates. 2020 is a perfect example of the shortfalls of looking at only marginal tax rates, due to COVID-19 unemployment is growing which means salaries and wages is unlikely to be as high as one would predict. It is

also probable that there would be diminishing returns to changing the variables. For example, looking at the variable median corporate rate times base the regression suggests for every one increase to this variable salaries and wages increase by $87.71 however if you were to keep raising this variable it would be unreasonable to say that the continuous rise this variable would increase salaries and wages an infinite amount.

A more reasonable idea to take from this is if the government needed to increase revenue with corporate taxation but didn't want to harm salaries and wages they should increase the median rate and the level of income which that median rate begins applying rather than the maximum rate. Looking at the individual tax rates a reasonable conclusion to draw would be that in order to increase tax revenue the lower quartile rate times base should be kept as low. If the goal was to increase taxes this should be done by increasing the rates and bases that are above the median but below the maximum. This would increase the mean rate times base as well as the amount paid if AGI is equal to the maximum tax base which are the variables based on individual tax rates with positive coefficients. Changes to the highest rate times base have different effects depending on the year because you also must consider how it would affect the mean rate times mean base so a general conclusion cannot be made for all years.

It can be an interesting topic to consider why changes to tax rates have any effect on earned income. Possible explanation for how increasing taxes also could increase salaries and wages would be that the additional revenue to the government is used to create more jobs. Having that increase in revenue could lead to increases to hiring within the government and in government agencies. Also, it could increase the amount the government is spending on contracts with private businesses which could cause them to hire more people that are compensated with salaries and wages. Right now, these statements are speculations and it would

take a different analysis to find out if they are true. They are only stated because this analysis raises further questions about why this model predicts the effects that it did.

Just as there are possible explanations for how increasing taxes increases salaries and wages there are possible explanations to why increasing some of these variables decrease salaries and wages. Two variables in specific are highest individual rate times highest individual base and highest corporate rate times highest base. When looking at highest individual rates and their bases its notable the most Americans don't reach these levels of income. The people that do reach those level of income would have more of a means to invest in businesses so their after-tax earnings could change their level of investment which may influence salaries and wages.

For corporate tax rates it would make sense that if you increase the tax expense, corporations will try to offset that cost. One cost that a company has a lot of influence over is their compensation expense which could be a reason why aggregate salaries and wages lower as highest corporate tax rate is increases. Now this doesn't address why salaries and wages increase when the median rate times base is increased. I haven't been able to come up with an explanation for why on this because this research only shows what happens not why it happens.

All the explanations offered are only theories because the purpose of this was to determine what the effects of changing marginal rates was. While it has been successful at serving that purpose, it opens the doors to an abundance for further research on why this happened. Each theory made to try to explain what changes occur could become new research projects where they could be either proven or disproven.

The question of what the effect of changes in tax rates are on earned income are could also be taken further. Some of the limitations of this were there wasn't an adjustment for changes to what was considered income or what kind of deductions were available. These are

improvements that could be addressed in further research. Also looking at only marginal tax rates limited how much of a conclusion could be drawn from the coefficients in the regression. A limitless amount time could be spent finding new variables to try to use to predict salaries and wages. Taking this idea further would involve finding more of those variables that have a significant impact on salaries and wages to improve this model. I believe that doing this would improve the accuracy of the effect of increasing any of the variables by one. The last limitation that I would address in furthering this research is that it only involved salaries and wages and not all forms of earned income as the initial question stated. Going forward it would be interesting to explore if the effects changed when looking into the less common forms of earned income.

The final take away to this project was that changing tax rates will influence salaries and wages. Increasing the lower quartile individual rate times base, the median individual rate times base, the highest individual rate times bases, and the highest corporate rate times base will decrease salaries and wages. Increasing the mean individual rate times base, the amount paid if AGI is equal to the highest base, and the median corporate rate times base will increase salaries and wages.

Citations

Beattie, A. (2019, March 12). A Concise History Of Changes In U.S. Tax Law. Retrieved October 7, 2019, from https://www.investopedia.com/articles/tax/10/concise-history-tax-changes.asp.

Beers, B. (2020, February 19). What P-Value Tells Us. Retrieved from https://www.investopedia.com/terms/p/p-value.asp

Blakey, R., & Blakey, G. (1935). The Revenue Act of 1935. The American Economic Review, 25(4), 673-690. Retrieved from www.jstor.org/stable/1807804

Ducasse, S. G. (2019, October 2). RMSE: Root Mean Square Error. Retrieved from https://www.statisticshowto.com/rmse/

Federal Corporate Income Tax Rates - Corporate Tax Rate & Jobs - ProCon.org. (n.d.). Retrieved from https://corporatetax.procon.org/federal-corporate-income-tax-rates/

Frost, J. (2019, March 30). Overfitting Regression Models: Problems, Detection, and Avoidance. Retrieved from https://statisticsbyjim.com/regression/overfitting-regression-models/

Historical Sources of Income and Tax Items. (2019, October 2). Retrieved from https://www.taxpolicycenter.org/statistics/historical-sources-income-and-tax-items

Income and Wealth Taxes (1934-1941). (n.d.). Retrieved from https://livingnewdeal.org/glossary/income-and-wealth-taxes-1934-1941/.

Kagan, J. (2019, November 18). Tax Reform Act of 1986. Retrieved from https://www.investopedia.com/terms/t/taxreformact1986.asp.

Kagan, J. (2020, January 29). Earned Income. Retrieved from https://www.investopedia.com/terms/e/earnedincome.asp

Moody, J. (2019, September 6). What does RMSE really mean? Retrieved from https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e

Wages and salaries were 92 percent of income before taxes for consumers ages 25 to 34 in 2014. (2016, June 28). Retrieved from https://www.bls.gov/opub/ted/2016/wages-and-salaries-were-92-percent-of-income-before-taxes-for-consumers-ages-25-to-34-in-2014.htm

WEISS, S. (1970). SUMMARY OF TAX REFORM ACT OF 1969. The Business Lawyer, 25(3), 973-999. Retrieved from www.jstor.org/stable/40684568

What is Linear Regression? (n.d.). Retrieved from https://www.statisticssolutions.com/what-is-linear-regression/