Determinants of Unemployment Rate: Final Project Report

Ashley Clarke

Link to GitHub Repository.

Contents

Executive Summary

Problem

Unemployment rates fluctuate over time as the economy goes through periods of recession and depression. Every month the Bureau of Labor Statistics releases an "Unemployment Situation Summary" and a "Gross Domestic Product Summary". While it is often assumed that factors in these two reports are linked to the United States unemployment rate, analysts often do not know which of the thousands of variables in these reports they should focus their attention on. Variables examined range from net government lending to the number of employees by industry sector. While using data from 1960 to 2020 inherently ignores shifts in employment patterns across time, the purpose of this report is to determine which factors have historically been predictive of the U.S. unemployment rates, not to predict future unemployment rates. Therefore, this report aims to identify and explain how different factors in these two reports relate to the U.S. unemployment rate.

Data

(1) Employment Situation Summary[1]

(2) Gross Domestic Product[2]

(3) Additional Federal Reserve of Economic Data Variables: Unemployment Rate, Inflation, Federal Funds Rate

My dataset combines data from two sources. First, I pulled the "Unemployment Situation Summary" and the "Gross Domestic Product Summary" time-series data sets from the U.S. Bureau of Labor Statistics from 1954 to 2021. Next, I pulled three additional variables from the Federal Reserve Bank of St. Louis (unemployment rate, inflation, and federal funds rate). For the two time-series data sets, I pulled every key variable available. My primary response variable of interest was the unemployment rate, which is defined as the number of unemployed persons as a percentage of the labor force. While the original data set has 804 observations and 2004 variables, the clean data set has 732 observations and 160 variables.

Analysis

Before exploring the dataset or running any analyses, I split the data into train and test data sets. I trained my model using the training data set and reserved the test dataset for measuring model performance. Next, I accessed correlations between variables and found multicollinearity issues, which I systematically dealt with. Next, I explored my data to look for relationships between variables and the response and checked to make sure all linear regression assumptions were met. To determine which features are predictive of the unemployment rate, I built six different cross-validated models: ordinary least squares, ridge regression, LASSO regression, elastic net regression, random forest, and boosting. Of the regression models, the ordinary least squares (OLS) regression has the lowest test error. The boosted model had the lowest test error of all models (and of the tree-based models). Finally, I drew conclusions based on what I learned from my models.

Conclusions

I found that both regression and tree-based methods found similar variables to be strong predictors of the unemployment rate. Specifically, the boosted model found depreciation of fixed assets and the number of mining/logging employees to be the strongest predictors, revealing how changes in the number of employees for "blue-collar" professions are more predictive than changes in the number of employees for "white collar" professions. I also found that unemployment rises as government debt and spending rise. Additionally, I found that the unemployment rate falls when net exports as a percentage of GDP, money invested in fixed assets, and corporate profits increase. I hope this analysis can reframe how economists and analysts think

[1] https://www.bls.gov/news.release/empsit.nr0.htm
[2] https://www.bea.gov/data/gdp/gross-domestic-product

about unemployment, both in the context of what signals potentially high future unemployment and the effects of unemployment.

Introduction

Background information

People are classified as unemployed if they do not have a job, have actively looked for work in the prior four weeks, and are currently available for work[3]. Only people classified as part of the civilian labor force are included in the unemployment rate metric.

Unemployment rates fluctuate over time. During periods of recession and depression, unemployment is high. During periods of economic growth, unemployment tends to be lower. For example, the unemployment rate was 25% during the Great Depression, 10.8% in November 1982, and 14.7% in April 2020.[4]. Unemployment rates in the late 1990s and into the mid-2000s were low by historical standards. The unemployment rate was below 5% from 1997 to 2000 and near 5% during almost all of 2006–2007[5].

Unemployment generally falls during periods of economic prosperity and rises during recessions, creating significant pressure on public finances as tax revenue falls and social safety net costs increase. Government spending and taxation decisions (fiscal policy) and U.S. Federal Reserve interest rate adjustments (monetary policy) are crucial tools for managing the unemployment rate. There may be an economic trade-off between unemployment and inflation, as policies designed to reduce unemployment can create inflationary pressure, and vice versa. The U.S. Federal Reserve (the Fed) has a dual mandate to achieve full employment while maintaining a low rate of inflation as shown in (Figure 1). Historically, the Fed has targeted a 5% unemployment rate and 2% inflation.

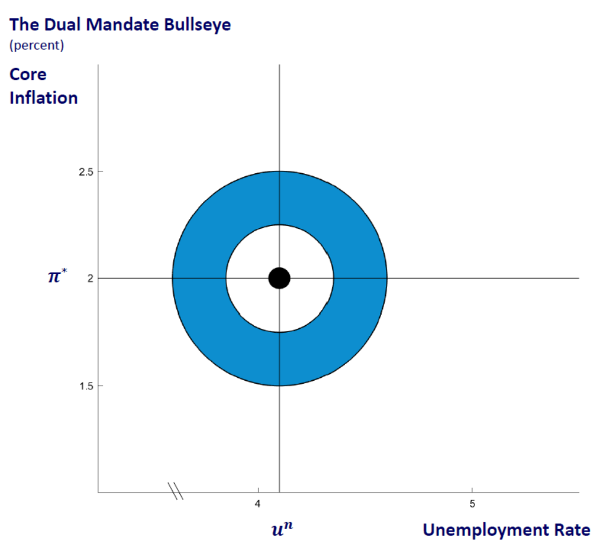


Figure 1: The Federal Reserve has a dual mandate to maintain low inflation and to achieve full employment

There are a variety of measures used to track the state of the U.S. labor market. The Bureau of Labor

[3] "How the Government Measures Unemployment". www.bls.gov.

[4] "Unemployment is nearing Great Depression levels. Here's how the eras are similar — and different". https://www.cnbc.com/2020/05/19/unemployment-today-vs-the-great-depression-how-do-the-eras-compare.html

[5] "Patterns of Unemployment",https://opentextbc.ca/principlesofeconomics/chapter/21-2-patterns-of-unemployment/

Statistics provides a "chartbook" displaying the major employment-related variables in the economy [6]. Members of the Federal Reserve also give speeches and Congressional testimony that explain their views of the economy, including the labor market.[7].

This report focuses on two reports released by the Bureau of Labor Statistics. For background, U.S. employment statistics are reported by government and private primary sources monthly, and these metrics are widely quoted in the news. These sources use a variety of sampling techniques that yield different measures. While the U.S. Bureau of Labor Statistics (BLS) provides a monthly "Employment Situation Summary", analysts often do not know which factors, amongst the hundreds provided, drive unemployment rates. Additionally, employment trends can be analyzed by looking at the state of the economy. The BLS also provides a "Gross Domestic Product Summary" monthly, which provides the real gross domestic product (GDP) and other features related to GDP.

Analysis goals

Given my knowledge that unemployment rates rise and fall in response to economic conditions, I sought to investigate which factors in the "Employment Situation Summary" and "Gross Domestic Product Summary" are predictive of unemployment. Additionally, I determined whether these factors were more or less predictive than inflation and the federal funds rate. The purpose of this analysis is not necessarily to predict future unemployment rates. Instead, my goal is to determine which features have been predictive of the U.S. unemployment rate in the past. To achieve this goal I will predict based on the 158 variables outlined in Table **??**, these variables consist of employment-related variables and economic-related variables. I will consider my analysis a success if I can identify a set of variables that are found to be predictive of the U.S. unemployment rate using multiple modeling techniques. Additionally, I will determine which model best matches the underlying trend in the data by calculating test error. In addition to calculating test root-mean-square-error (RMSE) for each of my models, I will calculate the test RMSE error for an intercept-only model as a baseline.

Significance

This analysis goal is important to address in the context of application because determining which factors are linked to high unemployment can help shape fiscal and monetary policy. Additionally, I hope my analysis will allow analysts and policymakers to recognize signals of high unemployment, which can lead to significant economic downturns. Since many explanatory variables in this analysis relate to specific industry sectors, my analysis also aims to identify which sectors of the economy are most affected by changes to the unemployment rate. Lastly, I hope to improve the interpretability of the BLS's Employment Situation and GDP reports for all interested parties.

Data

Data sources

The raw dataset includes merged data from three sources from June 1954 to November 2021

(1) Unemployment Situation Summary[8]

Each month, the Bureau of Labor Statistics publishes the Employment Situation Summary report based on information from the prior month. The data for the report is derived primarily from two sources: a survey of approximately 60,000 households, or about 110,000 individuals (household survey), and an establishment survey of over 651,000 worksites.

(2) Gross Domestic Product Summary[9]

[6] "Current Population Survey (CPS)". Bls.gov.
[7] "The Fed - Speeches of Federal Reserve Officials". Federalreserve.gov.
[8] https://www.bls.gov/news.release/empsit.nr0.htm
[9] https://www.bea.gov/data/gdp/gross-domestic-product

Each month, the Bureau of Labor Statistics publishes the Gross Domestic Product Summary report based on information from the prior month. The report includes Gross Domestic Product (GDP), which is a comprehensive measure of U.S. economic activity. GDP measures the value of the final goods and services produced in the United States (without double counting the intermediate goods and services used up to produce them). The report includes variables relating to GDP.

(3) Additional Federal Reserve of Economic Data Variables: Unemployment Rate[10], Inflation[11], Federal Funds Rate[12]

Process of downloading data

The data was collected using the fredr package[13], which provides a complete set of R bindings to the Federal Reserve of Economic Data (FRED) RESTful API, provided by the Federal Reserve Bank of St. Louis. The fredr package allowed me to search for and fetch time series observations as well as associated metadata within the FRED database. Since FRED organizes their data using variable ids, I downloaded time series observations from all variable ids in the Employment Situation and Gross Domestic Products reports, which represents over 2000 variables, from June 1954-November 2021. Additonally, I downloaded the U.S. unemployment rate (response variable), inflation rate, and federal funds rate. Before cleaning, the data set consisted of 804 observations and 2004 variables.

Due to the size of the data set, the data set takes around 5-10 minutes to download from FRED. Also, since the fredr package has a limit of 120 requests / minute, it might take longer than expected for the data to load.

Data cleaning

Three critical issues were resolved during the data cleaning phase:

(1) Features have not been reported all years

(2) Features are reported in different time increments: monthly, quarterly, and yearly

(3) Many of the features are highly correlated with each other since duplicate features are included

1. Timeframe Issues: Not every feature has been reported since 1954

While the unemployment rate has been reported monthly since June 1954, many other features have not been reported for the entire timeframe. Additionally, certain metrics have not yet been reported for 2021. Therefore, I decided to keep only observations from January 1960 to December 2020. Features that have not been reported since 1960 were dropped.

2. Reporting frequency: Not all features are reported monthly

While both the Employment Situation and Gross Domestic Product reports are released monthly, not every feature is updated monthly. Many features are reported either quarterly or yearly. This issue was identified by examining the number of observations per feature. I noticed that many features had eith 61 or 244 complete observations as shown in (Figure 2).

This makes sense because the time frame of the dataset corresponds to 61 years and 244 quarters (3 months each). To impute the missing values for yearly data, I set every month in the yearly equal to the yearly metric. For missing quarterly data, I set the next two months equal to the quarterly metric. While I recognize that this is not a perfect way to impute these features, I believe it is better than dropping entire columns or rows. After imputing missing values for quarterly and yearly data, I dropped all columns with NA values, which left me with 831 features.

[10] https://fred.stlouisfed.org/series/UNRATE
[11] https://fred.stlouisfed.org/series/FPCPITOTLZGUSA
[12] https://minds.wisconsin.edu/bitstream/handle/1793/77330/Federal%20Funds%20Rate.pdf?sequence=1&isAllowed=y
[13] https://cran.r-project.org/web/packages/fredr/vignettes/fredr.html

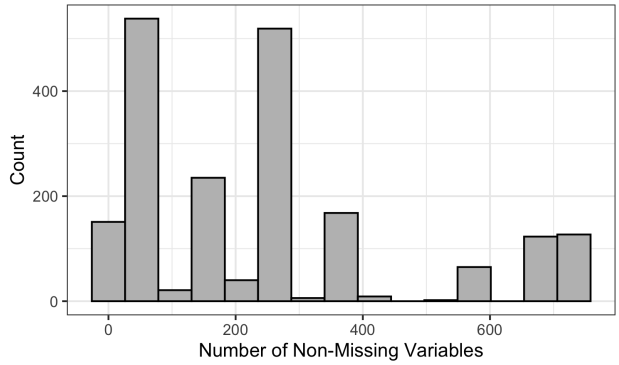


Figure 2: Histogram of the number of observations per feature that are complete. There are peaks at 61 (number of years) and 244 (number of quarters)

3. Duplicate features: Some features represent the same metric with minor adjustments

In their monthly reports, FRED makes minor adjustments to metrics and reports them as separate features. For instance, both seasonally adjusted and non seasonally adjusted numbers are reported for most metrics. To eliminate the double-counting variables, if a variable and another variable have a higher than 0.9 correlation with each other, I only kept one of the variables. Correlation is more informative than covariance when deciding which variables to remove because it does not depend on the scales of the features and always lies in the interval [-1,1]. The closer the correlation is to the endpoints of this interval, the more strongly the features are related. If the correlation is very high, it is likely the variables measure the same thing. Additionally, due to potential multicolinearity issues, I decided that it was better to remove features that are highly correlated with each other. After removing variables that are highly correlated, there are 160 remaining features.

Removed columns were the standard deviation is equal to zero

I calculated the standard deviations of all variables. If a variable had a standard deviation of 0, I removed it because it is a meaningless feature.

Next, I examined the independence of observations

Since each observation is indexed by month, I was concerned that samples of consecutive months would be strongly correlated (i.e. the unemployment rate of a given month would depend largely on the unemployment rate from the previous month). Intuitively, a time series is weakly dependent if events in the past have only a small influence on the value of the time series at the present moment[14].

While the unemployment rate yesterday has a large influence on the unemployment rate today, the unemployment rate last month has only a small influence on the unemployment rate this month. To test this hypothesis, I calculated the mean percent change between observations to be 0.225%. Since the standard deviation is 1.68%, there is most likely some dependence between observations, but this dependence is not too high. Also, Brookings finds that monthly unemployment rates behave similarly to independent random variables, since they can spike or drop at any time[15]. Therefore, I can still conduct my analysis.

In an optimal world, I would be able to time-average the data over a larger time frame. However, time-averaging would have produced too small of a data frame here. Thus, I proceeded with caution, knowing that the independent and identically distributed or i.i.d. assumption was not 100% met.

Data description

Observations: The cleaned data set has a total of 732 observations, corresponding to each of the 732 months between January 1960 and December 2020.

[14] https://towardsdatascience.com/time-series-analysis-part-i-3be41995d9ad
[15] https://www.brookings.edu/wp-content/uploads/2016/07/2013b_coibion_unemployment_persistence.pdf

Response Variable: Unemployment Rate (UNRATE)[16] is the response variable and is continuous. The unemployment rate represents the number of unemployed as a percentage of the labor force. Labor force data are restricted to people 16 years of age and older, who currently reside in 1 of the 50 states or the District of Columbia, who do not reside in institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces. The response variable is reported monthly and is seasonally adjusted.

Explanatory Variables

The cleaned data set includes 158 features, and documentation of each feature can be found in Table **??**. 156 of the explanatory variables consist of employment-related variables and economic-related variables that are pulled from the employment situation and GDP reports. Additionally, I included both inflation and the federal funds rate as features. All features are continuous.

Detailed descriptions of added variables:

Inflation: According to economic theory, as unemployment rates fall, the rate of inflation rises. This has been formalized according to what is known as "the Phillips Curve", which is shown in Figure 3.

Inflation (FPCPITOTLZGUSA)[17] as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used. This metric is not seasonally adjusted and is recorded annually.

Federal Funds Rate: It is thought that the unemployment rate and federal funds rate have a negative contemporaneous relationship. I expect that when the unemployment rate is at its highest, the federal funds rate will be at its lowest. This likely happens because there is a lower federal funds rate in a weak economy[18].

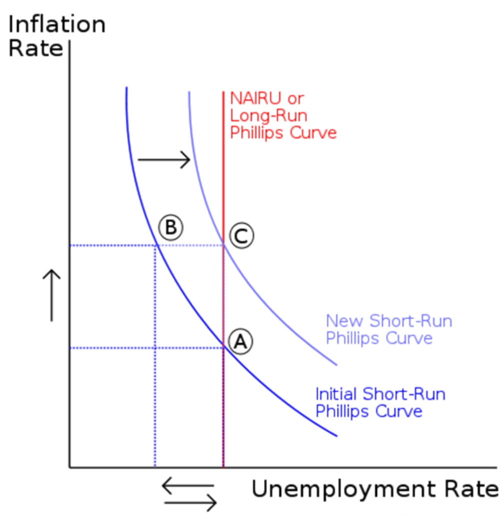


Figure 3: Phillips Curve

The federal funds rate (FEDFUNDS)[19] is the interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight. When a depository institution has surplus balances in its reserve account, it lends to other banks in need of larger balances. In simpler terms, a bank with excess cash, which is often referred to as liquidity, will lend to another bank that needs to quickly raise liquidity.

[16] https://fred.stlouisfed.org/series/UNRATE
[17] https://fred.stlouisfed.org/series/FPCPITOTLZGUSA
[18] https://minds.wisconsin.edu/bitstream/handle/1793/77330/Federal%20Funds%20Rate.pdf?sequence=1&isAllowed=y
[19] https://fred.stlouisfed.org/series/FEDFUNDS

Data allocation

To allocate my data, I used an 80-20 split, such that the training dataset consists of 80% of observations and the test data set consists of 20% of observations. Observations were allocated randomly. The same train-test split was used for each class of methods. Additionally, data exploration used solely the train data set. Thus, there are 585 observations in test dataset and 147 observations in the train dataset.

Data exporation

Response Variable

First, I looked at the response variable's distribution. As seen in the histogram of the unemployment rate variable (Figure 4), the data appears to be right-skewed, with some months having an unemployment rate that exceeds 10%. The median unemployment rate is 5.6%.

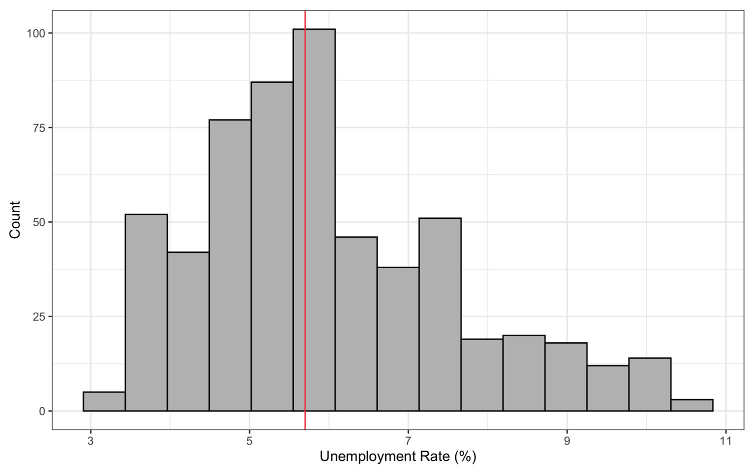


Figure 4: Histogram of Unemployment Rate

Next, I looked at which years have the most extreme unemployment rates and determined that those months corresponded to recessionary periods. Figure 5 shows that when unemployment rates are aggregated across each year, the highest unemployment rates occur in 1975-1976, 1981-1984, and 2009-2012, which are all recession years.

year	UNRATE
1983	9.6
1982	9.5
2010	9.5
2009	9.4
2011	8.9
1975	8.5
2012	8.1
1976	7.7
1981	7.6
1984	7.5

Figure 5: Years with the Highest Average Monthly Unemployment

Then, I looked at the mean unemployment rate from 1960-2020 and from 2010-2020, which are reported in Figure 6. I found that recent unemployment rates are relatively consistent with unemployment rates across the entire time frame.

Time	Unemployment Rate
1960-2020	5.98
2010-2020	6.17

Figure 6: Mean Unemployment Rate by Time Period

Finally, I plotted unemployment rate by date for all years (Figure 7) to visualize how the unemployment rate has changed over time.

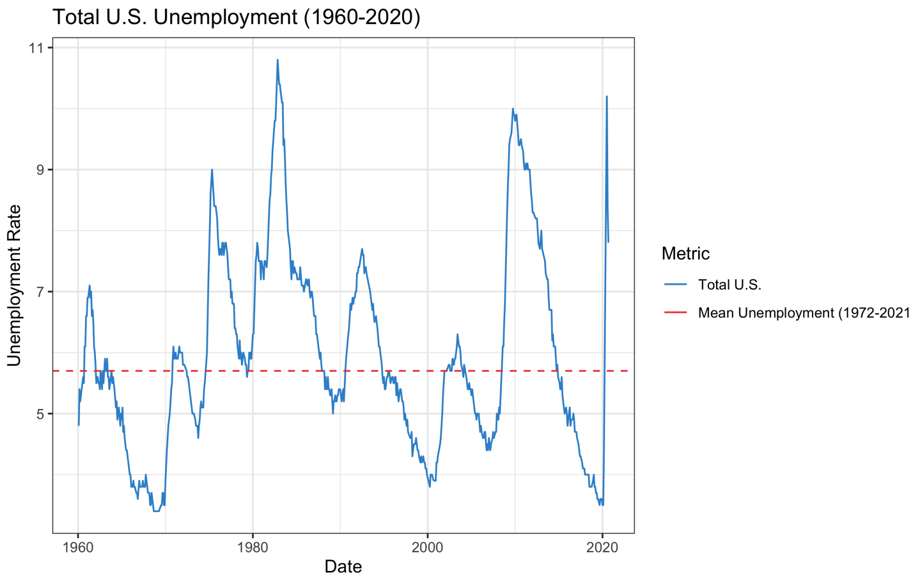


Figure 7: Total U.S. Unemployment by Date

Covariation between features

As mentioned above, many of the original features had high covariation since FRED reports include adjusted and unadjusted metrics. After cleaning the data to adjust for multicollinearity issues, no features in the data set have a correlation of 0.9 or above with each other.

While there are far too many features to create a correlation plot with every feature, I randomly subsampled 30 features to create a correlation matrix. Figure 8 demonstrates how the majority of the features are not highly correlated with each other. Though I do not believe multicollinearity is a major cause for concern after cleaning the data set, penalized regression techniques (e.g. lasso, ridge) will help adjust for possible multicollinearity issues.

Additionally, in Figure 9, I made a correlation plot that shows the correlation plot between the unemployment rate and all explanatory variables. The plot does not indicate that any variables are strongly positively / negatively correlated with unemployment rate. Therefore, I collect the five features with the hightest absolute correlation with the response variable in Table 1.

Table 1 indicates annual and quarterly depreciation of fixed assets/consumption of fixed capital (A024RL1A225NBEA, A024RL1Q225SBEA) are negatively correlated with the unemployment rate meaning as depreciation increases, the U.S. unemployment rate decreases. Also, as the number of mining and logging employees (CES1000000006) and the number of female mining and logging employees (CES1000000010) increases, so does the U.S. unemployment rate. This potentially suggets that people turn to mining/logging jobs when they cannot find jobs elsewhere. Finally, the average hours of production in the manufacturing

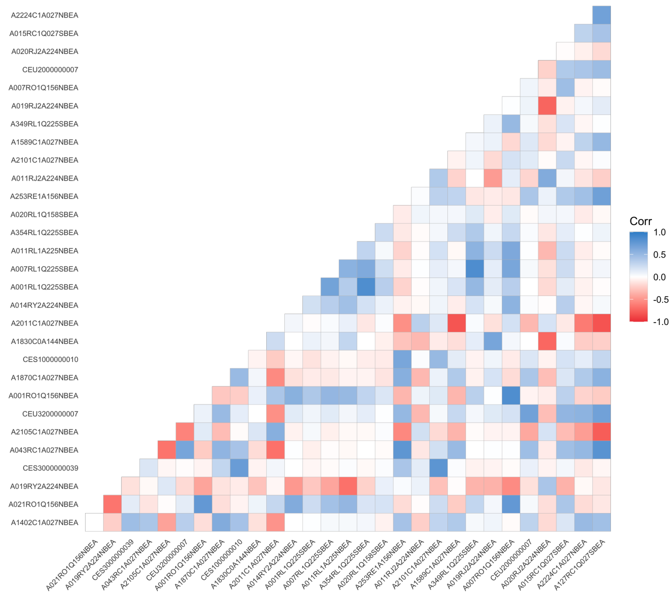


Figure 8: Correlation plot for a subsample of 30 randomly selected variables

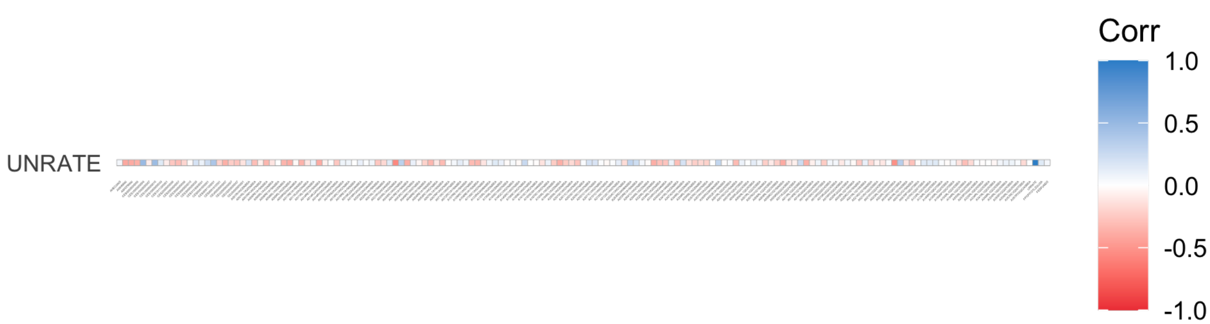


Figure 9: Plot of correlations between unemployment rate and all variables

Table 1: Coefficents with the highest correlation to unemployment rate

Feature	Correlation	Description
A024RL1A225NBEA	-0.57	Real Consumption of Fixed Capital: Private
A024RL1Q225SBEA	-0.53	Real Consumption of Fixed Capital: Private
CES1000000006	0.49	Production and Nonsupervisory Employees, Mining and Logging
CES1000000010	0.48	Women Employees, Mining and Logging
AWHMAN	-0.41	Average Weekly Hours of Production and Nonsupervisory Employees, Manufacturing

sector (AWHMAN) is negatively correlated with unemployment rate, suggesting as manufacturing employees work more hours, the unemployment rate declines.

Initial Insights: Employment Situation Features

0.0.1 Number of Employees by Industry

Next, I wanted to see how the number of employees across sectors is correlated with the unemployment rate Specifically, I looked at the logging, shipping/boating, information, and federal sectors. Due to the recent increase in the number of information jobs, I also plotted information vs. unemployment for 2000-2020. Figure 10 suggests that as the number of employees declines in any industry, the unemployment rate rises. However, the number of federal employees stays relatively constant despite changes in the unemployment rate. It is important to note that changes in the number of employees per industry might not be due to changes in the unemployment rate. Instead, demand for workers across sectors might have shifted during the time frame of the dataset.

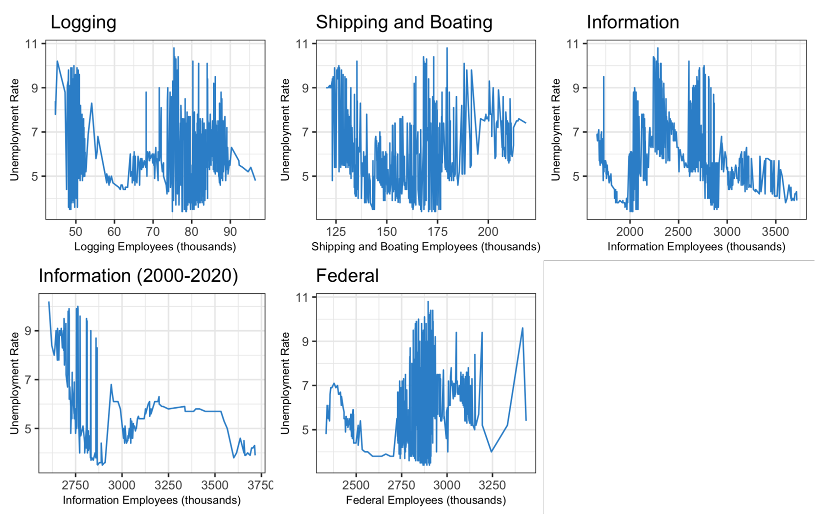


Figure 10: Unemployment Rate by Number of Employees by Industry

0.0.2 Average Weekly Hours of Production by Industry

I would predict that as the average number of hours of production increases, the unemployment rate decreases. I looked at the manufacturing, mining/logging, and construction industries. Figure 11 shows that my prediction holds. Out of selected variables, the manufacturing industry has the strongest correlation between an increase in average weekly hours of production and a decrease in the unemployment rate.

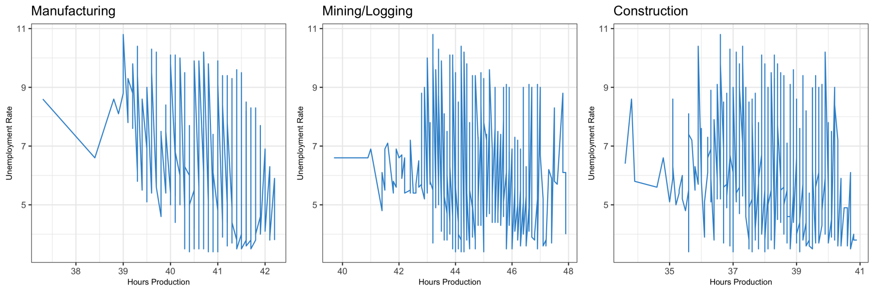


Figure 11: Unemployment Rate by Average Weekly Hours of Production by Industry

Initial Insights: Gross Domestic Product Summary Features

I also examined features that were present in the GDP report. I chose to plot four variables: social benefits to persons, net government saving, real consumption of fixed capital, and real disposable personal income in (Figure 12). Unemployment rate increases when government social benefits increase, which likely occurs because unemployment is a qualifying factor for many of these benefits. Also, as real consumption of fixed capital and real disposable personal income increases, the unemployment rate generally falls. This makes sense because corporations are willing to invest more in fixed assets (e.g. buildings) when the economy is doing well. Additionally, disposable personal income increases when someone is employed and when economic growth is strong.

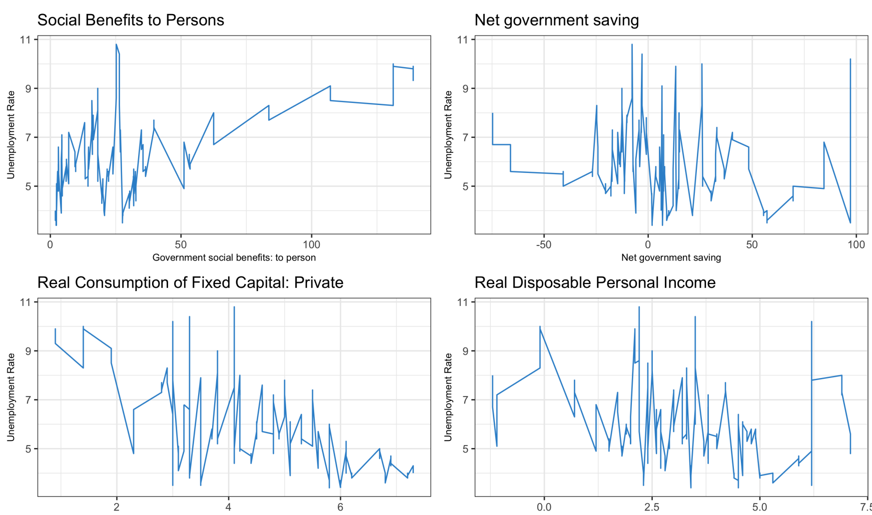


Figure 12: Unemployment Rate by Variables from GDP Report

Analysis of Inflation and Federal Funds Rate

0.0.3 Inflation

I would expect inflation to rise as the unemployment rate decreases due to the Phillips Curve (Figure 3). However, as the unemployment rate falls, inflation does not appear to rise. There is no apparent relationship between these two variables.

0.0.4 Federal Funds

Next, I examined whether the unemployment rate is negatively correlated with the federal funds rate. I only looked at the last 20 years because line plots with all 585 observations can be hard to interpret due to a

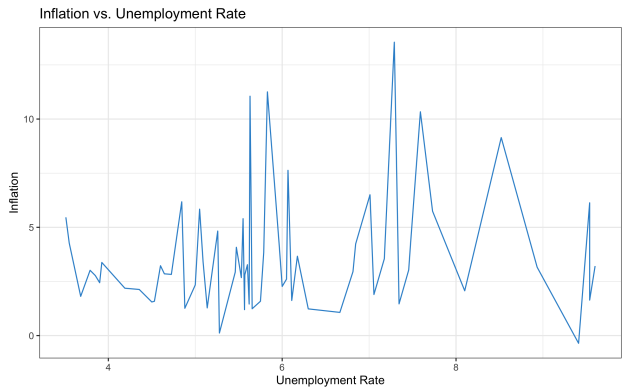


Figure 13: Inflation vs. Unemployment Rate. Replicates the Phillips Curve for dataset.

few outliers. Figure 14 implies a direct, negative linear relationship between the federal funds rate and the unemployment rate. Thus, Figure 14 implies when the unemployment rate is high, the federal funds rate will be low and vice versa.

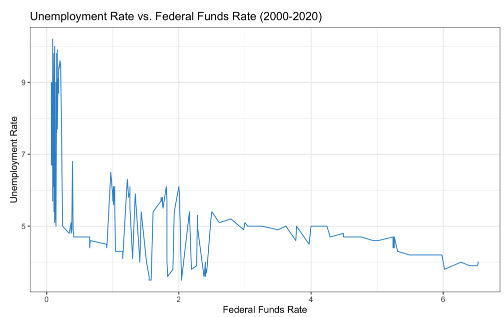


Figure 14: Unemploy by Federal Funds Rate

Variation in features

I also also curious to examine the variation in features. I randomly selected four features to look at: the number of goods producing employees (CES0600000006), the change in real motor vechile output (A133RL1Q225SBEA), the change in private inventories (A014RE1A156NBEA), and the change in real gross national product (A001RO1Q156NBEA). These features are plotted in Figure 15.While three out of four of the features appear to be approximately normally distributed, the change in real motor vechicle output is skewed to the right because of two large outliars. If I had more time, I would look at the varation in every feature.

Modeling

Modeling Class 1: Regression Methods

Ordinary least squares

To start, I created an ordinary least squares model that included all explanatory variables. Based on the regression summary, the ordinary least squares model has an r-squared value of 0.991, which implies that the features explain about 99.1% of the variation in the unemployment rate. Additionally, there are 34 statistically significant features in the model at a 0.05 threshold. (Table 2) shows the features with the lowest p-values.

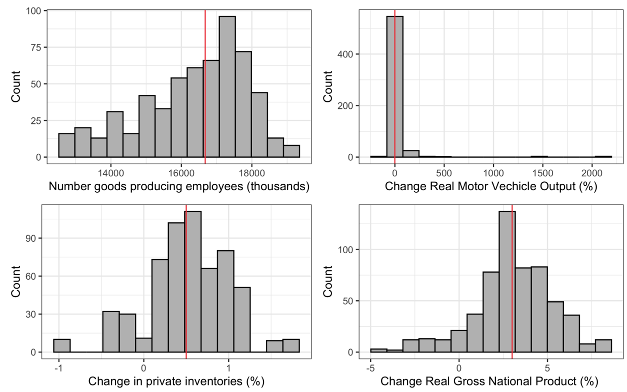


Figure 15: Histograms that show variation of four randomly selected features, where the red vertical line represents the median value

Table 2: Coefficents with the lowest p-value for ordinary least squares model

Feature	p-value	Description
A136RL1Q225SBEA	8.00e-08	Real Motor Vehicle Output: Final Sales of Domestic Product: Personal Consumption Expenditures: New Motor Vehicles: Autos
A354RL1Q225SBEA	3.34e-06	Real Gross Domestic Product: Goods: Durable Goods
A014RY2A224NBEA	4.01e-05	Contributions to percent change in real gross domestic product: Gross private domestic investment: Change in private inventories
A023RL1Q225SBEA	7.57e-05	Real Gross National Income
A2009L1A225NBEA	2.23e-04	Real Gross Housing Value Added
A030RC1Q027SBEA	2.30e-04	Net lending or net borrowing (-), NIPAs: Government: Statistical discrepancy
CES0600000006	2.67e-04	Production and Nonsupervisory Employees, Goods-Producing
A043RC1A027NBEA	3.43e-04	National income: Proprietors' income with IVA: Farm
A019RE1A156NBEA	4.32e-04	Shares of gross domestic product: Net exports of goods and services
A356RL1Q225SBEA	9.60e-04	Real Gross Domestic Product: Goods: Nondurable Goods

Next, I checked to make sure the necessary assumptions for linear regression are met in Figure 16. First, I checked the linearity of the data in the Residuals vs. Fitted plot. Though there are a few outliers, the residual plot shows no fitted pattern, and the red line is approximately horizontal at zero. Next, I check the homogeneity of residuals variance by looking at the Scale-Location plot. This plot shows that residuals are spread equally along the ranges of predictors. The QQ plot of residuals is used to check the normality assumption. The normal probability plot of residuals should approximately follow a straight line, which it does. Finally, I checked to see if the residuals have a constant variance in the Residuals vs. Leverage plot. Since there are only 2 extreme outliers, it is okay to assume this assumption is met.

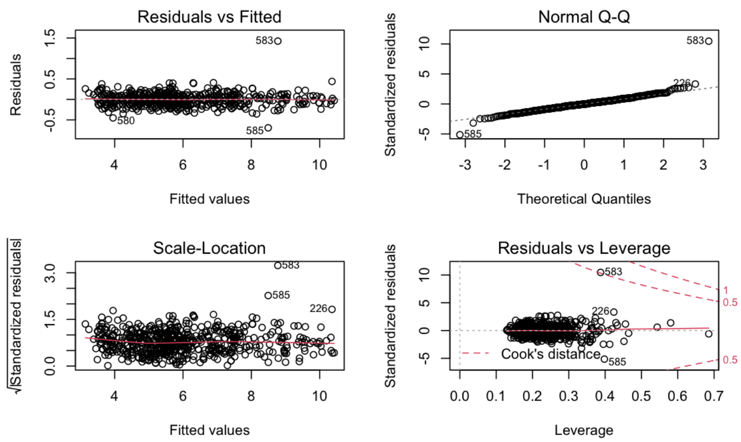


Figure 16: Check assumptions for ordinary least squares model

While the ordinary least squares method seems to work well, training a model with such few observations on so many explanatory variables might lead to overfitting. Due to a high r-squared value, I suspect that bias is low, but the variance is high in the ordinary least squares model. If there are many features, like there are in my least-squares model, the coefficients can become very large, which creates high variance. To temper the effects of large coefficients, I wanted to disincentivize large values. Therefore, I decided to build and evaluate shrinkage models with the hopes of getting a more interpretable and accurate model. I ran three cross-validated regressions for which optimal values of lambda were chosen according to the one-standard-error rule: ridge, LASSO (Least Absolute Shrinkage and Selection Operator), and elastic net.

Ridge Regression

The ridge regression adds a penalty to coefficients to disincentivize large values. The larger lambda is, the more of a penalty there is. If lambda equals 0, the ridge regression produces the same results as the ordinary least squares model. When lambda approaches infinity, only the intercept is left. Since lambda controls the flexibility of the ridge fit, I tuned my model to find the optimal value of lambda. For ridge regression, all features are assumed on the same scale. However, feature standardization is done behind the scenes in R. Another benefit of the ridge model is that while linear regression coefficients for correlated features tend to be unstable, penalized regressions provide a more stable way of "splitting the credit" amongst variables.

To train my model, I fit a 10-fold cross-validated ridge regression model to the training data. In (Figure 17) we have the CV plot for the model. Corresponding to the right vertical dashed line on the plot (on the log scale), the value of lambda selected according to the one-standard-error rule is about 0.111.

Next, I used plot_glmnet to visualize the ridge regression fitted coefficients, highlighting the six features with the highest absolute standardized coefficients. It appears that none of the coefficients change signs as lambda increases.

Table 3 shows the features with the highest standardized coefficients and provides their descriptions. Coefficents in Table 3 can be interpreted as follows: for every 1 standard deviation increase in the given featue there is a

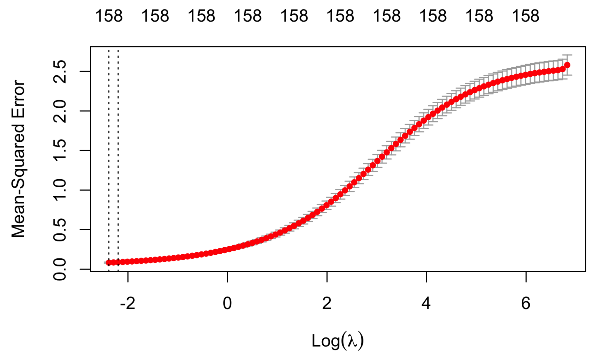


Figure 17: Ridge CV plot: This is the CV plot for the 10-fold cross-validated ridge regression model on the training data

"value of the coefficent" standard deviation change in the unemployment rate. Therefore, positive coefficients correspond to increases in the unemployment rate and negative coefficients correspond to decreases in the unemployment rate. One of the limitations of the ridge model is that I cannot obtain p-values from the ridge regression. Thus, the ridge model is only a prediction tool and cannot be used to determine statistical significance.

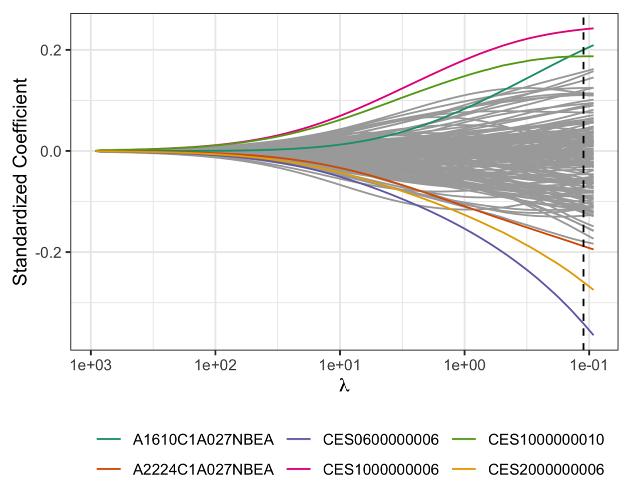


Figure 18: Ridge trace plot with 6 features with highest magnitude highlighted

Four of the features describe the number of employees by sector. As the number of goods-producing (CES0600000006) and construction (CES2000000006) employees increases, unemployment falls. However, when the number of mining/logging employees (CES1000000006) and the number of female mining/logging employees (CES1000000010) increases, the unemployment rate rises. This trend might imply that people are more willing to enter the mining/logging field when it is difficult to find a job in another sector. Likewise, as the number of manufacturing overtime hours (A2224C1A027NBEA) increases, unemployment decreases, likely because overtime hours signal a tight labor market.

Additionally, as the balance on current accounts adjustment for U.S. territories (A1610C1A027NBEA) increases, the unemployment rate increases. The current account balance of payments is a record of a country's international transactions with the rest of the world[20]. As the United States territories trade more

[20] https://data.oecd.org/trade/current-account-balance.htm

Table 3: Standardized coefficients for features in the ridge model based on the one-standard-error rule.

Feature	Coefficient	Description
CES0600000006	-0.34	Production and Nonsupervisory Employees, Goods-Producing
CES2000000006	-0.26	Production and Nonsupervisory Employees, Construction
CES1000000006	0.24	Production and Nonsupervisory Employees, Mining and Logging
A1610C1A027NBEA	0.20	Balance on current account: Adjustment for U.S. territories and Puerto Rico
A2224C1A027NBEA	-0.19	Net corporate dividend payments: Rest of the world
CES1000000010	0.19	Women Employees, Mining and Logging
AWOTMAN	-0.18	Average Weekly Overtime Hours of Production and Nonsupervisory Employees, Manufacturing
A048RC1A027NBEA	-0.16	Rental income of persons with capital consumption adjustment
A2051C1A027NBEA	0.16	Gross domestic product: Farm products consumed on farms
A1589C1A027NBEA	0.15	Government social benefits: to persons: Federal: Benefits from social insurance funds: Unemployment insurance

with foreign countries, there is a smaller demand for products from mainland United States, which drives up the unemployment rate. This occurs because U.S. territories are not included in the unemployment rate.

Three features are signals of the state of the economy. As corporate dividends (A2224C1A027NBEA) and rental income (A048RC1A027NBEA) increase, the unemployment rate decreases. When unemployment is low, corporations are most likely performing well. Thus, they have more retained earnings to pay out as dividends. Likewise, when individuals have jobs, they can afford higher rents, which drives up landlord income. Finally, as government social benefits/ unemployment insurance (A1589C1A027NBEA) increases, the unemployment rate increases. More individuals qualify for unemployment insurance as unemployment rates rise.

LASSO Regression

Like the ridge regression model, the LASSO model includes a penalty term that biases coefficients toward zero, which reduces variance. However, instead of merely shrinking the coefficients, the penalty term leads many of the coefficients to be 0. Since LASSO solutions are sparse, LASSO is a variable selection tool. However, like the ridge model, I cannot attach a measure of statistical significance to the selected variables. Thus, LASSO is just a prediction method. While feature scaling is necessary for LASSO, R takes care of this behind the scenes. One drawback of LASSO is that coefficients are unstable for correlated features because LASSO selects which of the correlated features to keep arbitrarily. Lambda controls the flexibility of the LASSO regression fit. Thus, I tuned my model to find the optimal lambda.

I fit a 10-fold cross-validated LASSO regression to the training data. In Figure 19, I have the CV plot for the model. Corresponding to the right vertical dashed line on the plot (on the log scale), the value of lambda selected according to the one-standard-error rule is about 0.00316. 80 features (excluding the intercept) are selected if lambda is chosen according to the one- standard-error rule.

Figure 20 shows the LASSO trace plot with the first 6 features to enter the model highlighted and Table 4 shows the top 10 selected features with the highest standardized coefficents. Coefficents in Table 3 can be interpreted as follows: for every 1 standard deviation increase in the given featue there is a "value of the coefficent" standard deviation change in the unemployment rate. Therefore, positive coefficients correspond to increases in the unemployment rate and negative coefficients correspond to decreases in the unemployment rate.

Examining the LASSO trace plot in Figure 20, I see that lambda decreases from left to right. Thus, depreciation of fixed assets / real consumption of fixed capital (A024RL1A225NBEA) is the first feature to enter the model as lambda decreases, and the coefficient is negative. Other features with negative standardized that enter the mdoel in the first six are the private domestic investment in nonresidential structures (A009RO1Q156NBE) and the average weekly hours of production for manufacturing employees (AWHMAN). Features with positive standardized coefficients include net government lending/borrowing (A030RC1Q027SBE), the number of goods-producing employees (CES1000000006), and the number of female

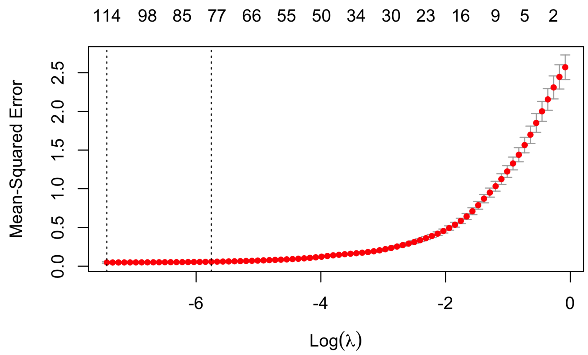


Figure 19: Lasso CV plot: This is the CV plot for the 10-fold cross-validated LASSO regression model on the training data.

mining and logging employees (CES1000000010).

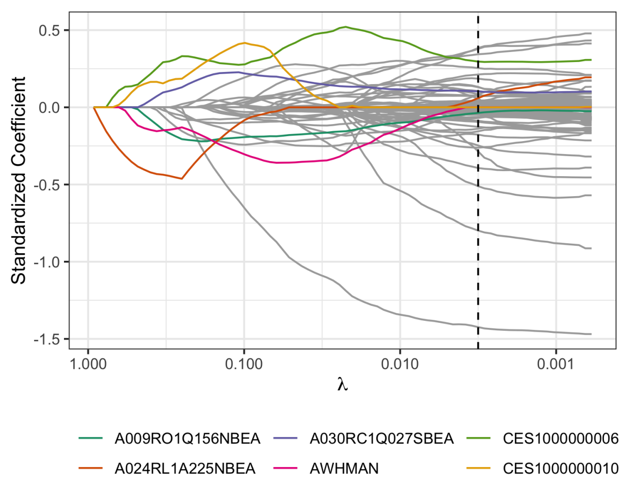


Figure 20: Lasso trace plot with first 6 features to enter the model highlighted

Table 4 shows the features with the highest standardized coefficents in the LASSO model. Many of the coefficients for the LASSO model are consistent with the ridge model. Three of the features describe the number of employees by sector. The LASSO model finds that as the number of goods-producing (CES0600000006) and construction employees (CES2000000006) increases, unemployment falls. Also, when the number of mining/logging for all employees (CES1000000006) increases, the unemployment rate rises. As rental income (A048RC1A027NBEA) increases, the unemployment rate decreases. As farm products consumed on farms (A2051C1A027NBEA) increase, so does the unemployment rate. Finally, as the amount of government social benefits (A1589C1A027NBEA) increases, the unemployment rate increases.

However, LASSO has four different features in the top 10 standardized coefficient list. As taxes on corporate income (A054RE1A156NBEA) increase, the unemployment rate decreases. This likely occurs because taxes are paid based on net income, which is higher when the economy is doing well. Thus, when corporate profits are higher, corporate taxes are higher.

While increased exportation of goods as of percent of GDP (A253RE1A156NBEA), is associated with higher unemployment, the net exportation of goods and services (A019RE1A156NBEA) is associated with lower unemployment. This relationship implies when the U.S. increases exports relative to imports more jobs are

Table 4: Standardized coefficients for features in the lasso model based on the one-standard-error rule.

Feature	Coefficient	Description
CES0600000006	-1.42	Production and Nonsupervisory Employees, Goods-Producing
A048RC1A027NBEA	-0.80	Rental income of persons with capital consumption adjustment
CES2000000006	-0.51	Production and Nonsupervisory Employees, Construction
A054RE1A156NBEA	-0.39	Shares of gross domestic income: Corporate profits with inventory valuation and capital consumption adjustments, domestic industries: Taxes on corporate income
A2051C1A027NBEA	0.38	Gross domestic product: Farm products consumed on farms
A253RE1A156NBEA	0.38	Shares of gross domestic product: Exports of goods
A1589C1A027NBEA	0.34	Government social benefits: to persons: Federal: Benefits from social insurance funds: Unemployment insurance
CES1000000006	0.30	Production and Nonsupervisory Employees, Mining and Logging
A019RE1A156NBEA	-0.29	Shares of gross domestic product: Net exports of goods and services
A2122C1A027NBEA	-0.26	Personal saving: Excluding imputations

being created in the United States. However, when the percent of GDP made up by the exportation of goods increases, the number of goods being imported is most likely increasing at a faster rate since the U.S. trade deficit has continued to widen since 1960 [21].

Finally, as personal savings (A2122C1A027NBEA) increase, unemployment declines. When more people have jobs, individuals can save more money.

Elastic Net Regression

The last regression method I used was the elastic net regression, which combines the penalties from ridge and LASSO regression techniques. When alpha = 0, a ridge regression model is produced. When alpha = 1, a LASSO regression is produced. In an elastic net, alpha is between 0 and 1, which creates a ridge-like shrinkage as well as a LASSO-like selection. As long as alpha is greater than 0, shrinkage will occur.

I need to tune the model for both alpha and lambda. To choose alpha, I cross-validated over alpha and lambda. First, I choose alpha to minimize CV error, and then I selected lambda according to the one-standard-error rule. Figure 21 shows CV error for each value of alpha. Alpha is set to 0.343, corresponding to the vertical dashed line on the plot, and lambda selected according to the one-standard-error rule is about 0.00579. Since alpha is closer to 0, the elastic net model is more similar to the ridge regression.

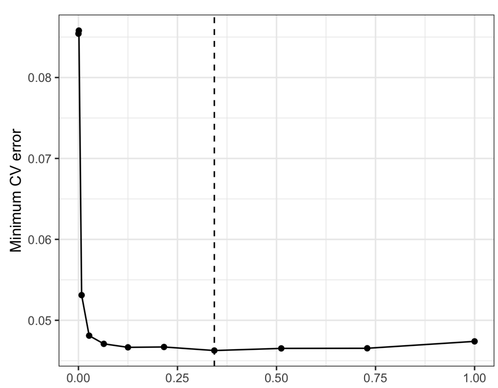


Figure 21: Elastic Net CV plot

Examining the elastic net trace plot in Figure 22, depreciation of fixed assets/ real consumption of fixed

[21] https://fred.stlouisfed.org/series/A019RE1A156NBEA

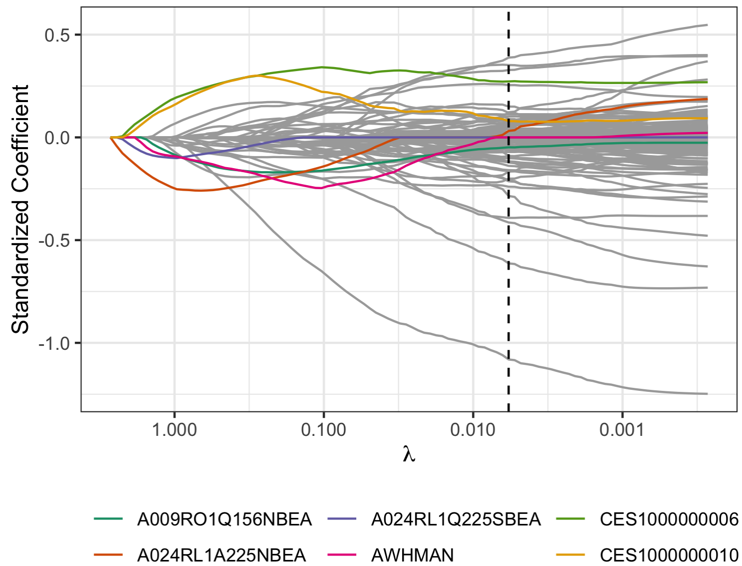


Figure 22: Elastic net trace plot with only first 6 enter the model highlighted

Table 5: Standardized coefficients for features in the elastic net model based on the one-standard-error rule.

Feature	Coefficient	Description
CES0600000006	-1.08	Production and Nonsupervisory Employees, Goods-Producing
CES2000000006	-0.61	Production and Nonsupervisory Employees, Construction
A048RC1A027NBEA	-0.41	Rental income of persons with capital consumption adjustment
A054RE1A156NBEA	-0.39	Shares of gross domestic income: Corporate profits with inventory valuation and capital consumption adjustments, domestic industries: Taxes on corporate income
A253RE1A156NBEA	0.39	Shares of gross domestic product: Exports of goods
A1589C1A027NBEA	0.35	Government social benefits: to persons: Federal: Benefits from social insurance funds: Unemployment insurance
A2051C1A027NBEA	0.33	Gross domestic product: Farm products consumed on farms
A019RE1A156NBEA	-0.29	Shares of gross domestic product: Net exports of goods and services
CES1000000006	0.27	Production and Nonsupervisory Employees, Mining and Logging
A1610C1A027NBEA	0.25	Balance on current account: Adjustment for U.S. territories and Puerto Rico

capital on a yearly and quarterly basis (A024RL1A225NBEA, A024RL1Q225SBE) are two of the first features to enter the model as lambda decreases, and the coefficients for these features are negative.

Similar to the LASSO trace plot, other features with negative standardized coefficents that enter in the first 6 are the private domestic investment in nonresidential structures(A009RO1Q156NBE) and the average weekly of hours of manufacturing production (AWHMAN). Features with positive standardized coefficients include the number of goods-producing (CES1000000006) and female mining/logging employees (CES1000000010). Both of these features were also found in the LASSO trace plot.

Table 4 shows the features with the highest standardized coefficients. For every 1 standard deviation increase in the given feature, there is a “value of the coefficient” standard deviation change in the unemployment rate. Given that alpha is set to 0.343, I would expect similar coefficients to show up in the elastic net as the ridge regression. Seven of the features in the elastic net regression are also present in the ridge regression.

Across all three models, as the number of goods-producing (CES0600000006) and construction employees (CES2000000006) increases, unemployment falls. However, when the number of mining/logging employees (CES1000000006) increases, the unemployment rate rises. In all three models, increases in rental income

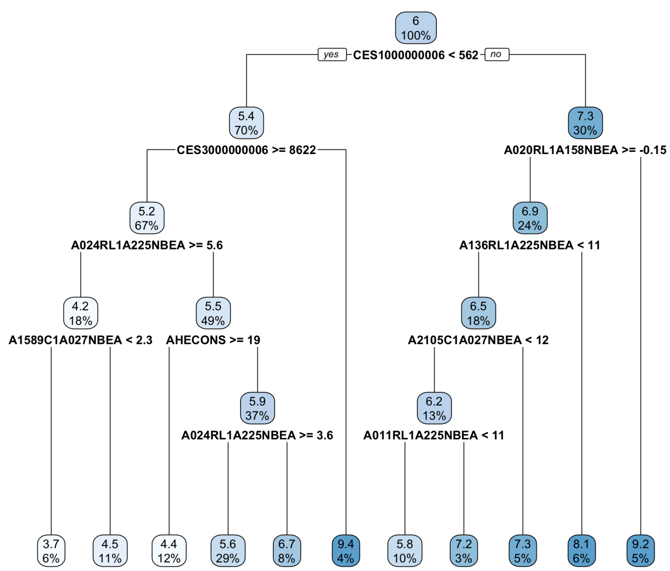
(A048RC1A027NBEA), government social benefits (A1589C1A027NBEA), and farm products consumed on farms (A2051C1A027NBEA) decrease the unemployment rate.

Like the ridge regression, the elastic net finds that as the balance on current accounts adjustment for U.S. territories (A1610C1A027NBEA) increases, the unemployment rate increases. Like LASSO, the elastic net finds that as the corporate income taxes (A054RE1A156NBEA) increase and as the share of GDP made up by the exportation of goods (A019RE1A156NBEA) increases, the unemployment rate increases. However, as the net share of GDP made by the net exportation of goods and services (A253RE1A156NBEA) increases, the unemployment rate falls. This interaction is explained in the LASSO section of this report.

Modeling Class 2: Tree-based Methods

Next, I will predict the U.S. unemployment rate using tree-based methods. Based on my knowledge of unemployment rates, I expect tree-based methods to outperform linear models since fiscal and monetary policy drive unemployment rates.

While I did not use a simple decision tree to predict the unemployment rate, Figure **??** visualizes what a decision tree looks like for the data set. Decision trees partition the feature space into axis-aligned nested rectangles, producing a constant prediction for feature vectors in each rectangle. Behind the scenes, the random forest model and boosted model create trees similar to the one in Figure **??**.



Random Forest

When it comes to prediction accuracy, simple decision trees suffer because of their high variance. Random forests attempt to reduce variance while keeping bias around the same by averaging many trees together. At each split point of each tree, a random subset of features is selected and the tree is split on the best feature among the subset. If the number of features selected at each split is larger, the random forest will have lower bias (since it can better fit the underlying trend) but higher variance (since there are more correlated trees). Therefore, I tuned the random forest-based on the number of variables randomly selected at each split (mtry).

Since there are 158 explanatory variables, mtry can range from 1 to 158. By default, mtry is set to 12, which is the square root of the number of variables. Figure 23 shows the out-of-bag error (OOB error) as a function of mtry. The best value of mtry according to Figure 23 is 17. Thus, I train my model using a mtry value of 17.

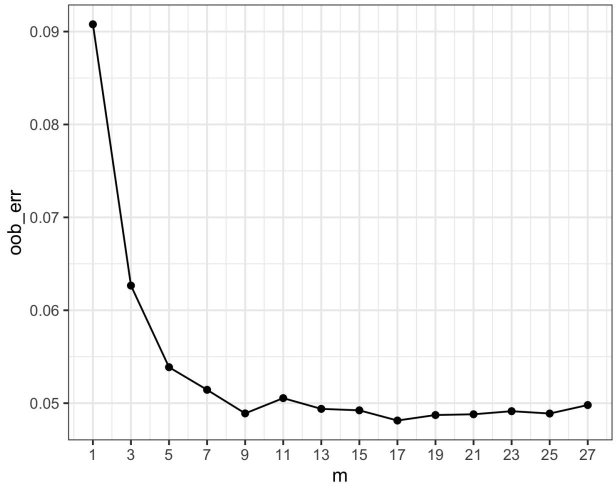


Figure 23: OOB error as a function of mtry

Next, I checked to see at what number of trees OOB stabilizes at. Figure 24 plots OOB error as a function of the number of trees. Figure 24 shows that OBB error stabilizes at around 50 trees, but I trained my random forest model on 300 trees to make sure I had enough.

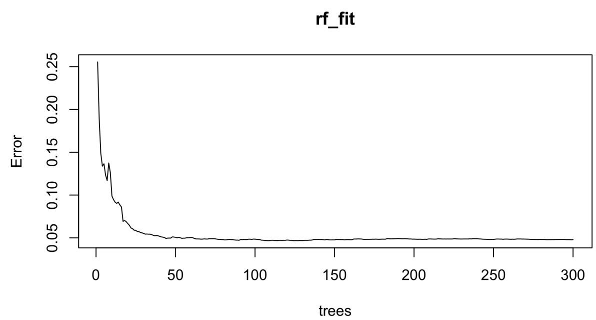


Figure 24: OBB error as a function of the number of trees

The tuned random forest model is trained using 300 trees and 17 random variables sampled from at each split

After training my random forest model, I determined which variables were most important in the model. Compared to regression methods and simple decision trees, the major drawback of random forests is reduced interpretability. However, I used variable importance measures to improve the interpretability of my random forest.

Two types of variable importance measures are used for random forests: > Purity based importance: how much improvement in node purity results from splitting on a feature

> OOB prediction based importance: how much deterioration in prediction accuracy results from scrambling a feature out of bag

Figure 25 shows the features with the top variable importance measures, using both types of measures. For this report, I will only look at OOB prediction-based importance, which is labeled %IncMSE in (Figure 25).

The ten variables with the highest variable importance are in Table 6. However, due to the limitations of the random forest model, how exactly these variables impact the unemployment rate is unknown.

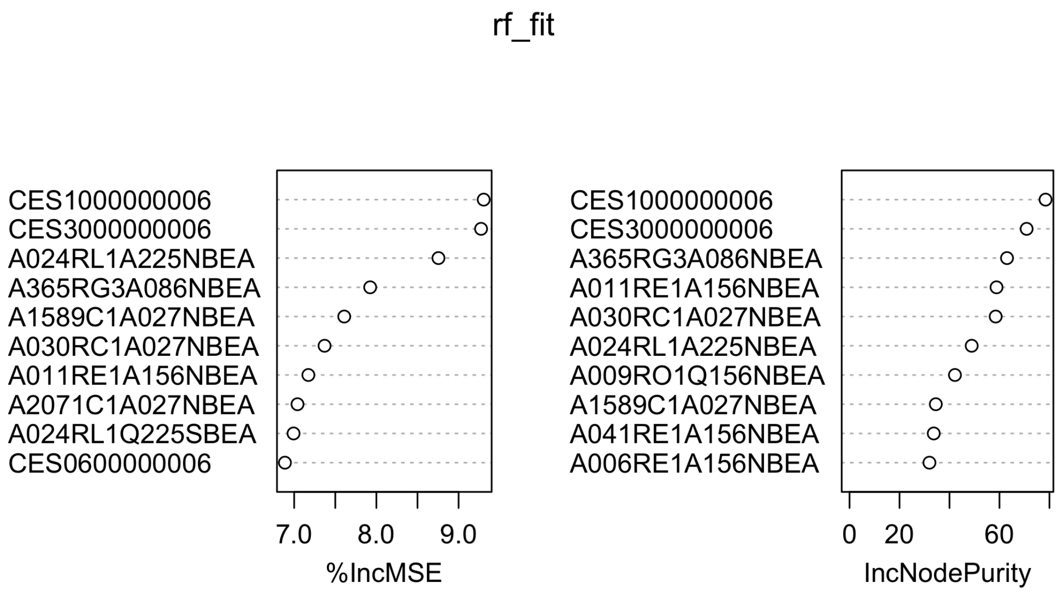


Figure 25: Variable importance for random forest model

Table 6: Top 10 relative importance for random forest model

Feature	Relative Importance	Description
CES1000000006	9.31	Production and Nonsupervisory Employees, Mining and Logging
CES3000000006	9.27	Production and Nonsupervisory Employees, Manufacturing
A024RL1A225NBEA	8.75	Real Consumption of Fixed Capital: Private
A365RG3A086NBEA	7.92	Farm output: Net farm value added (chain-type price index)
A1589C1A027NBEA	7.61	Government social benefits: to persons: Federal: Benefits from social insurance funds: Unemployment insurance
A030RC1A027NBEA	7.37	Net lending or net borrowing (-), NIPAs: Government: Statistical discrepancy
A011RE1A156NBEA	7.17	Shares of gross domestic product: Gross private domestic investment: Fixed investment: Residential
A2071C1A027NBEA	7.04	Monetary interest received: Government
A024RL1Q225SBEA	6.99	Real Consumption of Fixed Capital: Private
CES0600000006	6.89	Production and Nonsupervisory Employees, Goods-Producing

Boosting

While random forests grow deep decision trees in parallel, boosting grows shallow decision trees sequentially. Therefore, I need to consider how deep to grow my trees (interaction depth), how slow the boosting model should learn, and how many trees to grow.

First, I fit boosted tree models with interaction depths of 1, 2, and 3. For each tree model, I used a shrinkage factor of 0.1, 1000 trees, and 5-fold cross-validation. Figure 26 shows the cross-validation error plots for each tree with different interaction depths. The blue line on (Figure 26) has the lowest cross-validation error, so the optimal interaction depth is 3.

Next, I looked for the optimal number of trees to grow. Figure 27 shows that the optimal number of trees to grow is 990. As Figure 27 shows, if I choose a smaller number of trees, error would not increase dramatically. Finally, I left the subsampling fraction equal to 0.5.

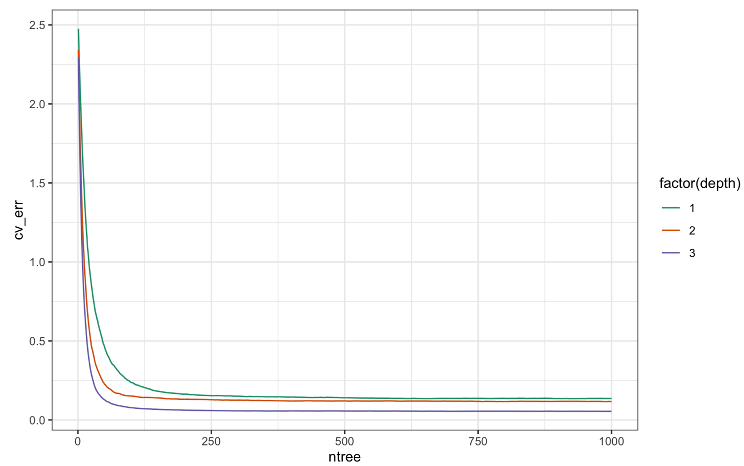


Figure 26: CV error plot for boosted tree model with interaction depth 1, 2, and 3

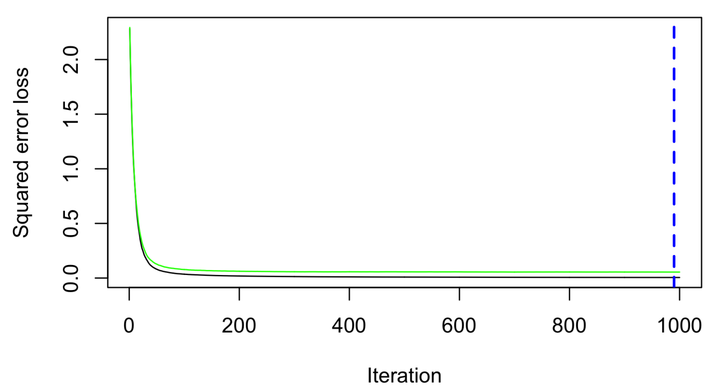


Figure 27: Squard error loss as a function of the number of trees, where blue dashed line represents optimal number of trees

To visualize the effects of the top ten features based on relative influence for the optimal boosting model, I used partial dependence plots (Figure 28).

The percent change in depreciation of fixed assets/consumption of fixed capital (A024RL1A225NBEA) is

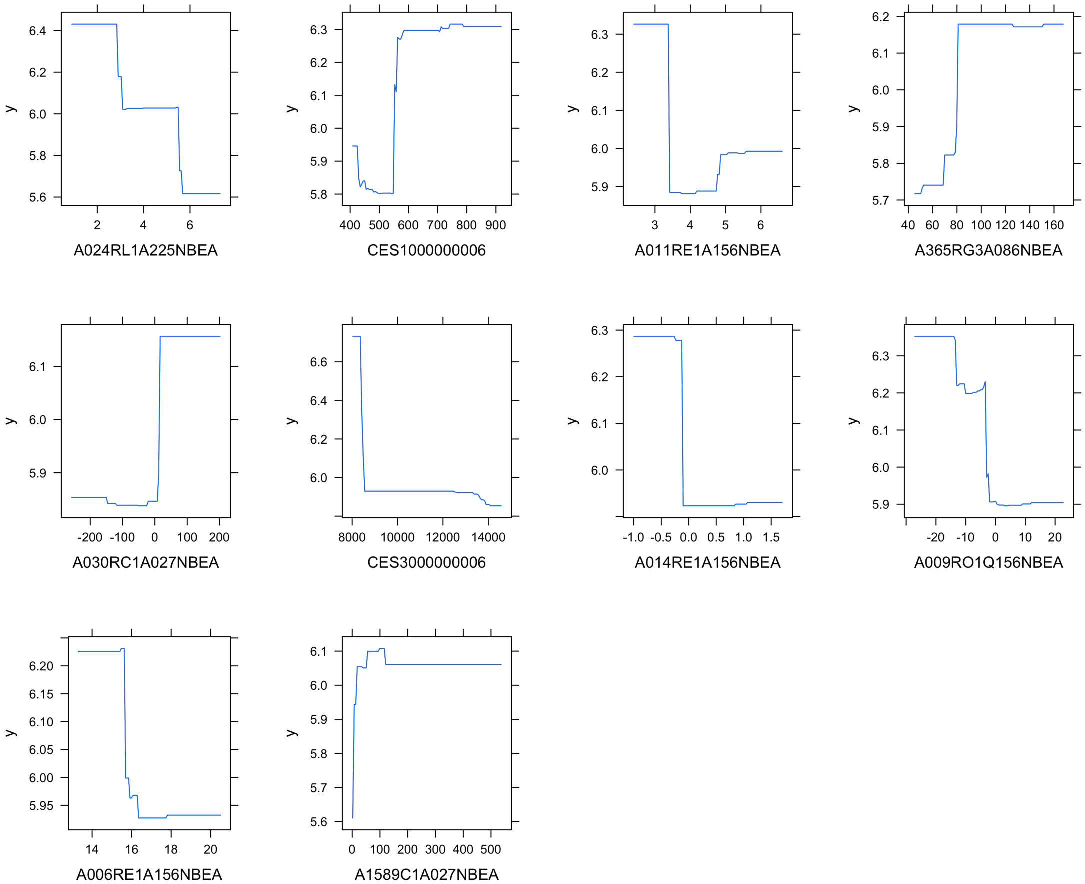


Figure 28: Partial dependence plots for top ten features based on relative influence for optimal boosting model

Table 7: Top 10 relative importance for boosting model, corresponding to partial dependence plots above

Feature	Relative Importance	Description
A024RL1A225NBEA	12.08	Real Consumption of Fixed Capital: Private
CES1000000006	10.70	Production and Nonsupervisory Employees, Mining and Logging
A011RE1A156NBEA	6.69	Shares of gross domestic product: Gross private domestic investment: Fixed investment: Residential
A365RG3A086NBEA	6.53	Farm output: Net farm value added (chain-type price index)
A030RC1A027NBEA	6.06	Net lending or net borrowing (-), NIPAs: Government: Statistical discrepancy
CES3000000006	5.97	Production and Nonsupervisory Employees, Manufacturing
A014RE1A156NBEA	5.41	Shares of gross domestic product: Gross private domestic investment: Change in private inventories
A009RO1Q156NBEA	3.97	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential: Structures
A006RE1A156NBEA	3.06	Shares of gross domestic product: Gross private domestic investment
A1589C1A027NBEA	2.90	Government social benefits: to persons: Federal: Benefits from social insurance funds: Unemployment insurance

the feature with the highest relative importance. As shown in Table 1, this feature also has the highest absolute correlation with the unemployment rate. Consumption of fixed capital is a term used to describe the depreciation of fixed assets. As companies invest more in fixed assets, the deprecition of these assets increases due to a higher depreciation base. As seen in Figure 28, as the consumption of fixed capital increases, the unemployment rate declines. This happens because companies are more likely to invest in buildings and other fixed assets when they expect future growth.

Two of the features describe the number of employees by sector. Unsurprisingly, the number of mining employees (CES1000000006) is the second most important feature. Like the regression methods found, as the number of mining employees increases, the unemployment rate also increases. However, this pattern is only observed after the number of mining employees reaches 550 thousand employees. Before reaching 550 thousand, the unemployment rate falls as the number of mining employees increases. The number of manufacturing employees (CES3000000006) shows up as a significant feature in both the random forest and boosted models but is not found in the regression models. As the number of manufacturing employees increases, the unemployment rate decreases. The unemployment rate levels off after the number of manufacturing employees reaches 8.5 million.

Fixed investment in residential as a percentage of GDP (A011RE1A156NBEA) is found in both the random forest and boosted models. As fixed investment in residential as a percentage of GDP increases, the unemployment rate decreases. This likely happens because real estate developers expect increased demand for housing (apartments, single-family homes, etc.) during good economic times.

Farm output (A365RG3A086NBEA) is also predictive of unemployment rates. As farm output increases, unemployment rate increases. While it is hard to justify why this behavior occurs, it is possible that working on a farm is undesirable when unemployment rates are low. Therefore, operating or working on a farm is more desirable when people cannot find jobs elsewhere. It is also possible that farm output is closely correlated with technological growth. As technology capability increases, fewer workers are needed.

Additionally, as governmental net lending/borrowing (A030RC1A027NBEA) increases, the unemployment rate increases. When unemployment rates are high, the government has to take on more debt to pay for social programs and experiences lower tax revenues. Likewise, the variable importance for unemployment insurance (A1589C1A027NBEA) is also high. As unemployment insurance increases, unemployment rates also increase. The ridge, lasso, elastic net, and random forest models all include unemployment insurance as a feature.

Three factors relate to private domestic investments: change in private inventories (A014RE1A156NBEA), private investment in nonresidential structures (A009RO1Q156NBEA), and gross private direct investment (A006RE1A156NBEA). Change in private inventories describes the increase or decrease in the stocks of final goods, intermediate goods, raw materials, and other inputs that businesses keep on hand to use in production. Figure 28 shows that when the change in private inventories is negative, the unemployment rate is higher than when the change in private inventories is positive. Lower inventories and raw materials on hand likely signal that a company expects lower sales or to perform worse in the future. When a company expects to perform poorly, they are likely to lay off workers. Similarly, as private investment in nonresidential structures and gross private direct investment increases, the unemployment rate falls, since corporations and individuals have more money to invest during periods of economic growth.

Conclusions

Method comparison

To determine which model performed the best, I calculated the test root mean squared error for each model. The lower the RMSE, the better the model is able to fit the data set. To provide a baseline, I also calculated the RMSE of an intercept-only model, where the unemployment rate is predicted to be the mean unemployment rate of the training data set. Table 8 shows the test RMSE for each model. The intercept-only model is the simplest possible model and has high bias and low variance.

Tree-based methods were more predictive than regression methods. I found that the boosting model performs marginally better than the random forest model. The RMSE for the boosted model is 0.882, which means that the square root of the variance of the residuals is 0.882. Thus, on average, the boosting model predicts the unemployment rate to be plus or minus 0.882% of the true unemployment rate for the month.

Looking at regression methods, the least-squares regression performs marginally better than all three penalized regression methods. The least squares, ridge, LASSO, and elastic net regressions all perform better than the intercept-only model. Since the LASSO model performs only slightly worse than the least-squares model, I believe the LASSO model is the best out of all of the regression models. Since it has 80 variables, instead of 158, the LASSO model is easier to interpret.

Table 8: RMSE by Model Type

Model type	Root mean squared error
Intercept Only	1.943
Least Squares	1.129
Ridge	1.295
Lasso	1.210
Elastic Net	1.227
Random Forest	0.882
Boosting	0.860

The boosted model may perform the best because it is better able to capture the underlying trends in the data. It is likely that the underlying trend in unemployment is not linear, since the unemployment rate normally does not dip below the natural rate of unemployment. Also, the Fed's dual mandate helps ensure that unemployment does not exceed a certain percentage. For example, if the unemployment rate gets too high, the Fed will use fiscal policy measures to lower it. Therefore, recursively partitioning the feature space better represents the underlying trends, which decreases bias. Additionally, since boosting aggregates multiple decision trees together, variance is reduced and prediction performance increases compared to a simple decision tree.

Takeaways for Stakeholders

While there were differences in test RMSEs across models, the methods overlap significantly in their identification of important variables from the larger data set. These coefficients also have the same directional effect in all models. Based on economic theory, inflation and the federal funds rate should be predictive of unemployment. However, these features were not in the top 10 features list for any of the models.

After examining which features show up in the top 10 feature list across models, I found that 15 features show up more than twice across the penalized regression models (ridge, LASSO, and elastic net) and the tree-based models (random forest and boosting). These coefficients can be found in Table 9. While I believe these features are the most predictive of the unemployment rate, it is important to note that these features are merely correlated with unemployment, and do not necessarily cause unemployment rates to rise or fall. Now, I will go over my five key takeaways to stakeholders.

> *(1) Changes in the number of employees for "blue-collar" professions are more predictive than changes in the number of employees for"white-collar" professions.*
>
> *(2) Government spending and government debt increases as a result of high unemployment.*
>
> *(3) If the United States exports more goods than it imports, unemployment will fall.*
>
> *(4) As more money is invested in fixed assets, the unemployment rate declines.*
>
> *(5) As corporation profits increase, the unemployment rate falls.*

Table 9: Features Found in Top 10 Across Models

Variable	Count	Ridge	Lasso	Elastic Net	Random Forest	Boosting
A1589C1A027NBEA	5	TRUE	TRUE	TRUE	TRUE	TRUE
CES1000000006	5	TRUE	TRUE	TRUE	TRUE	TRUE
CES0600000006	4	TRUE	TRUE	TRUE	TRUE	FALSE
A048RC1A027NBEA	3	TRUE	TRUE	TRUE	FALSE	FALSE
A2051C1A027NBEA	3	TRUE	TRUE	TRUE	FALSE	FALSE
CES2000000006	3	TRUE	TRUE	TRUE	FALSE	FALSE
A011RE1A156NBEA	2	FALSE	FALSE	FALSE	TRUE	TRUE
A019RE1A156NBEA	2	FALSE	TRUE	TRUE	FALSE	FALSE
A024RL1A225NBEA	2	FALSE	FALSE	FALSE	TRUE	TRUE
A030RC1A027NBEA	2	FALSE	FALSE	FALSE	TRUE	TRUE
A054RE1A156NBEA	2	FALSE	TRUE	TRUE	FALSE	FALSE
A1610C1A027NBEA	2	TRUE	FALSE	TRUE	FALSE	FALSE
A253RE1A156NBEA	2	FALSE	TRUE	TRUE	FALSE	FALSE
A365RG3A086NBEA	2	FALSE	FALSE	FALSE	TRUE	TRUE
CES3000000006	2	FALSE	FALSE	FALSE	TRUE	TRUE

Detailed explanations:

(1) Changes in the number of employees for “blue-collar” professions are more predictive than changes in the number of employees for“white-collar” professions.

Negative effect:

(1) Goods-Producing Employees (CES0600000006)

(2) Construction Employees (CES2000000006)

(3) Manufacturing Employees (CES3000000006)

Positive effect:

(1) Mining and Logging Employees (CES1000000006)

As the number of good-producing, construction, and manufacturing employees increases, the unemployment rate falls. This likely occurs because demand for workers in these industries moves closely with economic cycles. In contrast, the unemployment rate increases as the number of mining and logging employees increases. This is likely because more people are willing to work in mining and logging when it is hard to find a job elsewhere. While “blue-collar” professions are highly predictive of unemployment, “white-collar” professions, such as non-goods producing, information, and service employees, are not as predictive of the unemployment rate.

(2) Government spending and government debt increases as a result of high unemployment

Positive effect:

(1) Unemployment insurance (A1589C1A027NBEA)

(2) Government net lending or net borrowing (A030RC1A027NBEA)

When unemployment insurance increases and when the government borrows more money, the unemployment rate increases. Logically, more people file for unemployment insurance when more individuals are unemployed. Additionally, to pay for these social benefits, the government has to increase spending, which might increase borrowing. Borrowing might also be higher when the unemployment rate is low because the government collects less money in taxes during periods of high unemployment. Firstly, individuals pay fewer income taxes. Second, corporations typically decrease the number of employees they hire when they are performing poorly. Therefore, the government likely collects less corporate taxes during periods of high unemployment.

(3) If the United States exports more goods than it imports, unemployment will fall.

Positive effect

(1) Percent of GDP made up by exportation of goods (A253RE1A156NBEA)

(2) Adjustment to exports for U.S. territories and Puerto Rico (Balance on Current Accounts) (A1610C1A027NBEA)

Negative effect

(1) Percent of GDP made up by net exports of goods and services (A019RE1A156NBEA)

While increased exportation of goods is associated with higher unemployment, the net exportation of goods and services is associated with lower unemployment. The reason for this difference stems from the word "net". "Net" implies the difference between U.S. exports of goods and services and U.S. imports of goods and services, and this metric is positive when exports are greater than imports. Therefore, when the U.S. increases exports relative to imports, this signals that more jobs are being created in the United States. However, when the percent of GDP made up by the exportation of goods increases, the amount of goods being imported is most likely increasing at a faster rate, since the U.S. trade deficit has continued to widen since 1960 [22].

The adjustment to exports for U.S. territories and Puerto Rico is likely found in the model because U.S. territories are not included in the unemployment rate. Therefore, if more goods are exported from territories, fewer goods with be exported from the mainland United States.

(4) As more money is invested in fixed assets, the unemployment rate declines.

Negative effect:

Percent of GDP made up by gross private domestic investment in residential (A011RE1A156NBEA)

Consumption of fixed capital/ depreciation of fixed assets (A024RL1A225NBEA)

Investors and corporations are more likely to invest in fixed / illiquid assets when they believe they are in a growing or steady-state of their business cycle. When money is invested into residential properties, the developer believes the real estate market is growing and the economy is strong. Likewise, depreciation rises when more money as companies' depreciable base increases. Companies only buy more fixed assets when they believe they are likely to expand and be able to support these assets. When companies are doing well, they are likely to hire more workers.

(5) As corporation profits increase, the unemployment rate falls.

Negative effect:

Taxes on corporate income (A054RE1A156NBEA)

Rental income of persons (A048RC1A027NBEA)

When the net income of corporations is higher, they are more likely to hire more employees and not lay off existing ones. Corporations have higher income when they pay higher taxes. Additionally, landlords are likely to receive higher rent when more people have jobs and can afford rent. Thus, when corporate income and rental income increase, the unemployment rate falls.

Potential takeaway: increases in farm output correspond to an increase in unemployment since farm output is linked to technological innovation

Positive effect:

(1) Farm output (A365RG3A086NBEA)

[22] https://fred.stlouisfed.org/series/A1610C1A027NBEA

(2) Farm products consumed on farms (A2051C1A027NBEA)

It is hard to explain why the unemployment rate increases when farm output increases and when the amount of farm products consumed on farms increases. One possible explanation for this is that increases in farm output are associated with increased automation and technological advancements[23]. To be certain this conclusion is correct, I would need to do more analysis.

Limitations

0.0.5 Dataset limitations

There are four data-related limitations of my analysis:

(1) Removed NA values

Originally, my dataset had 2004 variables and 804 observations. However, not all variables have been reported since 1954 and many variables have not yet been reported for 2021. Therefore, I removed all observations before 1960 and after 2020. Also, even though I imputed the values for quarterly and yearly data, I had to remove many columns because many R packages require that no NA values are present in the data. It is possible that variables that started being reported after 1960 are more predictive of the unemployment rate than the variables in my model.

(2) Only focused on two monthly releases

For this report, I focused on two monthly releases: the Unemployment Situation Summary and the Gross Domestic Product Summary. However, other monthly releases may include variables that are more predictive of the unemployment rate. The results of the analysis might change dramatically if I were to incorporate other variables found in other releases. However, due to limited computing power and the fact that the FRED API can only process 120 requests/minute, I was not able to examine more variables.

(3) Observations might not be independent of each other

Additionally, observations are treated as independent of each other. However, it is possible that which features are predictive of the U.S. unemployment rate have shifted over time. For instance, while the number of technology works might not have a major impact on the unemployment rate of months in 1960, it likely has a larger impact on the unemployment rate of 2020. Thus, the variables that are most predictive of the unemployment rate across the entire period may not be as predictive today. Thus, the interpretation of my analysis might need to be taken with a grain of salt.

(4) Dropping features with high multicollinearity

Since many variables had high multicollinearity, I systematically dropped variables with higher than 0.9 correlation with each other. While I kept 1 of the correlated variables, I might have dropped variables that were more predictive of the unemployment rate during this process. However, since the variables dropped are highly correlated with variables left in the analysis, I do not believe this is a major issue.

0.0.6 Analysis limitations

Since I used data from 1960 to 2020 to train my models, I cannot be certain that the relationships identified in my model will still be true in the future. Additionally, while splitting the data into training and testing datasets allows for a more unbiased test of the models, it also adds randomness to my analysis since observations are randomly assigned to either the test or train data set. It is possible that if I split the data again, without using the same seed, the variables selected by each model could have changed. Next, although I trained my models of 158 variables, my analysis incorporates only variables found in two reports plus inflation and the federal funds rate. The model predicts high unemployment rates in 2020 due to economic factors such as net government lending and the number of manufacturing employees. However, high unemployment in 2020 is likely due to health variables not found in my data set.

[23] https://www.asme.org/topics-resources/content/automating-the-risk-out-of-farming

Follow-ups

To compensate for the limitations mentioned above, I would like to use a more up-to-date timeframe. While it is impossible to increase the number of monthly observations for U.S. unemployment rates, I could reproduce my analysis at a county level. To conduct this analysis, I would have to use variables from county-level data sets, instead of data from the Employment Situation or Gross Domestic Product reports. Additionally, if given more time, I would examine which variables were deleted due to missing values. If possible, I would impute these variables for the missing months so that I do not have to delete so many variables. If I had more time, I would also examine my assumption that monthly unemployment rates are independent of each other. To do so, I would examine patterns in the explanatory variables across time and flag any features that appear to have dramatically changed with time. Then, I would either remove these variables or index them. Conducting this analysis would make my conclusions more robust.

A Appendix: Descriptions of features

id	title	frequency	units
AHECONS	Average Hourly Earnings of Production and Nonsupervisory Employees, Construction	Monthly	Dollars per Hour
AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees, Manufacturing	Monthly	Hours
AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees, Manufacturing	Monthly	Hours
CES0600000006	Production and Nonsupervisory Employees, Goods-Producing	Monthly	Thousands of Persons
CES1000000006	Production and Nonsupervisory Employees, Mining and Logging	Monthly	Thousands of Persons
CES1000000007	Average Weekly Hours of Production and Nonsupervisory Employees, Mining and Logging	Monthly	Hours
CES1000000010	Women Employees, Mining and Logging	Monthly	Thousands of Persons
CES1000000039	Women Employees-To-All Employees Ratio: Mining and Logging	Monthly	Percent
CES1011330001	All Employees, Logging	Monthly	Thousands of Persons
CES2000000006	Production and Nonsupervisory Employees, Construction	Monthly	Thousands of Persons
CES2000000007	Average Weekly Hours of Production and Nonsupervisory Employees, Construction	Monthly	Hours
CES3000000006	Production and Nonsupervisory Employees, Manufacturing	Monthly	Thousands of Persons
CES3000000010	Women Employees, Manufacturing	Monthly	Thousands of Persons
CES3000000039	Women Employees-To-All Employees Ratio: Manufacturing	Monthly	Percent
CES3133660001	All Employees, Ship and Boat Building	Monthly	Thousands of Persons
CES9091000001	All Employees, Federal	Monthly	Thousands of Persons
CES9091100001	All Employees, Federal, Except U.S. Postal Service	Monthly	Thousands of Persons
CES9091912001	All Employees, U.S. Postal Service	Monthly	Thousands of Persons
CEU0600000007	Average Weekly Hours of Production and Nonsupervisory Employees, Goods-Producing	Monthly	Hours
CEU2000000007	Average Weekly Hours of Production and Nonsupervisory Employees, Construction	Monthly	Hours
CEU3200000007	Average Weekly Hours of Production and Nonsupervisory Employees, Nondurable Goods	Monthly	Hours
CEU5000000001	All Employees, Information	Monthly	Thousands of Persons
A001RI1A225NBEA	Gross National Product: Implicit Price Deflator	Annual	Percent Change from Preceding Period
A001RL1A225NBEA	Real Gross National Product	Annual	Percent Change from Preceding Period
A001RL1Q225SBEA	Real Gross National Product	Quarterly	Percent Change from Preceding Period
A001RO1Q156NBEA	Real Gross National Product	Quarterly	Percent Change from Quarter One Year Ago
A001RP1A027NBEA	Gross National Product	Annual	Percent Change from Preceding Period
A001RP1Q027SBEA	Gross National Product	Quarterly	Percent Change from Preceding Period
A006RE1A156NBEA	Shares of gross domestic product: Gross private domestic investment	Annual	Percent
A006RJ2Q224SBEA	Contributions to percent change in gross domestic product price index: Gross private domestic investment	Quarterly	Percentage Points at Annual Rate
A006RL1A225NBEA	Real Gross Private Domestic Investment	Annual	Percent Change from Preceding Period
A006RL1Q225SBEA	Real Gross Private Domestic Investment	Quarterly	Percent Change from Preceding Period
A007RL1Q225SBEA	Real Gross Private Domestic Investment: Fixed Investment	Quarterly	Percent Change from Preceding Period
A007RO1Q156NBEA	Real Gross Private Domestic Investment: Fixed Investment	Quarterly	Percent Change from Quarter One Year Ago
A008RE1A156NBEA	Shares of gross domestic product: Gross private domestic investment: Fixed investment: Nonresidential	Annual	Percent
A008RL1A225NBEA	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential	Annual	Percent Change from Preceding Period
A008RL1Q225SBEA	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential	Quarterly	Percent Change from Preceding Period

id	title	frequency	units
A009RL1A225NBEA	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential: Structures	Annual	Percent Change from Preceding Period
A009RL1Q225SBEA	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential: Structures	Quarterly	Percent Change from Preceding Period
A009RO1Q156NBEA	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential: Structures	Quarterly	Percent Change from Quarter One Year Ago
A009RS2A224NBEA	Contributions to percent change in gross domestic purchases: Gross private domestic investment: Fixed investment: Nonresidential: Structures	Annual	Percentage Points at Annual Rate
A009RS2Q224SBEA	Contributions to percent change in gross domestic purchases: Gross private domestic investment: Fixed investment: Nonresidential: Structures	Quarterly	Percentage Points at Annual Rate
A011RE1A156NBEA	Shares of gross domestic product: Gross private domestic investment: Fixed investment: Residential	Annual	Percent
A011RJ2A224NBEA	Contributions to percent change in gross domestic product price index: Gross private domestic investment: Fixed investment: Residential	Annual	Percentage Points at Annual Rate
A011RJ2Q224SBEA	Contributions to percent change in gross domestic product price index: Gross private domestic investment: Fixed investment: Residential	Quarterly	Percentage Points at Annual Rate
A011RL1A225NBEA	Real Gross Private Domestic Investment: Fixed Investment: Residential	Annual	Percent Change from Preceding Period
A011RL1Q225SBEA	Real Gross Private Domestic Investment: Fixed Investment: Residential	Quarterly	Percent Change from Preceding Period
A014RE1A156NBEA	Shares of gross domestic product: Gross private domestic investment: Change in private inventories	Annual	Percent
A014RE1Q156NBEA	Shares of gross domestic product: Gross private domestic investment: Change in private inventories	Quarterly	Percent
A014RJ2A224NBEA	Contributions to percent change in gross domestic product price index: Gross private domestic investment: Change in private inventories	Annual	Percentage Points at Annual Rate
A014RJ2Q224SBEA	Contributions to percent change in gross domestic product price index: Gross private domestic investment: Change in private inventories	Quarterly	Percentage Points at Annual Rate
A014RY2A224NBEA	Contributions to percent change in real gross domestic product: Gross private domestic investment: Change in private inventories	Annual	Percentage Points at Annual Rate
A014RY2Q224SBEA	Contributions to percent change in real gross domestic product: Gross private domestic investment: Change in private inventories	Quarterly	Percentage Points at Annual Rate
A015RC1A027NBEA	Change in private inventories: Nonfarm	Annual	Billions of Dollars
A015RC1Q027SBEA	Change in private inventories: Nonfarm	Quarterly	Billions of Dollars
A015RS2A224NBEA	Contributions to percent change in gross domestic purchases: Gross private domestic investment: Change in private inventories: Nonfarm	Annual	Percentage Points at Annual Rate
A015RS2Q224SBEA	Contributions to percent change in gross domestic purchases: Gross private domestic investment: Change in private inventories: Nonfarm	Quarterly	Percentage Points at Annual Rate
A019RC1A027NBEA	Net exports of goods and services	Annual	Billions of Dollars
A019RE1A156NBEA	Shares of gross domestic product: Net exports of goods and services	Annual	Percent
A019RJ2A224NBEA	Contributions to percent change in gross domestic product price index: Net exports of goods and services	Annual	Percentage Points at Annual Rate
A019RJ2Q224SBEA	Contributions to percent change in gross domestic product price index: Net exports of goods and services	Quarterly	Percentage Points at Annual Rate
A019RY2A224NBEA	Contributions to percent change in real gross domestic product: Net exports of goods and services	Annual	Percentage Points at Annual Rate
A019RY2Q224SBEA	Contributions to percent change in real gross domestic product: Net exports of goods and services	Quarterly	Percentage Points at Annual Rate
A020RJ2A224NBEA	Contributions to percent change in gross domestic product price index: Exports of goods and services	Annual	Percentage Points at Annual Rate
A020RJ2Q224SBEA	Contributions to percent change in gross domestic product price index: Exports of goods and services	Quarterly	Percentage Points at Annual Rate
A020RL1A158NBEA	Real Exports of Goods and Services	Annual	Percent Change from Preceding Period
A020RL1Q158SBEA	Real Exports of Goods and Services	Quarterly	Percent Change from Preceding Period
A020RO1Q156NBEA	Real Exports of Goods and Services	Quarterly	Percent Change from Quarter One Year Ago
A021RL1A158NBEA	Real Imports of Goods and Services	Annual	Percent Change from Preceding Period
A021RL1Q158SBEA	Real Imports of Goods and Services	Quarterly	Percent Change from Preceding Period

id	title	frequency	units
A021RO1Q156NBEA	Real Imports of Goods and Services	Quarterly	Percent Change from Quarter One Year Ago
A021RV1Q225SBEA	Imports of Goods and Services (chain-type price index)	Quarterly	Percent Change from Preceding Period
A021RY2A224NBEA	Contributions to percent change in real gross domestic product: Imports of goods and services	Annual	Percentage Points at Annual Rate
A021RY2Q224SBEA	Contributions to percent change in real gross domestic product: Imports of goods and services	Quarterly	Percentage Points at Annual Rate
A023RL1Q225SBEA	Real Gross National Income	Quarterly	Percent Change from Preceding Period
A024RL1A225NBEA	Real Consumption of Fixed Capital: Private	Annual	Percent Change from Preceding Period
A024RL1Q225SBEA	Real Consumption of Fixed Capital: Private	Quarterly	Percent Change from Preceding Period
A030RC1A027NBEA	Net lending or net borrowing (-), NIPAs: Government: Statistical discrepancy	Annual	Billions of Dollars
A030RC1Q027SBEA	Net lending or net borrowing (-), NIPAs: Government: Statistical discrepancy	Quarterly	Billions of Dollars
A041RE1A156NBEA	Shares of gross domestic income: Net operating surplus: Private enterprises: Proprietors' income with inventory valuation and capital consumption adjustments	Annual	Percent
A043RC1A027NBEA	National income: Proprietors' income with IVA: Farm	Annual	Billions of Dollars
A048RC1A027NBEA	Rental income of persons with capital consumption adjustment	Annual	Billions of Dollars
A048RE1A156NBEA	Shares of gross domestic income: Net operating surplus: Private enterprises: Rental income of persons with capital consumption adjustment	Annual	Percent
A054RE1A156NBEA	Shares of gross domestic income: Corporate profits with inventory valuation and capital consumption adjustments, domestic industries: Taxes on corporate income	Annual	Percent
A059RC1A027NBEA	Net private saving: Domestic business: Capital consumption adjustment, corporate	Annual	Billions of Dollars
A067RL1A156NBEA	Real Disposable Personal Income	Annual	Percent Change from Preceding Period
A067RL1Q156SBEA	Real Disposable Personal Income	Quarterly	Percent Change from Preceding Period
A067RO1Q156NBEA	Real Disposable Personal Income	Quarterly	Percent Change from Quarter One Year Ago
A067RP1A027NBEA	Disposable Personal Income	Annual	Percent Change from Preceding Period
A067RP1Q027SBEA	Disposable Personal Income	Quarterly	Percent Change from Preceding Period
A071RC1A027NBEA	Personal saving	Annual	Billions of Dollars
A072RC1A156NBEA	Personal saving as a percentage of disposable personal income	Annual	Percent
A107RE1A156NBEA	Shares of gross domestic income: Subsidies	Annual	Percent
A108RC1A027NBEA	Current surplus of government enterprises	Annual	Billions of Dollars
A108RE1A156NBEA	Shares of gross domestic income: Current surplus of government enterprises	Annual	Percent
A1100C1A027NBEA	Monetary interest paid: Domestic: Corporate business: Financial	Annual	Billions of Dollars
A127RC1Q027SBEA	Net private saving: Domestic business	Quarterly	Billions of Dollars
A133RA3A086NBEA	Real automobile output (chain-type quantity index)	Annual	Index 2012=100
A133RL1A225NBEA	Real Motor Vehicle Output: Auto Output	Annual	Percent Change from Preceding Period
A133RL1Q225SBEA	Real Motor Vehicle Output: Auto Output	Quarterly	Percent Change from Preceding Period

id	title	frequency	units
A136RL1A225NBEA	Real Motor Vehicle Output: Final Sales of Domestic Product: Personal Consumption Expenditures: New Motor Vehicles: Autos	Annual	Percent Change from Preceding Period
A136RL1Q225SBEA	Real Motor Vehicle Output: Final Sales of Domestic Product: Personal Consumption Expenditures: New Motor Vehicles: Autos	Quarterly	Percent Change from Preceding Period
A1400C1A027NBEA	Capital consumption adjustment: for consistent accounting at historical cost	Annual	Billions of Dollars
A1402C1A027NBEA	Capital consumption adjustment: Domestic corporate business: for consistent accounting at historical cost	Annual	Billions of Dollars
A145RC1A027NBEA	Automobile output: Change in private inventories of new and used autos	Annual	Billions of Dollars
A145RC1Q027SBEA	Automobile output: Change in private inventories of new and used autos	Quarterly	Billions of Dollars
A1589C1A027NBEA	Government social benefits: to persons: Federal: Benefits from social insurance funds: Unemployment insurance	Annual	Billions of Dollars
A1608C1A027NBEA	Balance on goods and services and income: Adjustment for U.S. territories and Puerto Rico	Annual	Billions of Dollars
A1610C1A027NBEA	Balance on current account: Adjustment for U.S. territories and Puerto Rico	Annual	Billions of Dollars
A1830C0A144NBEA	Inventory valuation adjustment to nonfarm incomes	Annual	Millions of Dollars
A1870C1A027NBEA	Business current transfer payments (net): Insurance payments to persons by business	Annual	Billions of Dollars
A190RL1Q225SBEA	Real Final Sales of Domestic Product	Quarterly	Percent Change from Preceding Period
A190RP1Q027SBEA	Final Sales of Domestic Product	Quarterly	Percent Change from Preceding Period
A193RL1A225NBEA	Real Gross Value Added: Gross Domestic Product: Households and Institutions	Annual	Percent Change from Preceding Period
A193RL1Q225SBEA	Real Gross Value Added: Gross Domestic Product: Households and Institutions	Quarterly	Percent Change from Preceding Period
A194RL1A225NBEA	Real Gross Output of General Government: Value Added: Compensation of General Government Employees	Annual	Percent Change from Preceding Period
A194RL1Q225SBEA	Real Gross Output of General Government: Value Added: Compensation of General Government Employees	Quarterly	Percent Change from Preceding Period
A2009L1A225NBEA	Real Gross Housing Value Added	Annual	Percent Change from Preceding Period
A2009L1Q225SBEA	Real Gross Housing Value Added	Quarterly	Percent Change from Preceding Period
A2010C1A027NBEA	Net government saving: Imputations	Annual	Billions of Dollars
A2011C1A027NBEA	Net government saving: Excluding imputations	Annual	Billions of Dollars
A2033C1A027NBEA	Balance on goods and services and income: Statistical differences , International Transactions Accounts vs. NIPAs	Annual	Billions of Dollars
A2036C1A027NBEA	Balance on current account: Statistical differences, International Transactions Accounts vs. NIPAs	Annual	Billions of Dollars
A2051C1A027NBEA	Gross domestic product: Farm products consumed on farms	Annual	Billions of Dollars
A2071C1A027NBEA	Monetary interest received: Government	Annual	Billions of Dollars
A2101C1A027NBEA	Gross domestic income: Proprietors' income with inventory valuation and capital consumption adjustments: Imputations	Annual	Billions of Dollars
A2105C1A027NBEA	Gross domestic income: Net interest and miscellaneous payments: Excluding imputations	Annual	Billions of Dollars
A2122C1A027NBEA	Personal saving: Excluding imputations	Annual	Billions of Dollars
A2224C1A027NBEA	Net corporate dividend payments: Rest of the world	Annual	Billions of Dollars
A253RD3A086NBEA	Exports of goods (implicit price deflator)	Annual	Index 2012=100
A253RE1A156NBEA	Shares of gross domestic product: Exports of goods	Annual	Percent
A253RY2Q224SBEA	Contributions to percent change in real gross domestic product: Exports of goods	Quarterly	Percentage Points at Annual Rate
A2641C1A027NBEA	Employment-related imputations: Employees' lodging	Annual	Billions of Dollars
A264RL1A225NBEA	Real Consumption of Fixed Capital: Government	Annual	Percent Change from Preceding Period
A325RL1A225NBEA	Real Gross Domestic Product: Goods: Final Sales	Annual	Percent Change from Preceding Period

id	title	frequency	units
A325RL1Q225SBEA	Real Gross Domestic Product: Goods: Final Sales	Quarterly	Percent Change from Preceding Period
A334RL1A225NBEA	Real Gross Domestic Product: Goods: Nondurable Goods: Final Sales	Annual	Percent Change from Preceding Period
A334RL1Q225SBEA	Real Gross Domestic Product: Goods: Nondurable Goods: Final Sales	Quarterly	Percent Change from Preceding Period
A334RY2A224NBEA	Contributions to percent change in real gross domestic product: Nondurable goods: Final sales	Annual	Percentage Points at Annual Rate
A3402C0A144NBEA	Undistributed corporate profits: Domestic industries	Annual	Millions of Dollars
A341RL1A225NBEA	Real Gross Domestic Product: Services	Annual	Percent Change from Preceding Period
A341RL1Q225SBEA	Real Gross Domestic Product: Services	Quarterly	Percent Change from Preceding Period
A3474C0A144NBEA	Undistributed corporate profits: Rest of the world	Annual	Millions of Dollars
A349RL1A225NBEA	Real Private Fixed Investment: Private Fixed Investment in Structures	Annual	Percent Change from Preceding Period
A349RL1Q225SBEA	Real Private Fixed Investment: Private Fixed Investment in Structures	Quarterly	Percent Change from Preceding Period
A354RL1A225NBEA	Real Gross Domestic Product: Goods: Durable Goods	Annual	Percent Change from Preceding Period
A354RL1Q225SBEA	Real Gross Domestic Product: Goods: Durable Goods	Quarterly	Percent Change from Preceding Period
A355RY2Q224SBEA	Contributions to percent change in real gross domestic product: Durable goods: Change in private inventories	Quarterly	Percentage Points at Annual Rate
A356RL1A225NBEA	Real Gross Domestic Product: Goods: Nondurable Goods	Annual	Percent Change from Preceding Period
A356RL1Q225SBEA	Real Gross Domestic Product: Goods: Nondurable Goods	Quarterly	Percent Change from Preceding Period
A357RC1A027NBEA	Gross domestic product: Nondurable goods: Change in private inventories	Annual	Billions of Dollars
A357RC1Q027SBEA	Gross domestic product: Nondurable goods: Change in private inventories	Quarterly	Billions of Dollars
A357RY2A224NBEA	Contributions to percent change in real gross domestic product: Nondurable goods: Change in private inventories	Annual	Percentage Points at Annual Rate
A357RY2Q224SBEA	Contributions to percent change in real gross domestic product: Nondurable goods: Change in private inventories	Quarterly	Percentage Points at Annual Rate
A365RG3A086NBEA	Farm output: Net farm value added (chain-type price index)	Annual	Index 2009=100
A4002E1A156NBEA	Shares of gross domestic income: Compensation of employees, paid	Annual	Percent
FEDFUNDS	Federal Funds Effective Rate	Monthly	Percent
FPCPITOTLZGUSA	Inflation, consumer prices for the United States	Yearly	Percent
UNRATE	Unemployment Rate.	Monthly	Percent