UC San Diego

ECE-225A Final Project

# Analysis: Unemployment Rate in the United States, 1990-2016

Time-Series Approach

**Arth Dharaskar**

**Aaron Hanna**

*Electrical and Computer Engineering, UC San Diego*

December 21, 2020

# Contents

**1 Abstract** **2**

**2 Introduction** **2**

**3 Data** **2**

**4 Materials and Methods** **3**
4.1 Time Series Forecasting . . . . . 3
4.2 Geo Plotting . . . . . 3

**5 Results** **3**
5.1 Predictor Performance . . . . . 4
5.2 Political Party Correlation . . . . . 5
5.3 Historical Events . . . . . 5
5.4 Individual Counties . . . . . 6

**6 Conclusion** **7**

**References** **8**

# 1 Abstract

This article uses linear time series data of the United States Unemployment Rate by County between 1990 and 2016. A time series model is used to forecast unemployment by state or county. Special attention is given to large events such as the housing market bubble and recession with a careful look at overall impact vs impact by county. In addition, correlations between State and County political party and unemployment rate are analyzed.

# 2 Introduction

During the pandemic, Unemployment rate has seen a large increase due to many factors and affecting people and businesses all around the world. It is especially important given the political and economic climate to analyze the US Unemployment to observe any common trends. People and businesses in the US are more vulnerable than ever now and we expect to see similar trends in major events such as the Housing Market Crash as we are seeing now. We suspect that different counties will be more or less vulnerable given their geographic location and political affiliation. There are many other socio-economic factors that contribute to and affect the Unemployment Rate and health of the economy, such as Federal Taxes Paid vs. Taxes Received [1], but we will only analyze Unemployment Rate and political affiliation.

We seek to address questions such as:

- Is there a correlation between unemployment rate and county? Positive? Negative?
- Is there a trend to be seen over time that is independent of location?
- What events, if any, have a significant impact on unemployment rate in certain counties?
- Can a model be generated that can accurately predict unemployment rate by county?

# 3 Data

In this analysis, we observe a kaggle dataset [3] containing Unemployment Rates in the US for counties excluding in Hawaii, Louisiana, Georgia, Florida and Alaska. The dataset doesn't have consistent data for all states and often times is missing a states data for a given month/year. The feature we are analyzing is time series of unemployment rate in the US. Data is represented in the format as seen in Figure 1 after cleaning and the parameters can be seen in Figure 2. The data we have is also available from the US Department of Labor Statistics (BLS).

| Date | State | County | Rate |
|---|---|---|---|
| 1990-01-01 | New York | Wyoming County | 8.0 |
| 1990-01-01 | New York | Chenango County | 6.5 |
| 1990-01-01 | New York | Westchester County | 3.5 |
| 1990-01-01 | New York | Otsego County | 6.8 |
| 1990-01-01 | New York | New York County | 6.6 |

**Figure 1:** Dataset Sample

| | Year | Rate |
|---|---|---|
| count | 885548.000000 | 885548.000000 |
| mean | 2003.000017 | 6.175010 |
| std | 7.824893 | 3.112535 |
| min | 1990.000000 | 0.000000 |
| 25% | 1996.000000 | 4.000000 |
| 50% | 2003.000000 | 5.500000 |
| 75% | 2010.000000 | 7.700000 |
| max | 2016.000000 | 58.400000 |

**Figure 2:** Dataset Parameters

# 4 Materials and Methods

## 4.1 Time Series Forecasting

Our dataset has a single variable and no additional features to create a model hence we decided to use Univariate Time Series forecasting using ARIMA(Autoregressive Integrated Moving Average) model. An ARIMA model uses lag variables and moving averages as its predictors:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_1 Y_{t-1} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

Here, Alpha is a constant, while Beta and Phi are the scaling factors for the Lag variables and Moving Averages respectively.

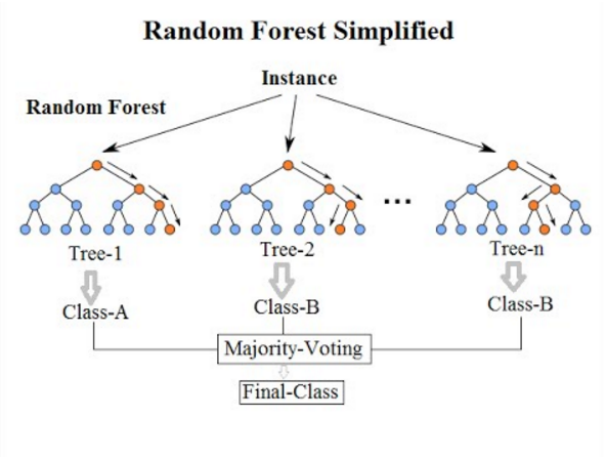


**Figure 3:** Random Forest Simplified [2]

We decided to use four lag variables and three rolling averages as features for the Random Regressor model which uses Bootstrap Aggregation ensemble clusters of Decision tree classifiers to make the required prediction.

While it is typical to use K-fold cross validation, given that our data is time series based, we cannot use that CV method as the estimate would not be representative of the training. Future random samples can leak data into the model for predictions.

Instead, we use a method known as the Time Series Split method provided by Scikit-learn.

This allows a non-overlapping, sliding window approach which return the splits for the dataset. We calculate cross validation score based on the *Negative Mean Absolute Error* and plot our **95%** confidence interval superimposed on the plots of predictions and original trend-lines.

## 4.2 Geo Plotting

To geographically visualize the unemployment rate in the US over time we used a combination of Geopandas, Bokeh and IPython in Juptyer Notebook. The geopandas library helps to organize the pandas dataframe to have geometry by using a .shp file of the polygons representing each county.

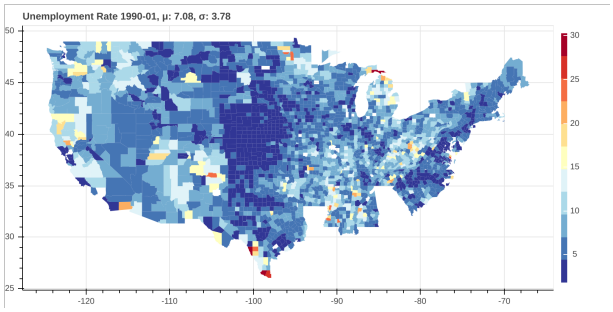


**Figure 4:** US Map Rates Jan 1990

The visualization portion was made easier with the use of Bokeh which uses an inline java feature and has the capability of producing html, png and servers for viewing. IPython was used to maneuver over time to look at significant dates or play an animation using the widgets capability.

# 5 Results

We begin by plotting the time series for New York, New York County as seen in Figure 5. We can already see a periodic pattern of roughly 4 years and centered slightly higher then the mean unemployment rate.

When plotting the autocorrelation function (ACF) of the same county we see that it in fact converges to zero which indicates the unemployment rates are roughly independent (Figure 5).

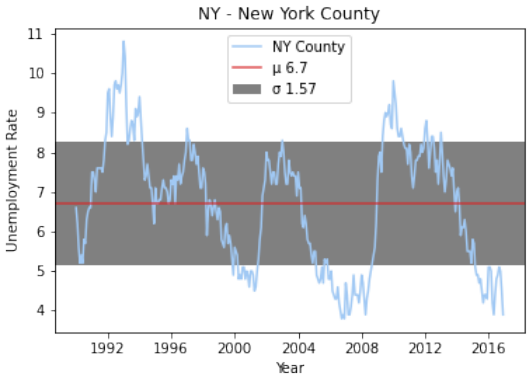


**Figure 5:** New York County Rate

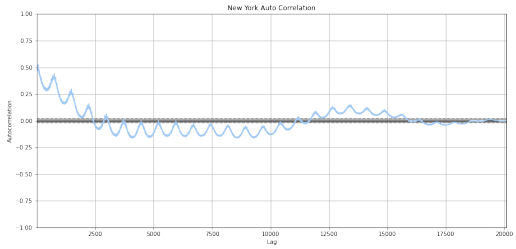


**Figure 6:** New York Auto Correlation

Given the large amount of variability of the economy and the Unemployment Rate in the US over time, it is surprising to see a periodic trend. In Figure 7 we plot the rolling average of New York county. The rolling average helps to smooth out short term changes over time and help visualize long term trends in financial time series datasets. We chose a window size of 10 months to better approximate the moving average and reduce overlap.

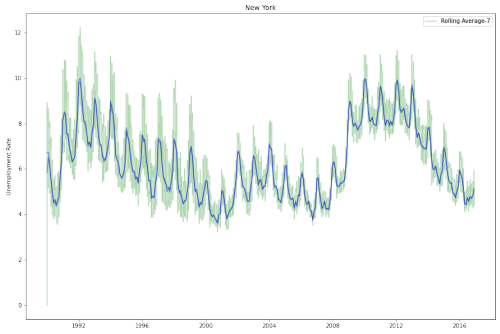


**Figure 7:** New York Rolling Average

## 5.1 Predictor Performance

In Figure 8 we have the prediction and actual New York county unemployment rate plotted together with a 95% confidence interval overlaid. Similarly, we see the Imperial County prediction on the right. We notice that we are able to get better bounds on New York in terms of variance while Imperial County had a higher variance in predictions, we discuss these results in the conclusion section. While there is high variance we are able to model the general trend lines well as seen in the plots.

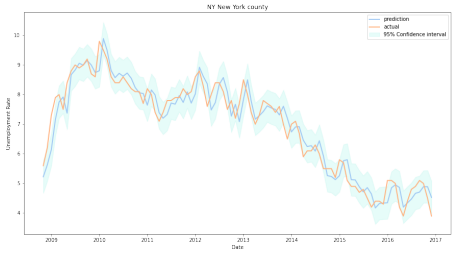


**Figure 8:** NY County Model

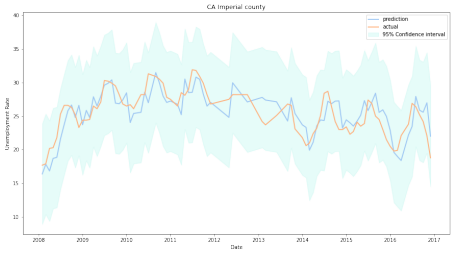


**Figure 9:** Imperial County Model

## 5.2 Political Party Correlation

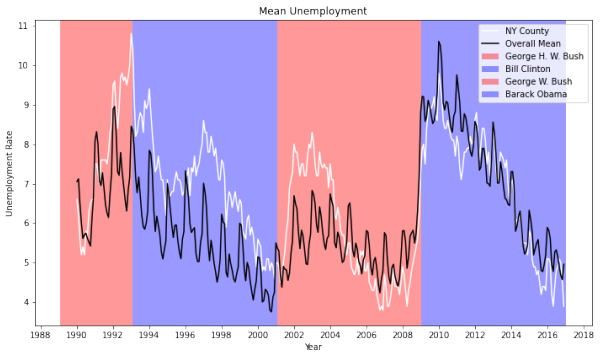


**Figure 10:** Party Affiliation New York County

Political party representation in the US is color coded by blue for Democrat and red for Republican. We visualize this with an overlay of the presidential terms in the US with the name of the President and party affiliation.

To begin we take another look at New York county which is a historically blue state. In Figure 9 we observe that during the Republican(red) terms the Unemployment Rate in New York county increases. Consequently, we see the opposite is true for the Democratic terms and the Unemployment rate decreases. The anomaly here can be seen during George W. Bush's term where the unemployment rate rises and falls within the term. However, the rate increases dramatically from 2008 to 2009 at the end of his term during the Housing Market bubble[4].

Unfortunately, this is a relatively small sample and furthermore we do not know the influence the federal elections have compared to state and municipal elections. Irregardless, we can derive their is a weak positive correlation between Republican Presidencies and a weak negative correlation between Democratic Presidencies.

## 5.3 Historical Events

The dot com and housing market bubble crash are significant economic events in US history occurring roughly around October 2002 and September 2008, respectively. We can see in Figure 11 that the unemployment rate has a downward trend over time and that the dot com bubble seems to have zero correlation. Consequently, the housing market bubble had a strong positive correlation with unemployment rate.

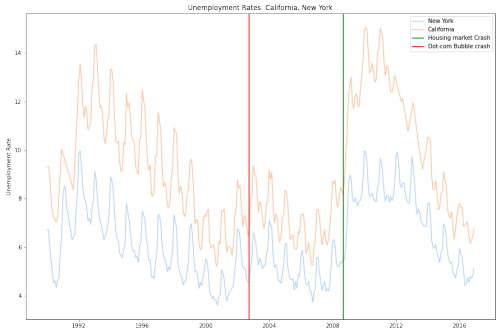


**Figure 11:** NY, CA: Dot Com and Housing

The unemployment rates in the US during the dot com can be seen in Figure 11 where the majority of the counties are below 10% unemployment. There is little variance overall with the exception of Imperial County - CA (16.9), Yuma County - AZ (17), Presidio and Reeves County - Texas (15.9, 16.9).

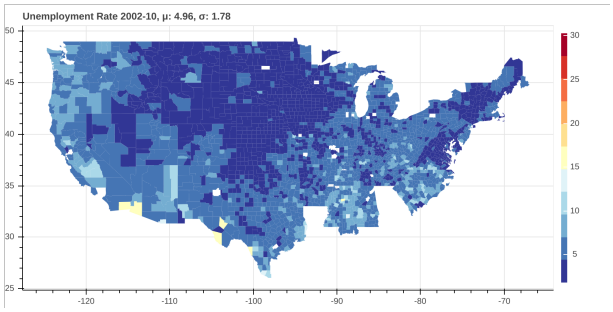


**Figure 12:** Dot Com Bubble

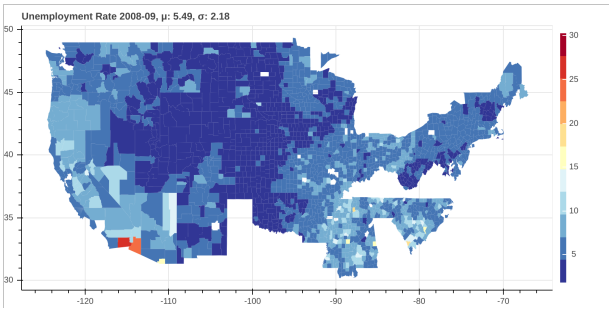


**Figure 13:** Housing Market Bubble

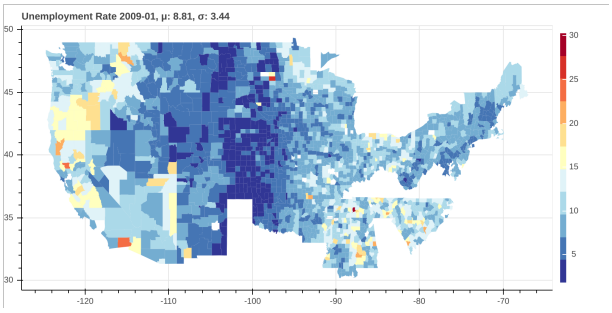


**Figure 14:** January 2009 Map

## 5.4 Individual Counties

Here we compare two counties Yuma, AZ and Imperial, CA to New York, NY. Imperial and Yuma are two counties that we noticed were consistently high in our graphs.

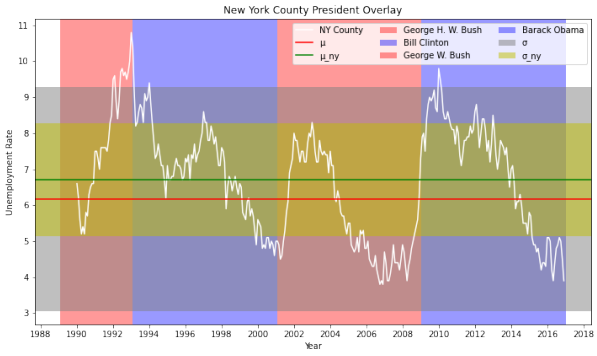


**Figure 15:** New York County

Here we see New York County is centered about the dataset mean and is roughly within one standard deviation of the dataset.

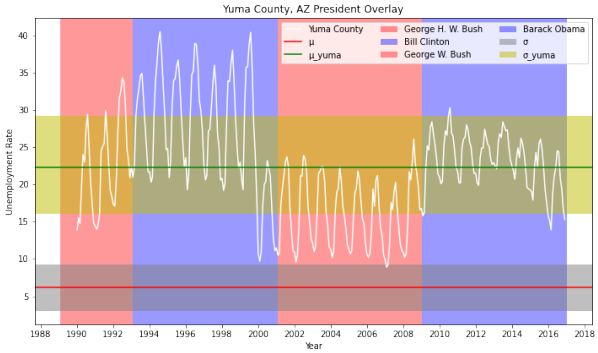


**Figure 16:** Yuma County

Consequently, in Yuma and Imperial county we that the mean unemployment rate of the county is offset by a significant amount compared to the dataset mean and standard deviation.

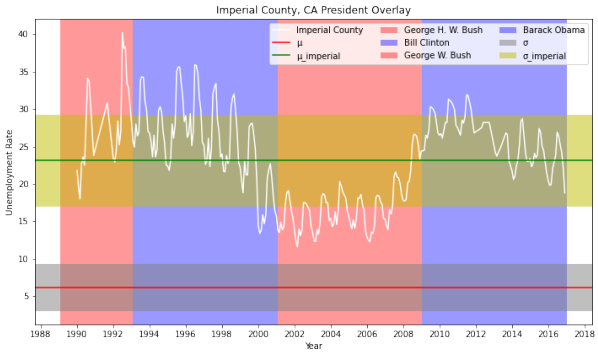


**Figure 17:** Imperial County

We also observe that while the Imperial and Yuma county long term trends are periodic, they are highly variant in the short term.

# 6 Conclusion

Since we were able to have a sensible prediction within our 95% confidence interval we can derive that past data is a useful predictor for future data with high correlation. Obviously, this predictor would be more robust with the addition of more features to better classify and approximate data. When tested tested on counties with abnormal parameters outside of the dataset trends we observed that the accuracy was maintained. This shows that Unemployment rate can be predicted with high accuracy given an appropriate training method.

The more stark difference in unemployment rate after the housing market when compared to the dot com bubble may be due to the broader population affected. During the dot com bubble many people were laid off but the housing market contributed more to financial and livelihood stability. In Figure 12 the map doesn't show much variance except for the coastal counties along the east and west coast. One thing important to note is that mid-western counties seem to be an independent subset of our dataset. The evidence for this can be seen by observing that the unemployment rate in the mid west remain largely unchanged over time. Contrary to that is the east and west coast states which have higher variability and fluctuate with financial events and has a periodic nature.

Comparing California to New York counties in Figure 11 we see that the trends are similar with a slight offset. We know that these to states are both typically Democratic and highly urbanized as opposed to their mid-western counterparts.

There are also larger wealth gaps and higher homelessness in San Diego that other counties do not experience. We can derive that San Diego and New York county have a stronger correlation than New York and Imperial. This makes sense since New York county and San Diego county are large financial hubs and wouldn't have strong correlations with counties with economy based in agriculture and other areas.

In addition, we observe that the New York and San Diego counties are more representative of the National unemployment rate which may be due to their large influence on the economy as compared to other counties.

Another important observation is that of individual counties. We observed that certain counties behave abnormally compared to the overall dataset trends. For example, Imperial County, CA unemployment rate is offset from the mean by approximately 17% and a higher variance. This is due to stimulus provided to agricultural counties.

Imperial County and Yuma county in Arizona both had similar offsets with higher variability while still maintaining a periodic trend of roughly 8 years. Comparing this to New York County shows that individual counties are unique and can be looked at in more detail in further studies.

From our observations in Figure 15 we can derive that there is a correlation between political party in government and unemployment rate over time. In New York county we see the unemployment rate drop during the democratic terms and rise during the republican terms.

The exception here being in George W. Bush's term where unemployment rate fell and the housing market crisis caused the rise in 2008. Overall, the political party of the President in this dataset is too small and too broad of an application given that often times state and local laws affect citizens in there area. This idea is reinforced by the different trends in New York county as compared to Imperial and Yuma county.

# References


[1] Tax Foundation. *Federal Taxes Paid vs. Federal Spending Received by State, 1981-2005*. URL: https://taxfoundation.org/federal-taxes-paid-vs-federal-spending-received-state-1981-2005/.

[2] Venkata Jagannath. *Random Forest Template for TIBCO Spotfire®*. URL: https://community.tibco.com/wiki/random-forest-template-tibco-spotfire.

[3] Jay Ravaliya. *US Unemployment Rate by County, 1990-2016*. URL: https://www.kaggle.com/jayrav13/unemployment-by-county-us.

[4] Wikipedia. *United States housing bubble*. URL: https://en.wikipedia.org/wiki/United_States_housing_bubble.