CPI's Effects on Employment Market

ABSTRACT

Explore Job market with Consumer Price Index (CPI), based on available data from US federal data sources between 1998 -2017.

Team:

Debabrata Kabiraj

Joseph Simone

Rajwant Mishra

CPI's Effects on Employment Market - Draft

Date: 12.2.2020

Team

Debabrata Kabiraj
Joseph Simone
Rajwant Mishra

Source Code

Github
Rpubs

Motivation

We are in totally different era of twenty first century, and it gives us very rare situation where any positive news would help the humankind. We want to use the historical CPI data and find the relation of it with the employment, in hope that we would have some positive news on employment by following the trend of the data in past.

We feel that lower CPI would result in more job opportunity, as the it gives space for more competition in small business across sectors.

Research Design

Types of Research Data: We will use existing data available from US government to do this analysis, with some aggression of the info to derive the missing info.

Research Philosophy: We don't plan to change or replace any of the data points to meet our goal. Since data is used as is with some grouping based on the predefined key, value pair we expect we would not add any bias in this process. We would use the data to amylase the fact about the data.

Types of Research: Since, the analysis knowledge gained would not add immediate practical implications of the finding, we feel its Fundamental research on the topic to bring the correlation and biasness of these two disjoined data points.

Research approach: We would be following an inductive approach, as we would like to find answer to specific research question(s) formulated in the beginning of the research process. Additionally, we will be also following a deductive approach, since we have chosen to achieve research objective(s) via testing hypotheses.

Research design: We would be doing following Conclusive research design, which provide final and conclusive answers to the research question.

Sampling: We would divide data in Train and Test set. Test set would include data from 2017 to till date data available. Train set would be further sampled on 70:30 (Train: Test) ratio in multiple Random sampling process.

Missing value: We would treat missing value as the average of last years and following years value only, instead of taking the whole years average. This would keep impact of missing value limited to neighboring years only.

Ethical Considerations: We would not infer any info apart from the given set of data point. We would try to properly take care of the outlier's impact on the data and report it. We will avoid any type of misleading information, as well as representation of primary data findings in a biased way.

Data History

CPI

The data which encompasses the consumer price index (CPI) provided by the U.S. Department of Labor Bureau of Labor Statistics. This metric is related to the Base of 100 from 1982-1984 published by the U.S. Bureau of Labor Statistics. When looking at the CPI, we can observe the change in the average cost of monetary goods dating back to 1913. The Bureau of Labor Statistics gathers the average prices paid by consumers for hundreds of different items each month. The average is then compared to a reference base period. That base period is an arbitrary date set by the federal government. Currently, the US uses the average of goods and services from 1982 to 1984 and considers that our reference base period with a factor of 100. Inflation measured by Consumer Price Index, CPI, is defined as the change in the prices of a basket of goods and services that are typically purchased by specific groups of households.

Employment by Industry

The CES program dates back to October 1915, when a small sample of manufacturers were asked to provide total employment and payroll data. In 1919, the Bureau of Labor Statistics originally published monthly data on employment and earnings for production workers in manufacturing (monthly average weekly hours data for these workers were added a few years later). That same year, CES began publishing annual employment data for various industries, including detailed industries in the goods-producing sector and in the service-providing sector, the latter of which included wholesale and retail trade, transportation and public utilities, and government. Historically employment is closely tied to recessions and that is no accident.

Data Definition

CPI

The consumer price index (CPI) measures the average level of prices of goods and services in the economy. The CPI formula is used to measure the change in prices by consumers for a representative basket of goods and services during a defined time period. CPI is a widely followed measure of inflation which is used by economists, policy makers, investors to guide economic policy, forecasting and investment decisions. The consumer price index is estimated as a series of summary measures of the period-to-period proportional change in the prices of a fixed set of consumer goods and services of constant quantity and characteristics, acquired, used or paid for by the reference population. Each summary measure is constructed as a weighted average of a large number of elementary aggregate indices. Each of the elementary aggregate indices is estimated using a sample of prices for a defined set of goods and services obtained in,

or by residents of, a specific region from a given set of outlets or other sources of consumption goods and services.

What is the CPI Formula?

CPI = (Cost of market basket in a given year / Cost of market basket in base year) x 100

How is the CPI market basket determined?

The CPI market basket is developed from detailed expenditure information provided by families and individuals on what they actually bought. There is a time lag between the expenditure survey and its use in the CPI. For example, CPI data in 2016 and 2017 was based on data collected from the Consumer Expenditure Surveys for 2013 and 2014. In each of those years, about 24,000 consumers from around the country provided information each quarter on their spending habits in the interview survey. To collect information on frequently purchased items, such as food and personal care products, another 12,000 consumers in each of these years kept diaries listing everything they bought during a 2-week period. Over the 2-year period, then, expenditure information came from approximately 24,000 weekly diaries and 48,000 quarterly interviews used to determine the importance, or weight, of the item categories in the CPI index structure.

Employment by Industry

When workers are unemployed, they, their families, and the country as a whole lose. Workers and their families lose wages, and the country loses the goods or services that could have been produced. In addition, the purchasing power of these workers is lost, which can lead to unemployment for yet other workers. The number of workers earning a wage or salary, along with those who are self-employed. These statistics count workers in private industries and in federal, state, and local government.

Employment rates are defined as a measure of the extent to which available labor resources, people available to work, are being used. They are calculated as the ratio of the employed to the working age population. Employment rates are sensitive to the economic cycle, but in the longer term they are significantly affected by governments' higher education and income support policies and by policies that facilitate employment of women and disadvantaged groups. Employed people are those aged 15 or over who report that they have worked in gainful employment for at least one hour in the previous week or who had a job but were absent from work during the reference week. The working age population refers to people aged 15 to 64. This indicator is seasonally adjusted and it is measured in terms of thousand persons aged 15 and over; and in numbers of employed persons aged 15 to 64 as a percentage of working age population.

Data tables produced annually for employment by industry include: the number of full-time and part-time employees in each industry, full-time equivalent employees, self-employed persons, and persons engaged in production (full-time equivalents plus the self-employed).

What are the basic concepts of employment and unemployment?

1. People with jobs are employed.
2. People who are jobless, looking for a job, and available for work are unemployed.

3. The labor force is made up of the employed and the unemployed.
4. People who are neither employed nor unemployed are not in the labor force.

Sample Data:

Sample of Employment Data:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	GeoFIPS	GeoName	Region	TableName	LineCode	IndustryCl	Description	Unit	1998	1999	2000	2001	2002
2	"02000"	Alaska	8	SAEMP25N	10	...	Total employment (number of jobs)	Number of jobs	382166	381307	389734	394565	402187
3	"02000"	Alaska	8	SAEMP25N	20	...	Wage and salary employment	Number of jobs	298940	299907	307474	313492	319545
4	"02000"	Alaska	8	SAEMP25N	40	...	Proprietors employment	Number of jobs	83226	81400	82260	81073	82642
5	"02000"	Alaska	8	SAEMP25N	50	...	Farm proprietors employment	Number of jobs	547	546	549	559	530
6	"02000"	Alaska	8	SAEMP25N	60	...	Nonfarm proprietors employment 2/	Number of jobs	82679	80854	81711	80514	82112
7	"02000"	Alaska	8	SAEMP25N	70	111-112	Farm employment	Number of jobs	848	897	786	775	744
8	"02000"	Alaska	8	SAEMP25N	80	...	Nonfarm employment	Number of jobs	381318	380410	388948	393790	401443
9	"02000"	Alaska	8	SAEMP25N	90	113-814	Private nonfarm employment	Number of jobs	285610	285404	292267	296462	300949
10	"02000"	Alaska	8	SAEMP25N	100	113-115	Forestry, fishing, and related activitie	Number of jobs	16295	14896	14077	(D)	13382
11	"02000"	Alaska	8	SAEMP25N	101	113	Forestry and logging	Number of jobs	1397	1455	1307	986	858
12	"02000"	Alaska	8	SAEMP25N	102	114	Fishing, hunting and trapping	Number of jobs	14399	12996	12372	(D)	(D)

Sample A

	LineCode	Year	Description	GeoName	Number of jobs
1	705	1998	Food and beverage stores	New York	206308
2	705	1999	Food and beverage stores	New York	209367
3	705	2000	Food and beverage stores	New York	210877

Sample B

Sample of Consumer Price Index (CPI):

9172 lines (9172 sloc) | 338 KB

Raw Blame

Search this file...

	Series ID	Year	Period	Label	Value
2	APU0000701111	1980	M01	1980 Jan	0.203
3	APU0000701111	1980	M02	1980 Feb	0.205
4	APU0000701111	1980	M03	1980 Mar	0.211
5	APU0000701111	1980	M04	1980 Apr	0.206
6	APU0000701111	1980	M05	1980 May	0.207
7	APU0000701111	1980	M06	1980 Jun	0.210
8	APU0000701111	1980	M07	1980 Jul	0.214

Sample C

Sample of CPI & Employment Data by state:

Year	Description	GeoName	Number of jobs	parent_item_code	area_code	area_name	cpi_value
1998	Food and beverage stores	New York	206308	SAF	A101	New York-Northern New Jersey-Long Island	10092.9
1999	Food and beverage stores	New York	209367	SAF	A101	New York-Northern New Jersey-Long Island	10371.2
2000	Food and beverage stores	New York	210877	SAF	A101	New York-Northern New Jersey-Long Island	10527.6
2001	Food and beverage stores	New York	209884	SAF	A101	New York-Northern New Jersey-Long Island	10812.8

Sample D

Data Summary

The group's first objective when working with both the Consumer Price Index and Employment datasets was mapping the two subsets by two common theme, industrial sections and geographical location. These industrial sectors range from a variety of fields, including but not limited to, Food & Beverage manufacturing to Wood Product manufacturing. In addition, during pre-processing, in order to try and detect an early trend with the data, the filtering by state or region was implemented using Geo-Coding to merge to the two sets together for visualization.

As to be expected, from the initial visualizations, shown above, there is no distinctive connection that to can be measured. Although, there does seem to appear, when looking at the

side by side color bar graph of the two labor statistics, that both seem to be increasing overtime. With the exceptions of the CPI value drop-off in 2017. Furthermore, with every increase in both the CPI and Employment Rate seem to happening over a period of time, rather than a significant increase with every year that passes. With this in mind, the team will further investigation the Consumer Price Index's impact on the Job Market.

Introduction

This paper strives to determine if a connection between Consumer Price Index, CPI, and Employment Rate exists. Furthermore, if that is the case which industries or sectors have the most effect on the Consumer Price Index. The vast data set provided by the Bureau of Labor Statistics provided insights into a large array of different industries and regions through the United States. However, during the course of this research, limit time constrains only allowed for the prepossessing of five major industries and 20 states or region. Throughout the deration of this paper, it will be made clear whether or not this was enough of a sample sign for statistically interpretation and warrant further investigations into other industries and other states or regions.

Related Works and Group Hypothesis

This section highlights pertinent resources used during the research phase of the project. As with many Academic Scholarly Articles and the IOT revolution of the 21st century, the use of collaborative research is at an all-time high. These relative works can be summarized into four separate sections. The first section relating to the public data collected from the Bureau of Labor Statistics for analysis [1-3]. The second piece of this project's references pertain to the metrics used to calculate the giving metrics of center observed in the data imported for modeling [4-11]. The third section of works citied refers to relevant studies regarding the main purpose of the article either being the use of CPI or Employment Data as a metric of economic growth, inflation calculation or the type of mathematical modeling used in conjunction with furthering research [5-8]. Finally, the last section includes the both the research and data analysis approaches used by this research team in order to maintain a certain level of scientific and ethical standards [12-13]. We feel CPI has some impact on the Employment Data by each sector and national level. We would validate this with our analysis in the project.

Null Hypothesis

Consumer price index (CPI) has no relation with Job Market (working Employment Data).

Alternate Hypothesis:

Employment is related to CPI.

Trying to answer following questions as well:

a. Which CPI parameter has most impact on the Job Market?
b. What are TOP 5 sector that gets most impacted by CPI?
c. Which sector has no impact of CPI?

Data source:

Data Of Job Market (ZIP File)
BEA.gov
Data of CPI

Literature Review

There has be some historical evidence highlighting that labor costs can be a key predictor of inflation. The Consumer Price Index, evaluates and determines the rate of change in the prices paid for within market bins of goods and services [1]. Whereas, by gaging the fluctuations of those prices, inflation reflects the rise in prices of the goods and services that all consumers buy. Characteristically, the price of goods and services grow over-time, however, there are times where the price can decrease, creating a situation called deflation.

The Consumer Price index is considered to be the one of the best indications of inflation [6]. Additionally, Regis Matthes and Christian Barnichon state that, "the natural rate of unemployment, or u-star, is the hypothetical unemployment rate that is consistent with stable inflation and aggregate production being at its long-run level [4]." In this literature review, the information gathered regarding the effect between CPI and the Employment Rate pertains to inflation's relationship with labor costs and the change in inflation over a thirty-year span in Nigeria.

Historical Attributes Pertaining to Inflation's Impact on the Global Economy

Another key indicator of inflation is commonly believed to be labor costs. This is due to the fact that labor costs characterize approximately two-thirds of entire costs to independent U.S. businesses. From this point of view suggests a Cost-Push Model of inflation, the theory that the main cause of higher prices are higher costs. Alternatively, some economists suggest that any change the market will tolerate, regardless of their real costs [11]. The degree of inflation has substantially risen in Nigeria, over the span of the past three decades in Nigeria.

The significance of this change in inflation could be attributed to the rapid growth of money supply motivated by the expansionary fiscal policies of the public sector. Making this a crucial finding due to the fact that detecting the probable relationship between inflation and economic growth might advance the progression of a comprehensive overhaul of realist policy options to be implemented towards accomplishing wide-ranging macroeconomic stability in Nigeria [7].

Hypothesis Test

- **Null Hypothesis:** CPI Value has no effect on Number of Employment.

  H_o: $\mu_1 = \mu_2 = \mu_3 \ldots = \mu_n$

- **Alternative Hypothesis:** CPI Value has effect on Number of Employment.

  $H_a \neq H_o$: $\mu_1 \neq \mu_2 \neq \mu_3 \ldots \neq \mu_n$

- **Rejection**: Reject H_o (Null Hypothesis) if the calculated value (P-Value) is less than the tabulated value (Table value = 0.05), otherwise do not reject H_o

- **Confidence Interval**: A 1-α confidence interval for a parameter is a range such that if the experiment were repeated many times, the parameter's true value would be outside this range at most α of the time.

Decision Method

Hypothesis Test can be done in various ways depending on the nature of the data. Levene Test is appropriate on categorical/qualitative data but is not appropriate with quantitative explanatory variables.

As part of our hypothesis we conducted many tests using following:

a) Linear Regression Test

Linear regression is used to predict the value of a continuous variable Y based on one or more input predictor variables X.

As the p-value is much less than 0.05, we reject the null hypothesis that $\beta = 0.$ with 95% confidence level. Hence there is a significant relationship between the variables (CPI and Number of Jobs) in the linear regression model of the data set faithful.

b) Stepwise Logistic Regression Test

Logistic regression is just like linear regression, except that it predicts the log odds of the response variable. Stepwise regression is a combination of the forward and backward selection techniques of Logistic Regression. To escape the problem of multicollinearity (correlation among independent variables) and to filter out essential variables/features from a large set of variables, a stepwise regression usually performed.

Stepwise regression is a modification of the forward selection so that after each step in which a variable was added, all candidate variables in the model are checked to see if their significance has been reduced below the specified tolerance level. If a nonsignificant variable is found, it is removed from the model. Stepwise regression requires two significance levels: one for adding variables and one for removing variables. The cutoff probability for adding variables should be less than the cutoff probability for removing variables so that the procedure does not get into an infinite loop.

The process starts with initially fitting all the variables and after that, with each iteration, it starts eliminating variables one by one if the variable does not improve the model fit. The AIC metric is used for checking model fit improvement.

c) T-Test

A two-sample t-test tests whether two sequences of real-valued samples come from a distribution with different means (population average) equal to some fixed value μ_0. The t-test is a parametric test; it assumes the data have a normal distribution

d) ANOVA Test

ANOVA (one-way ANalysis Of VAriance) is used to compare nested models in general, but it's often not necessary with continuous predictors as *lm* (Linear Model) gives a significance level for each parameter weight. Anova is useful with categorical predictors because each categorical predictor is usually associated with many parameters (one for each category, except for the baseline).

- Null hypothesis: the means of the different groups are the same
- Alternative hypothesis: At least one sample mean is not equal to the others.

ANOVA test can be applied only when:

i. The observations are obtained independently and randomly from the population defined by the factor levels
ii. The data of each factor level are normally distributed.
iii. These normal populations have a common variance

Anova calculated the significance of the interaction term in the linear model CPI ~ Jobs, and compare that value with the significance of the parameter associated with the interaction term computed by *lm*.

e) Chi-Squared Test

In order to establish that 2 categorical variables are dependent, the chi-squared statistic should be above a certain cutoff. This cutoff increases as the number of classes within the variable increases. The null hypothesis of the chi-squared test is that the two variables are independent and the alternate hypothesis is that they are related.

We have a high chi-squared value (***X-squared = 3446521***) and a p-value (0.000209) of less that 0.05 significance level. So we reject the null hypothesis

Method

Results

By rejecting the Null Hypothesis, it presented the opportunity to further investigation in the possibilities of this paper's Alternative Hypothesis. The United States' economy is not only comprised of a vast number of different industries, it also has State and Regional economies that measure those of other first world countries. In light of this, further analysis of these markets that we sampled for Modeling and Hypothesis Testing to see the contributing factors from the sampled data.

First when sampling the data for Hypothesis Testing, the extrapolation of both the Employment data and the Consumer Price Index from the U.S. Bureau of Labor Statistics, required the mapping of these two metrics by primary keys. For the purpose of this paper and analysis, this was accomplished through sub-setting the BLS's immense data set by states or regions and by five industries randomly selected. These five industries include *Transportation and Warehousing*, *Real Estate*, *Food and Beverage Stores*, *Health Care and Social Assistance* and *Education*. The states and regions sampled for this research contain 20 out of the 50 United States of America; *AK, AZ, CA, CO, FL, GA, HI, IL, KS, MA, MI, MN, MO, NY, OH, OR, PA, TX, WA, WI*.

Initially analyzing the five industries sampled from the BLS, the goal was to identify the industry with the highest impact on the impact on the Consumer Price Index. According to **Table 2**, of the five industries sampled, Real Estate reporting a CPI value of *333,754.24* in 2014, **Table 3**. As shown in **Fig. 4**, the CPI value from 2015 and 2016 was also around the value. On the other hand, both *Health Care* and *Education* proved the have the lowest CPI values, shown in **Table 2.**

Year Description	Number of Jobs	CPI Value
2014 Real estate and rental and leasing	626686	33754.24
2012 Transportation and warehousing	115015	27264.89
2016 Food and beverage stores	235573	15508.96
2016 Health care and social assistance	879671	5739.41
2013 Educational services	453675	4051.45

Table 2

Year Description	Number of Jobs	CPI Value
2014 Real estate and rental and leasing	626686	33754.24

Table 3

Upon the discovery that the *Real Estate* market has the highest impact on the Consumer Price Index of the five industries sample, the focus was shifted towards the 20 states and regions mapped out throughout the sampled industries. Of the 20 states the top five states, shown in Table 4, that contribute the *Real Estate* Market include NY, WA, MA, IL and CA. As shown in **Table 5**, NY has the highest CPI value among the states and regions of *333,754.24.* Consequently, that is the same *333,754.24* that appears in **Table 2**. Additionally, noting that, according to **Fig. 5**, the state of WA's *Real Estate* effect on the Consumer Price Index is vastly approach that of NY with every passing year.

State	Year Description	Number of Jobs	CPI Value
NY	2014 Real estate and rental and leasing	626686	33754.24
WA	2016 Real estate and rental and leasing	198644	32596.98
MA	2016 Real estate and rental and leasing	200471	30213.26
IL	2014 Real estate and rental and leasing	291445	29975.50
CA	2016 Real estate and rental and leasing	1153175	27725.69

Table 4

State	Year Description	Number of Jobs	CPI Value
NY	2014 Real estate and rental and leasing	626686	33754.24

Table 5

Fig 4, 5 is available in Appendix.

Discussion

Hypothesis Result:

We can reject Null Hypothesis (Ho) since P-Value is 0.000209 as per **Result Table 1** which is less than Tabulated value of 0.05 and thus accept the Alternative Hypothesis (Ha).

Linear Regression Test

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -9215  -4879  -3086   1892  27158
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.443e+03  2.108e+02  25.816  < 2e-16 ***
## No_Jobs     1.839e-03  4.950e-04   3.715 0.000209 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7121 on 1913 degrees of freedom
## Multiple R-squared:  0.007163,   Adjusted R-squared:  0.006644
## F-statistic:  13.8 on 1 and 1913 DF,  p-value: 0.0002089
```

Stepwise Logistic Regression Test

```
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -123.49   -80.03   -44.21    23.71   242.79
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 8.612e+00  3.822e-04 22530.9   <2e-16 ***
## No_Jobs     2.717e-07  7.928e-10   342.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 13266179  on 1914  degrees of freedom
## Residual deviance: 13158328  on 1913  degrees of freedom
## AIC: Inf
##
## Number of Fisher Scoring iterations: 5
```

ANOVA Test

```
##               Df    Sum Sq   Mean Sq F value   Pr(>F)
## No_Jobs        1 6.998e+08 699833372    13.8 0.000209 ***
## Residuals   1913 9.700e+10  50704223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

T-Test

```
##
##  Welch Two Sample t-test
##
## data:  cpi_emp_data_Combined$No_Jobs and cpi_emp_data_Combined$cpi_value
## t = 35.253, df = 1915.8, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  250193.1 279670.8
## sample estimates:
```

```
## mean of x mean of y
## 270873.36   5941.46

##
##  Pearson's product-moment correlation
##
## data:  cpi_emp_data_Combined$No_Jobs and cpi_emp_data_Combined$cpi_value
## t = 3.7151, df = 1913, p-value = 0.0002089
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03999459 0.12894081
## sample estimates:
##       cor
## 0.0846363
```

Chi-Squared Test

```
##
##  Pearson's Chi-squared test
##
## data:  cpi_emp_data_Combined$cpi_value and cpi_emp_data_Combined$No_Jobs
## X-squared = 3446521, df = 3444522, p-value = 0.000209
```

Result Table 1

When conducting a residual analysis, a "**residuals versus fits plot**" is the most frequently created plot. It is a scatter plot of residuals on the *y* axis and fitted values (estimated responses) on the *x* axis. The plot is used to detect non-linearity, unequal error variances, and outliers.

The plot under **Fig. 1** shows that the mean residual doesn't change with the fitted values (and so is doesn't change with x), but the spread of the residuals (and hence of the y's about the fitted line) is increasing as the fitted values (or x) changes. That is, the spread is not constant. i.e., Heteroskedasticity.

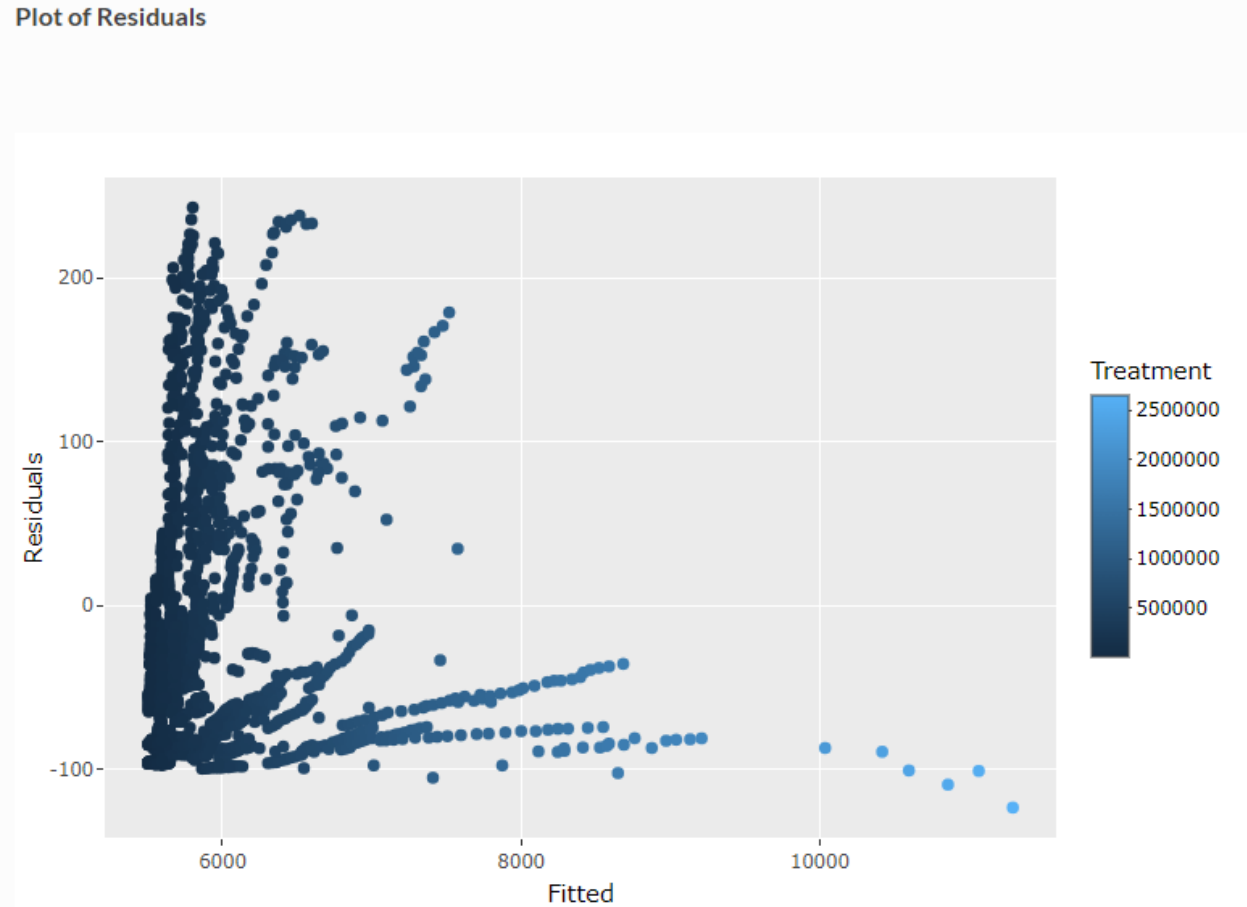


Fig. 1

Q-Q Plots (Quantile-Quantile plots) are plots of two quantiles against each other. A quantile is a fraction where certain values fall below that quantile. The purpose of Q-Q plots is to find out if two sets of data come from the same distribution.

The following normal quantile-quantile (QQ) plots under **Fig. 2** and **3** shows that the quantiles from a theoretical normal distribution on the horizontal x-axis compared to a set of residuals on the vertical y-axis. The points are not clustered on the 45 degree line, and in fact follow a curve, suggesting that the sample data is not normally distributed and has a right skewed or positively skewed distribution.

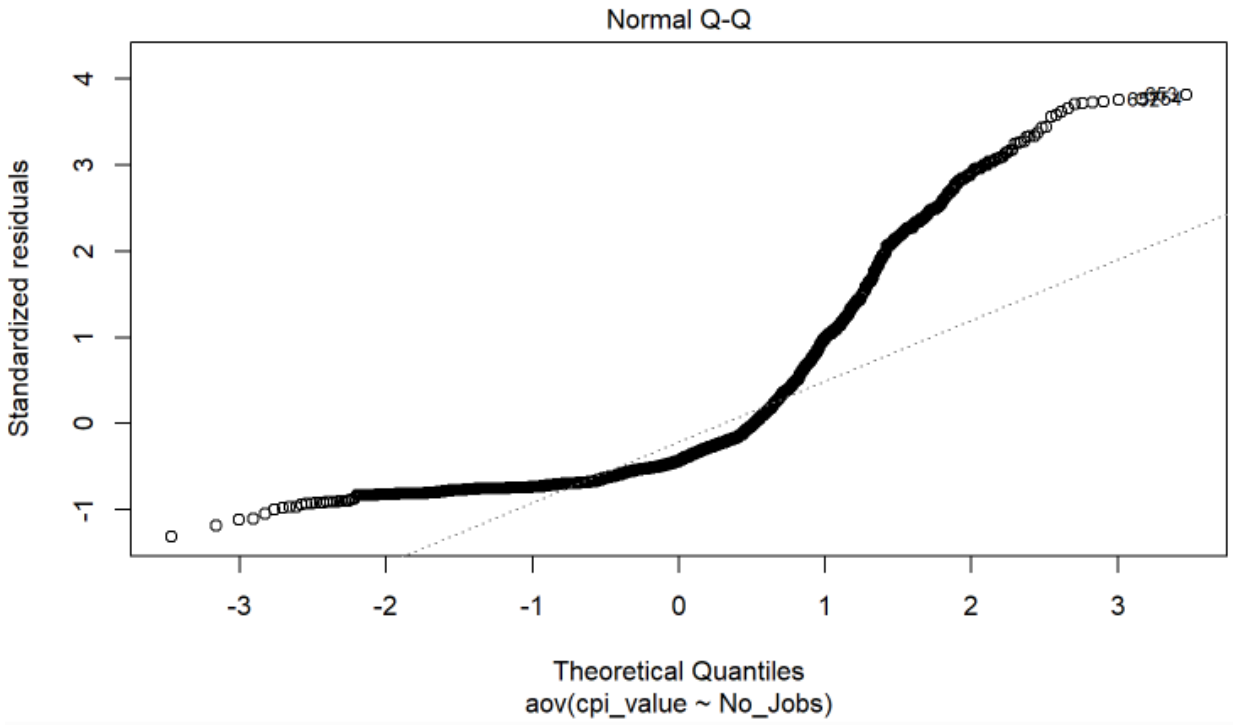

Normal Q-Q
Standardized residuals
Theoretical Quantiles
aov(cpi_value ~ No_Jobs)

Fig. 2

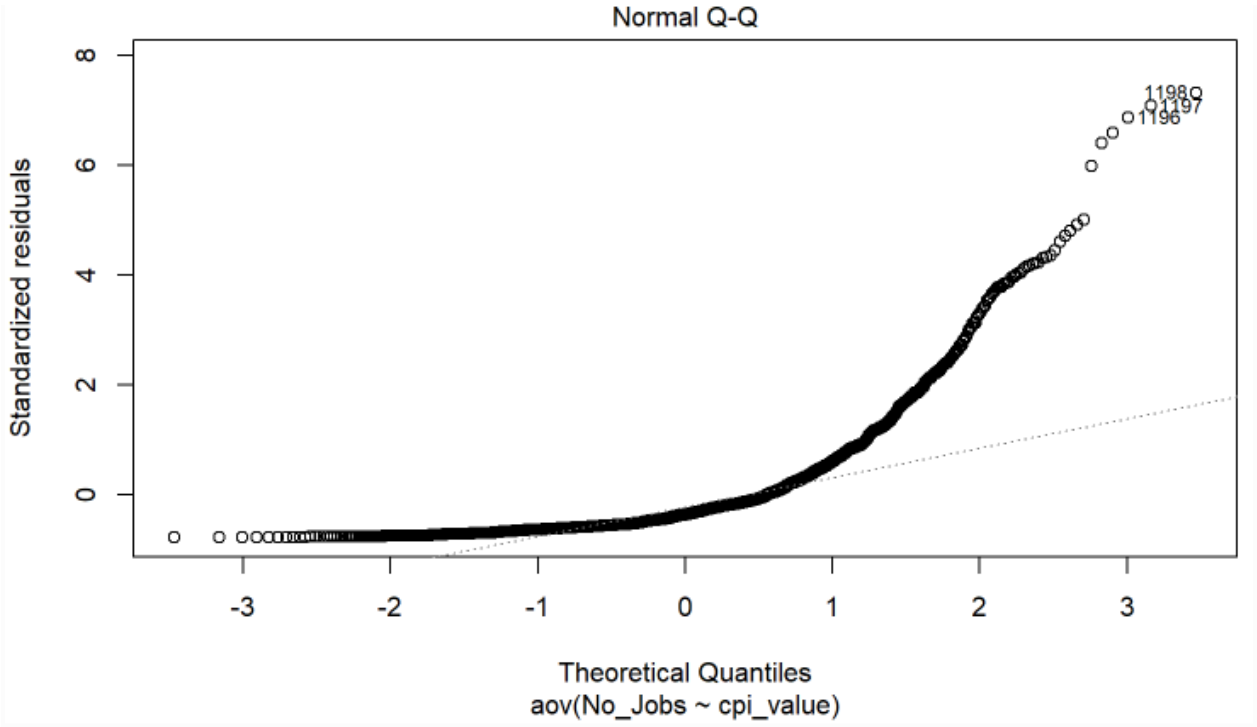

Normal Q-Q
Standardized residuals
1198
1197
1196
Theoretical Quantiles
aov(No_Jobs ~ cpi_value)

Fig. 3

Model /Prediction:

With such an important data of Jobs and CPI and Industry information for each state. We have following:

Predictor Variable: Consumer Price Index, Industry, State,
Response Variable: Number of Jobs

By looking at the **Fig 6**, which shows some pattern and correlation with No. of Jobs and CPI value (Y-axis) and Year.
The Fig 6, is New York data over years of the Food and Beverage industry and shows consistent growth with CPI for the same industry and the number of Jobs generated over the years.

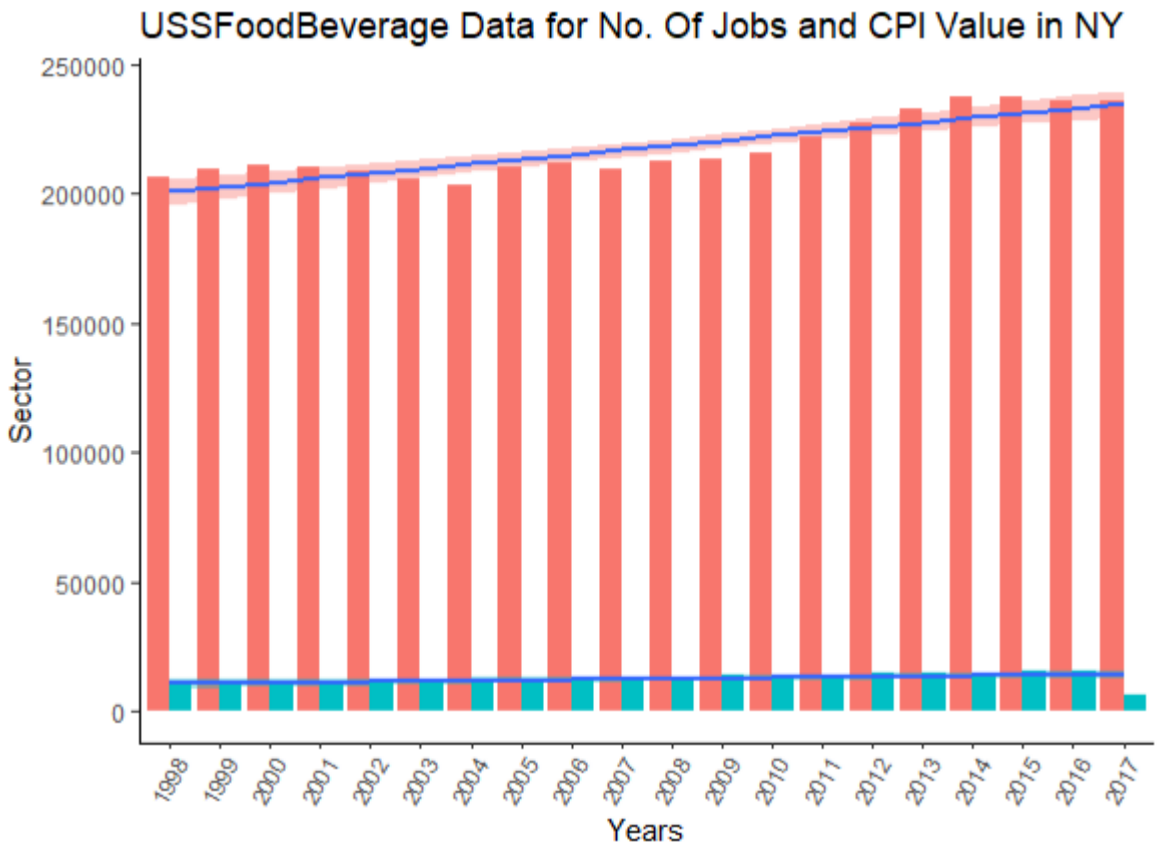


Fig 6

Model Building: In our initial hypothesis we found Jobs are dependent upon the CPI, which in turn depends upon location, states and its consumer behavior. We used this information from the dataset available to us, and created the Predicative Model, which can predict the Number of Jobs a state can generate if we know the Industry/CPI of the proposed Industry.

Assumption: We used our best to pair the data of the Employment and CPI data **(Table A.1).** Which shows how industry in respective datasets are joining to produce the result for our model. We have also mapped states to their closest possible city, to pair the Employment data with CPI data **(Table A.2**).

Model Selection:

MODEL A lm(Jobs ~ CPI+C_Item,ALL_train)	Multiple R-squared: 0.4133, Adjusted R-squared: 0.4114 p-value: < 2.2e-16
MODEL B lm(Jobs ~ (CPI*GeoName)+CPI+GeoName+ C_Item+Year,ALL_train)	Multiple R-squared: 0.781, Adjusted R-squared: 0.772 p-value: < 2.2e-16
MODEL C lm(Jobs~ (CPI* GeoName* C_Item) + CPI+ GeoName+ C_Item+Year, ALL_train)	Multiple R-squared: 0.9797, Adjusted R-squared: 0.9765 p-value: < 2.2e-16

We created many predictive data models using the regression techniques of Simple Liner Regression, Multiple Liner Regression, Partial Least Square and Generalized linear model. We observed some interesting result as you can see in **Table 2**.

Table 2

Model NAME	dAICc	df	weight
Model C	0	208	1
Model D	191.5	133	<0.001
Model B	3296.7	62	<0.001
Model A1	4693.8	7	<0.001
Model A2	4693.8	7	<0.001

Table 3

AIC check on multiple linear Model, (**Table 3**) shows how Model C is performing better, even though we noted both Model and C and D were having adjusted r-squired more than 97%.

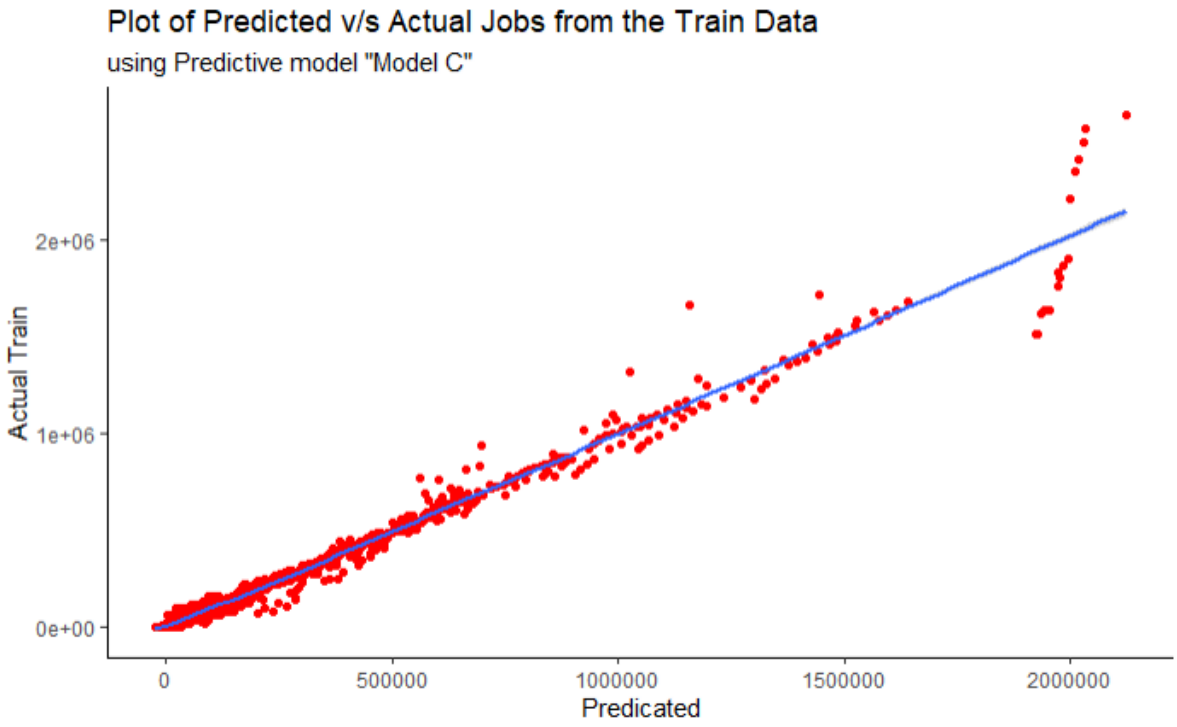
After this we used Model *lm_cpi_item_name2* (**Model C**), along with other model to check the performance of model on the test data. Model with better AIC gave better MAPE score and our results indicates **(Result Table 2**) the support for our Alternate Hypotheses where we see some relation with Consumer Price Index (CPI) on the Job market.

Model Names	RMSE	R-squared	MAE	MAPE
Model B	1.37E+05	7.95E-01	1.01E+05	1.368536
Model C	6.14E+04	9.63E-01	2.67E+04	0.2057522
Model D	6.07E+04	9.62E-01	3.16E+04	0.3190908
Model A1	1.40E+05	7.84E-01	1.03E+05	1.304226
Model A2	2.27E+05	4.00E-01	1.45E+05	1.27469
PLS MODEL	1.40E+05	7.839705e-01	1.03E+05	1.302576
GLM Model	1.40E+05	7.84E-01	1.03E+05	1.304226
Best	6.07E+04	4.00E-01	2.67E+04	2.06E-01

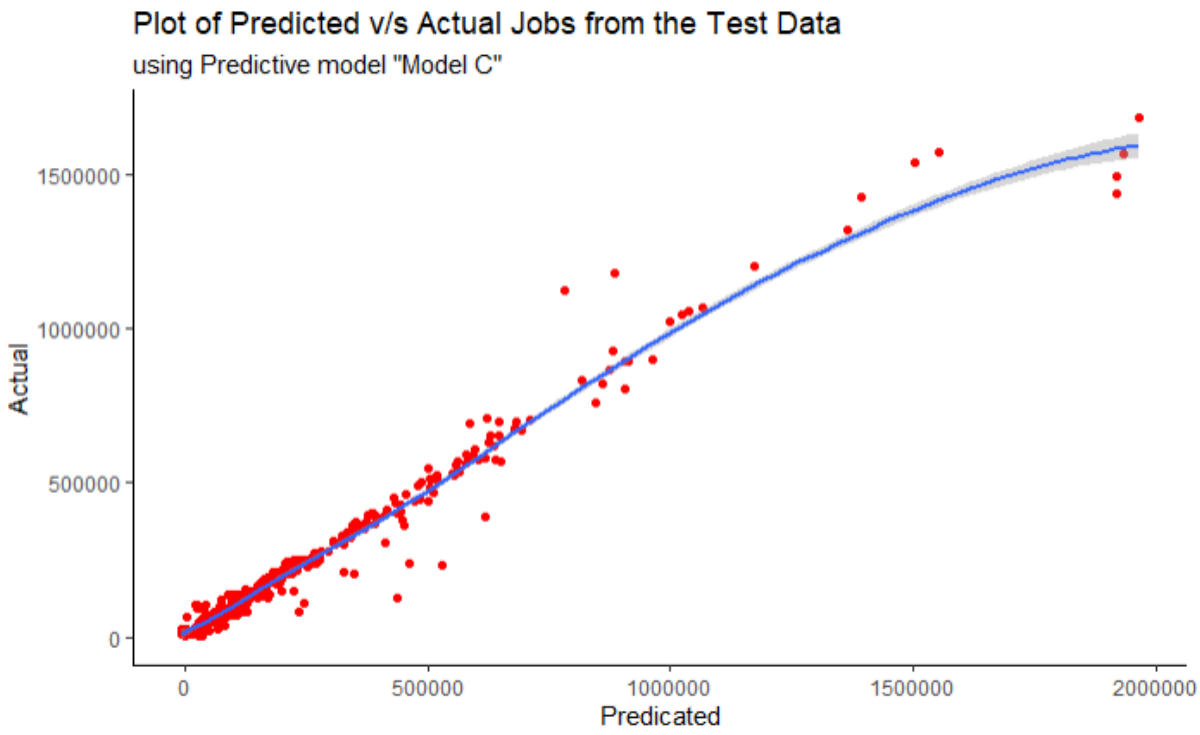
Result Table 2

Model Performance:

We do see that our model has underfitted the training data, with adjusted R-squared being close to 98%, their predication on train set was not as close as expected. As per **Fig 6**, the performance of the model on the test set was satisfactory, as we were able to project most of data points.



Training Model Performance
Fig. 6.1



Test Model Performance
Fig. 6.2

Below graph shows how our model can predict most of the data points (**Fig. 7**)
Fig.8,9,10 very clearly shows how model is predictive the NY, Florida and Texas data by sector.

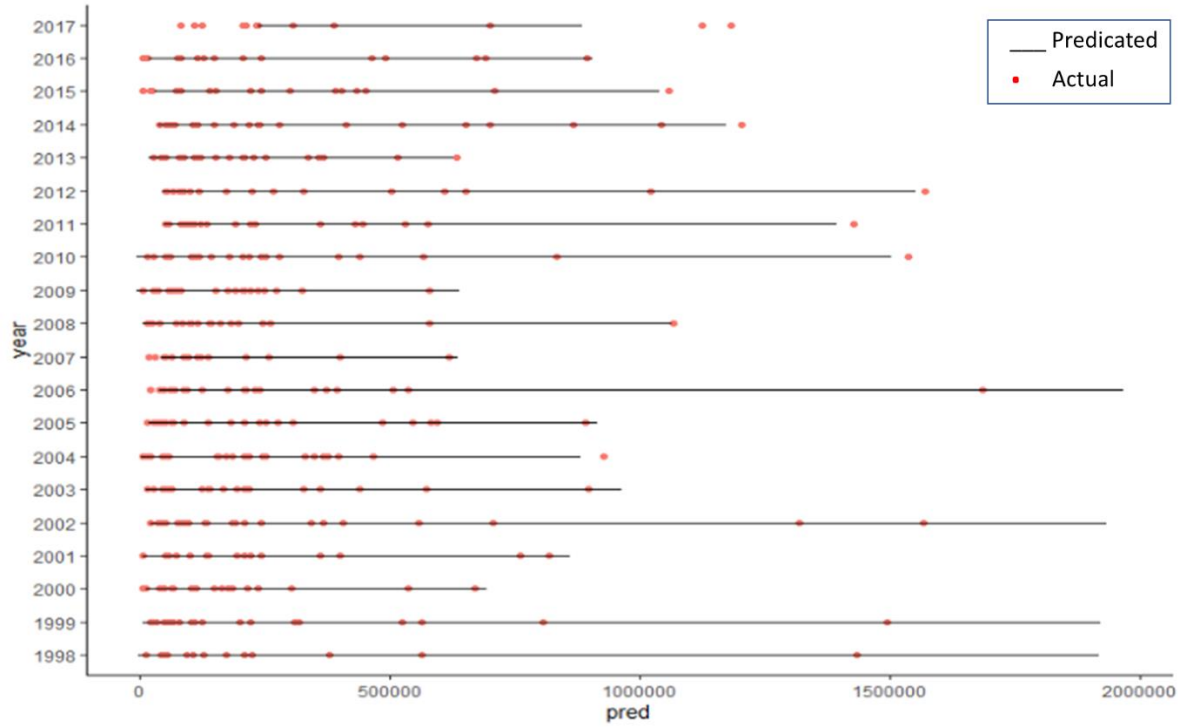


Fig.7

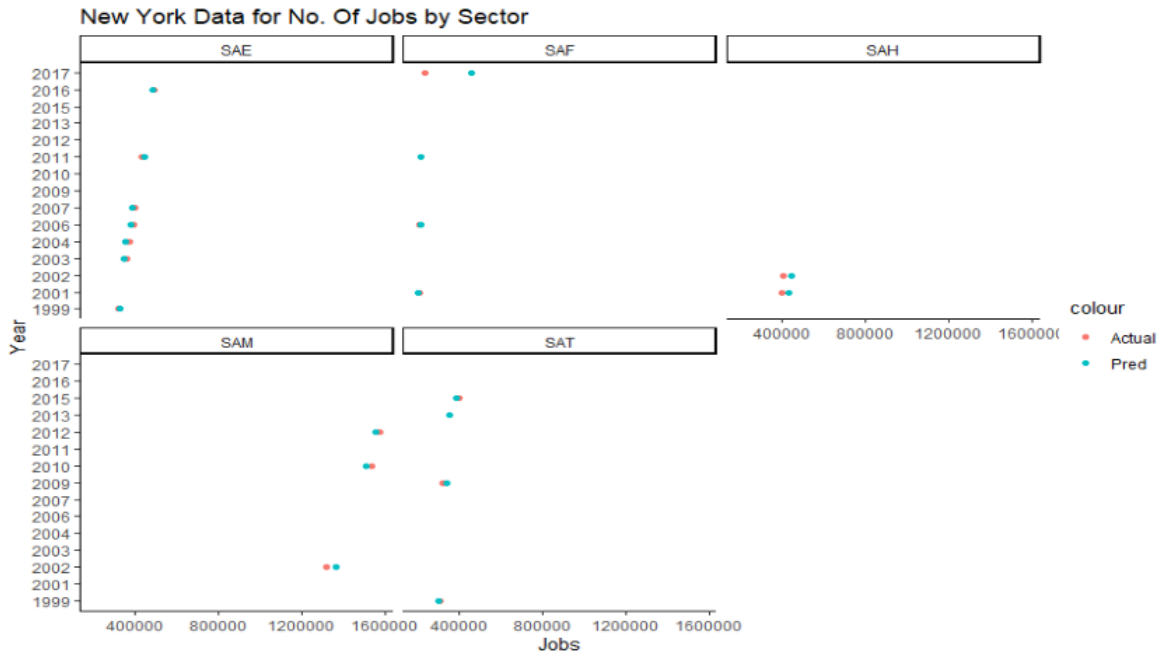


Fig.8

Fig 9 and Fig 10, available in appendix.

Limitation:

Data: Data was huge and not all the attributes were mapping with the information that our study was trying to look for. We had to do lot of data wrangling and *mapping* to meet our objective where we mapped City to States and Industry Codes from CPI data to Industry for Employee data.

Missing Value: We replaced all missing value with 0, as some of the data points were dropped from dataset from the source due to its nature of importance. We had an option to do some to take average of old data to predict this information, but frequency of such numbers was not too high.

Multicollinearity: We do see out model is multicollinear and it makes it hard to interpret our coefficients, and it reduces the power of your model to identify independent variables that are statistically significant. These are serious problems. However, sometimes we can't fix multicollinearity or finding ways to fix is not important if the objective is met.

Conclusion

The possible of finding a connection between the Consumer Price Index and the Employment Rate is feasible by comparing the Data culminated by the US Labor Bureau. Furthermore, a regression model created in congruence with a Support Vector Machine, SVD, can be executed while building a Financial Conditions Index, FCI. The idea behind this is to explore the connection between compound index of financial indicators and impending inflation

[9]. “Compared with the traditional econometric method, our model takes the advantage of the machine learning method to give a more accurate forecast of future CPI in small dataset [9].”

Beginning with the rejection of this paper’s Null Hypothesis given the **p-value of 0.000209** which is less than tabulated value of 0.05 and thus accept the Alternative Hypothesis. In addition, the five industries explored during the course of this paper includes *Transportation and Warehousing*, *Real Estate*, *Food and Beverage Stores*, *Health Care and Social Assistance* and *Education*. Furthermore, the states and regions sampled for this research contain 20 out of the 50 United States of America; *AK, AZ, CA, CO, FL, GA, HI, IL, KS, MA, MI, MN, MO, NY, OH, OR, PA, TX, WA, WI*. Real Estate reporting a CPI value of *333,754.24* in 2016. Consequently, the top five contributing states for Real *Estate* Market include NY, WA, MA, IL and CA. The largest contributor being the state of NY.

Future Work:

In future we expect to explore other industry and states. We see good potential in using this data able to predict future job market in such hard time.

This analysis helps in the general macro-economic idea that improving consumption among the public through various schemes like UBI or Stimulus especially in times of down-turn gives a major positive boost to the overall job market in a direct manner.

References

[1] “CPI Home,” U.S. Bureau of Labor Statistics. [Online]. Available: http://www.bls.gov/cpi/. [Accessed: 15-Oct-2020].

[2] Labor, U.S. Department of. "Bureau of Labor Statistics." 31 July 2020, https://www.bls.gov/news.release/pdf/eci.pdf. 9 October 2010.

[3] “How to Use the Consumer Price Index for Escalation,” U.S. Bureau of Labor Statistics. [Online]. Available: http://www.bls.gov/cpi/factsheets/escalation.htm. [Accessed: 15-Oct-2020].

[4] R. B. and C. Matthes, “The Natural Rate of Unemployment over the Past 100 Years,” Federal Reserve Bank of San Francisco, 14-Aug-2017. [Online]. Available: http://www.frbsf.org/economic-research/publications/economic-letter/2017/august/natural-rate-of-unemployment-over-past-100-years. [Accessed: 15-Oct-2020].

[5] Bartalotti, Otávio, “CPI and inflation: relationship between MoM and YoY values,” CPI and inflation: relationship between MoM and YoY values | Department of Economics. [Online]. Available: http://www.econ.iastate.edu/ask-an-economist/cpi-and-inflation-relationship-between-mom-and-yoy-values. [Accessed: 15-Oct-2020].

[6] “Inflation and Its Measurement: Explainer: Education.,” Reserve Bank of Australia . [Online]. Available: www.rba.gov.au/education/resources/explainers/inflation-and-its-measurement.html. [Accessed: 15-Oct-2020].

[7] A. Banerji, “The Relationship Between Labor Costs and Inflation: A Cyclical Viewpoint.’ Economic Cycle Research Institute,” Economic Cycle Research Institute, May 2005.

[8] “Effective Federal Funds Rate,” FRED, 13-Oct-2020. [Online]. Available: https://fred.stlouisfed.org/series/FEDFUNDS. [Accessed: 15-Oct-2020].

[9] Y. Wang, B. Wang, and X. Zhang, “A New Application of the Support Vector Regression on the Construction of Financial Conditions Index to CPI Prediction,” Procedia Computer Science, vol. 9, pp. 1263–1272, 2012.
[10] M. Bryan and S. Cecchetti, “The Consumer Price Index as a Measure of Inflation,” 1993.
[11] M. Idris and R. Bakar, “The Relationship between Inflation and Economic Growth in Nigeria: A Conceptual Approach,” Asian Research Journal of Arts & Social Sciences, vol. 3, no. 1, pp. 1 research-methodology.net/research-methods/.–15, 2017.
[12] Research Methods,” *Research*. [Online]. Available: https://research-methodology.net/research-methods/. [Accessed: 15-Oct-2020].
[13] Data Collection. [Online]. Available: https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html. [Accessed: 15-Oct-2020].

Appendix

Figure/ Graphs

Fig. 4

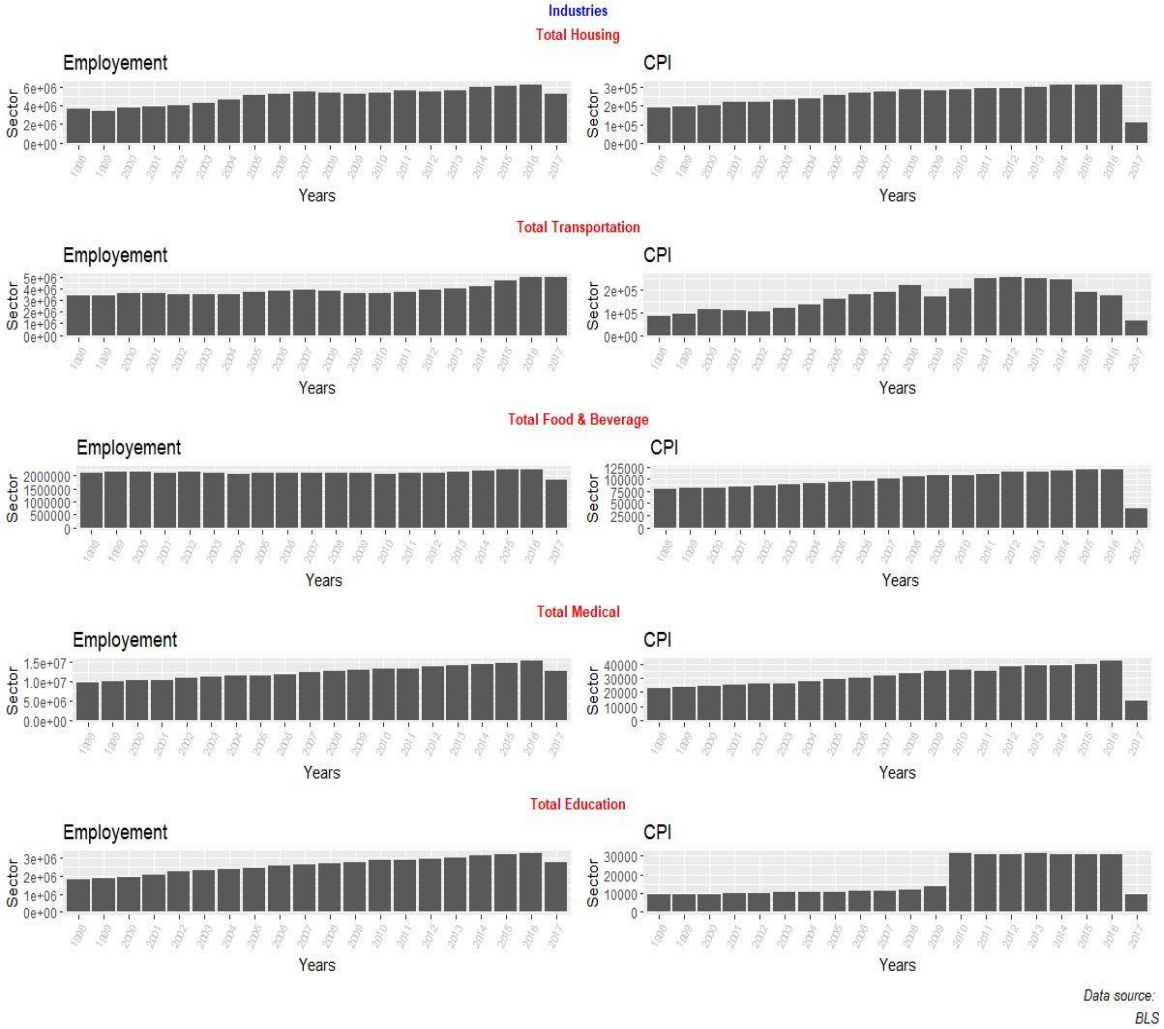

Industries
Total Housing
Employement
CPI
Sector
Years
Total Transportation
Employement
CPI
Sector
Years
Total Food & Beverage
Employement
CPI
Sector
Years
Total Medical
Employement
CPI
Sector
Years
Total Education
Employement
CPI
Sector
Years
Data source:
BLS

Fig. 5

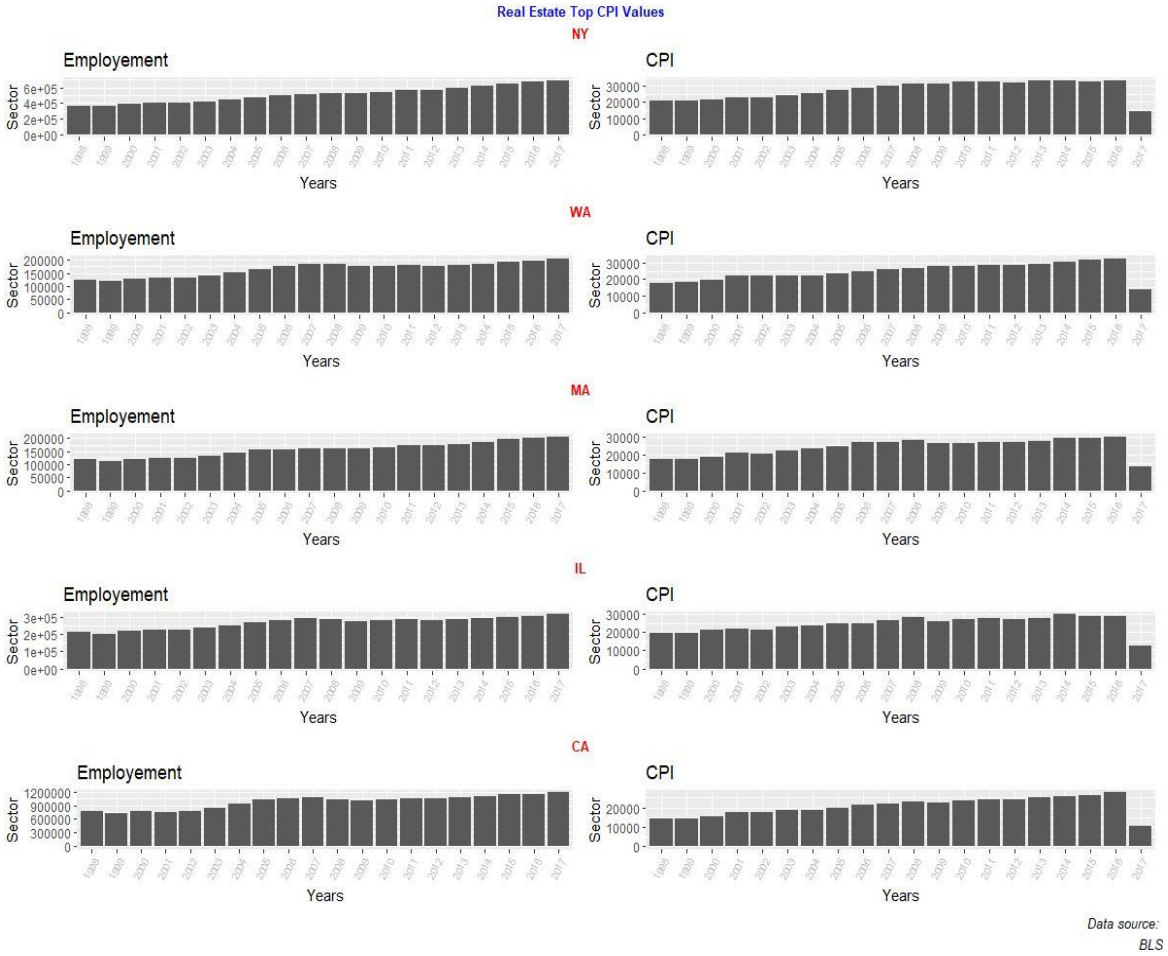

Real Estate Top CPI Values
NY
Employement
CPI
WA
MA
IL
CA
Sector
Years
Data source:
BLS

Fig. 9

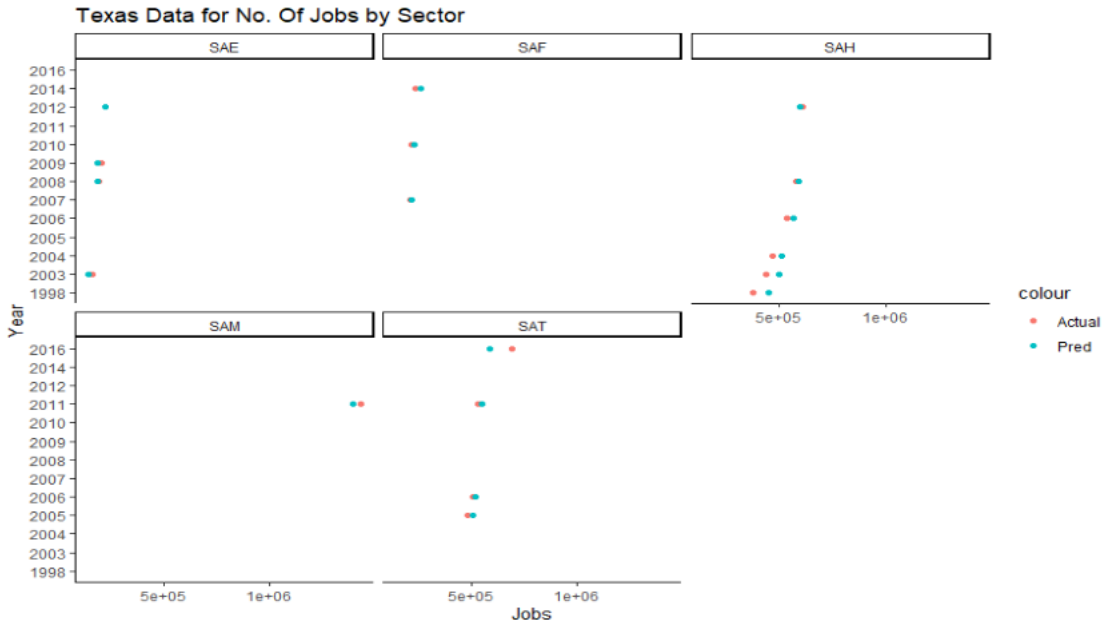
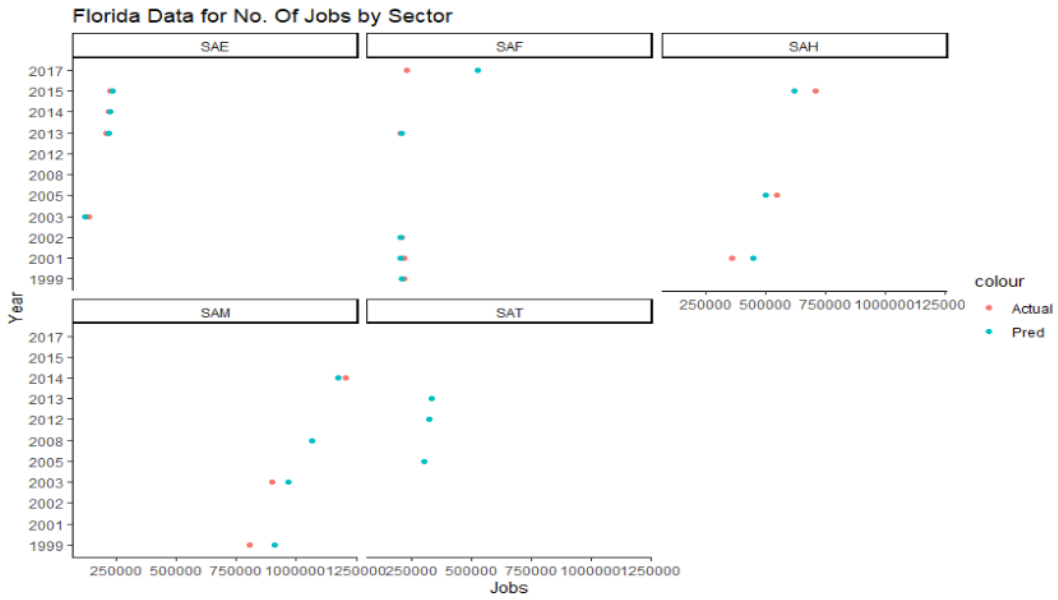

Texas Data for No. Of Jobs by Sector
SAE
SAF
SAH
SAM
SAT
Year
Jobs
colour
Actual
Pred

Fig. 10



Tables

Table A.1

item_code	item_name	display_level	selectab	sort_sequence	Industry
SAF	Food and beverages	0	T	3	USFoodBeverage
SAF1	Food	1	T	4	USFoodBeverage
SAF11	Food at home	2	T	5	USFoodBeverage
SAF111	Cereals and bakery products	3	T	6	USFoodBeverage
SEFA	Cereals and cereal products	4	T	7	USFoodBeverage
SEFA01	Flour and prepared flour mixes	5	T	8	USFoodBeverage
SEFA02	Breakfast cereal	5	T	9	USFoodBeverage
SS01031	Rice	6	T	11	USFoodBeverage
SEFB	Bakery products	4	T	12	USFoodBeverage
SEFB01	Bread	5	T	13	USFoodBeverage
SS02011	White bread	6	T	14	USFoodBeverage
SS02021	Bread other than white	6	T	15	USFoodBeverage
SS02041	Fresh cakes and cupcakes	6	T	18	USFoodBeverage
SS02042	Cookies	6	T	19	USFoodBeverage

Table A.2

State Code	GeoName in EMP9818.csv	Area_Name in cu.area.csv
	United States	U.S. city average
AL	Alabama	
AK	Alaska	Anchorage
AZ	Arizona	Phoenix-Mesa
AR	Arkansas	
CA	California	Los Angeles-Riverside-Orange County San Francisco-Oakland-San Jose San Diego
CO	Colorado	Denver-Boulder-Greeley
FL	Florida	Miami-Fort Lauderdale Tampa-St. Petersburg-Clearwater
GA	Georgia	Atlanta
	Great Lakes	
HI	Hawaii	Honolulu

Code

https://rpubs.com/Rajwantmishra/CAP_P3

https://rajwantmishra.shinyapps.io/EMP_DATA/