CPI’s Effects on Employment Market - Draft

**10.21.20**

# Team

Debabrata Kabiraj  
Joseph Simone  
Rajwant Mishra

# Source Code

Github: <https://github.com/josephsimone/DATA698>

Rpubs: <https://rpubs.com/Rajwantmishra/CAP_P2>

# 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.

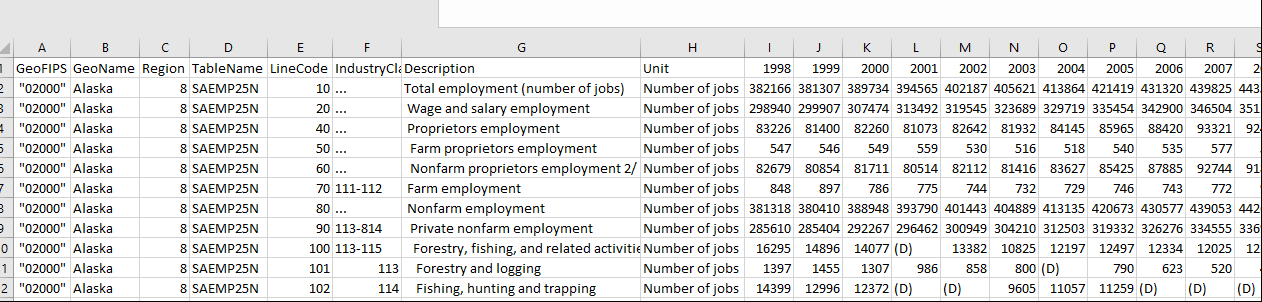
Employment rates are defined as a measure of the extent to which available labor resources, people available to work, are being used. They are calculated as the ratio of the employed to the working age population. Employment rates are sensitive to the economic cycle, but in the longer term they are significantly 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.

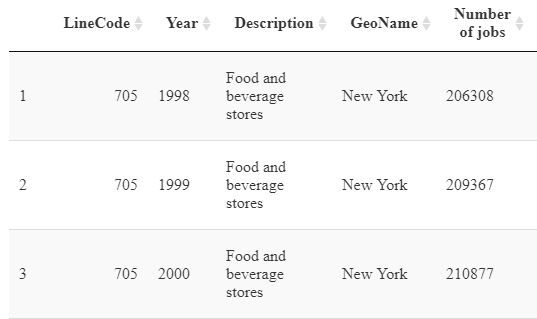
Data tables produced annually for employment by industry include: the number of full-time and part-time employees in each industry, full-time equivalent employees, self-employed persons, and persons engaged in production (full-time equivalents plus the self-employed).

***What are the basic concepts of employment and unemployment?***

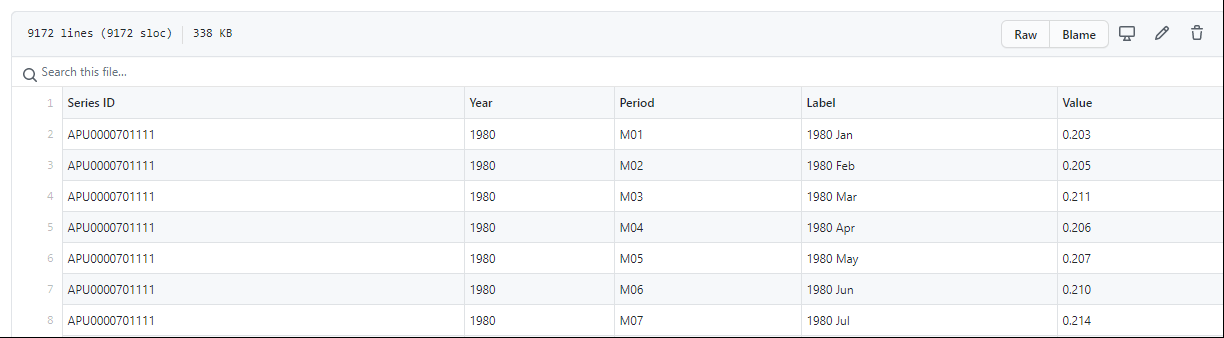
1. People with jobs are employed.
2. People who are jobless, looking for a job, and available for work are unemployed.
3. The labor force is made up of the employed and the unemployed.
4. People who are neither employed nor unemployed are not in the labor force.

# Sample Data:

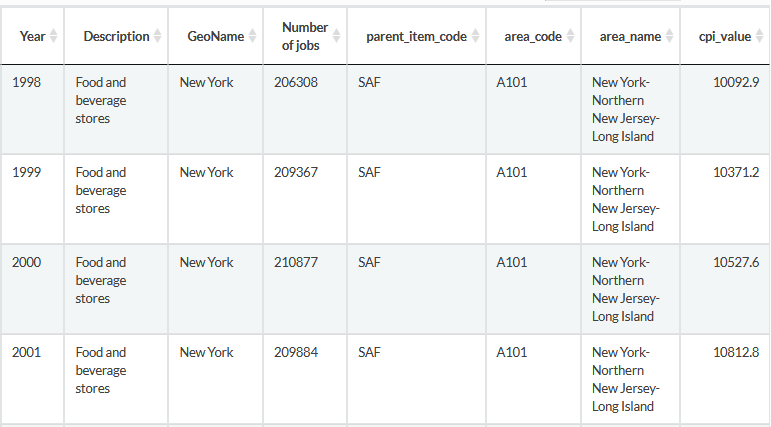
**Sample of Employment Data:**



**Sample of Consumer Price Index (CPI):**



**Sample of CPI & Employment Data by state:**



# 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.

# 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.

1. Which CPI parameter has most impact on the Job Market?
2. What are TOP 5 sector that gets most impacted by CPI?
3. Which sector has no impact of CPI?

# Literature Review

The purpose of this paper is 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 Employment Rate. Concurrently, 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.

Ho: μ1 = μ2 = μ3 ... = μn

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

Ha ≠ Ho : μ1 ≠ μ2 ≠ μ3 ... ≠ μn

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

## Decision

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

## 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

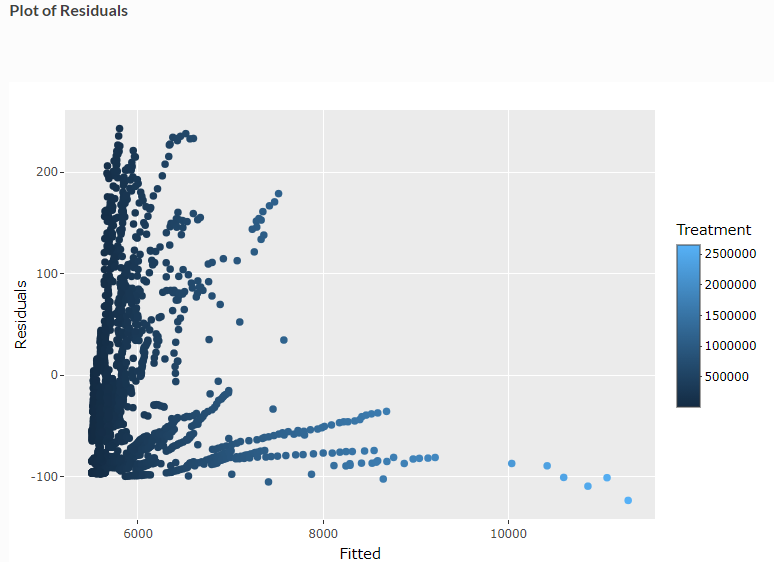
## Multiple R-squared: 0.007163, Adjusted R-squared: 0.006644

## F-statistic: 13.8 on 1 and 1913 DF, p-value: 0.0002089

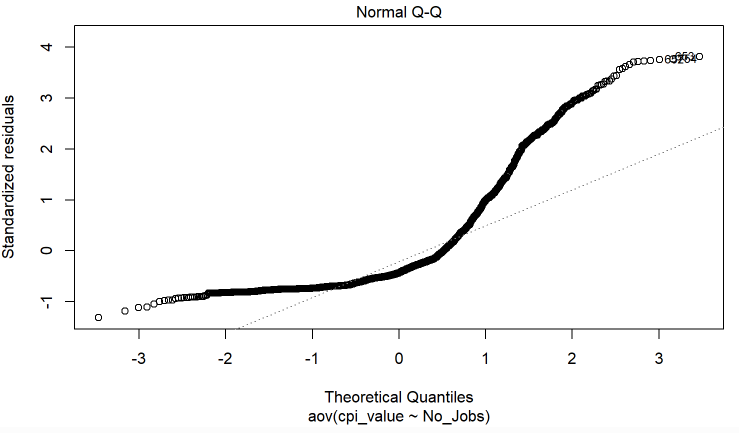
**Table 1**

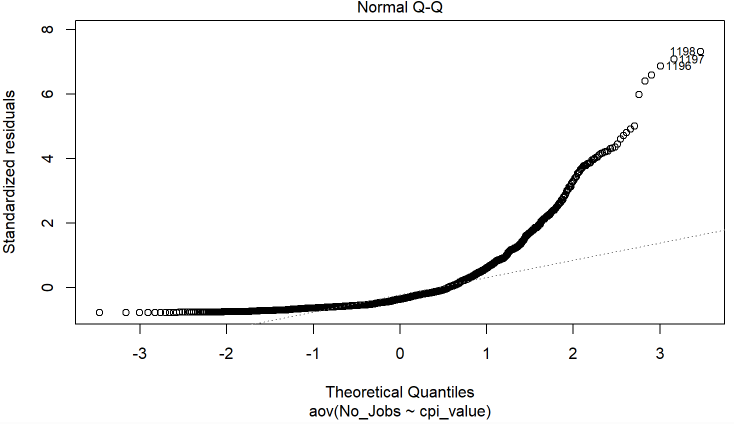
When conducting a residual analysis, a "**residuals versus fits plot**" is the most frequently created plot. It is a scatter plot of residuals on the *y* axis and fitted values (estimated responses) on the *x* axis. The plot is used to detect non-linearity, unequal error variances, and outliers.

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](https://www.statisticshowto.com/quantile-definition-find-easy-steps/)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.**Fig. 2**



**Fig. 3**

# Hypothesis Method

# Method

Results   
Needs to systematically and clearly articulate the study findings. If the results are unclear, the reviewer must decide whether the analysis of the data was poorly executed or whether the Results section is poorly organized. The Discussion Section

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.

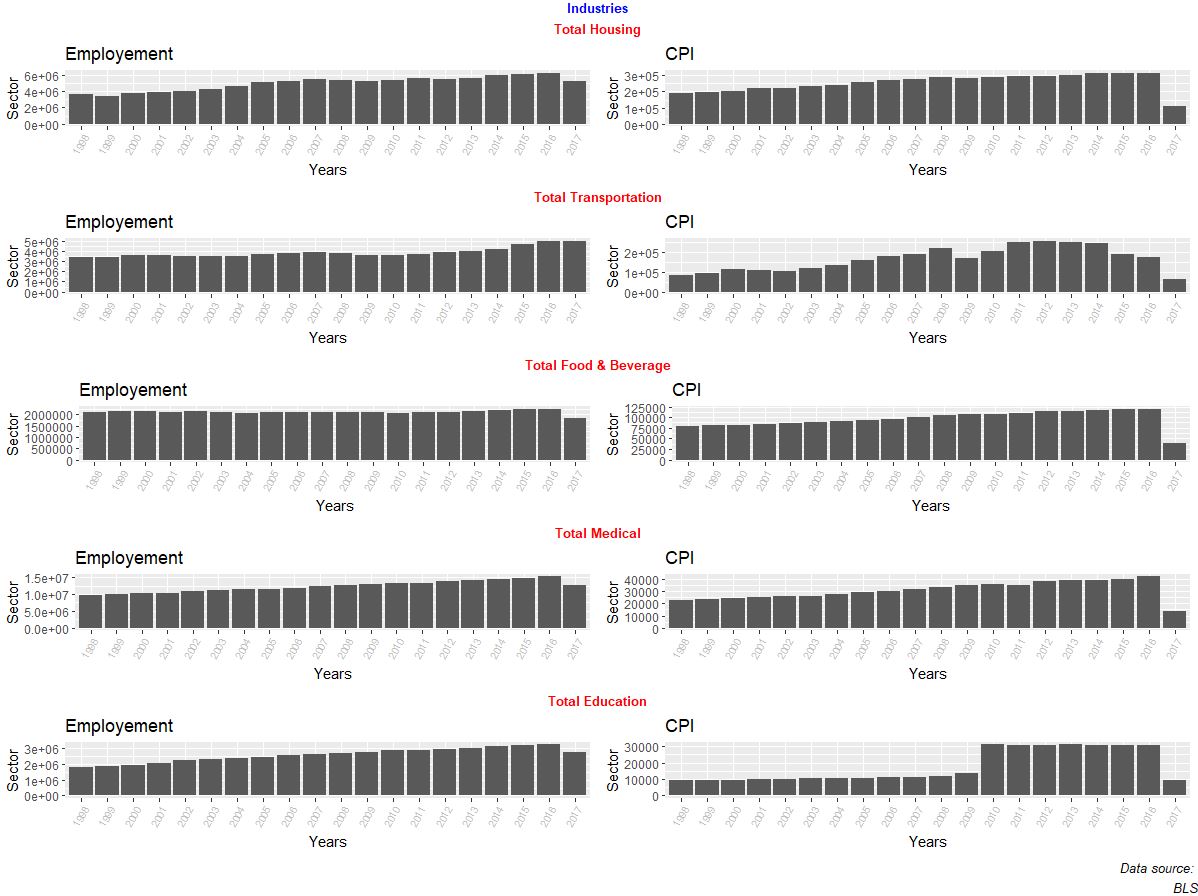
First when sampling the data for Hypothesis Testing, the extrapolation of both the Employment data and the Consumer Price Index from the U.S. Bureau of Labor Statistics, required the mapping of these two metrics by primary keys. For the purpose of this paper and analysis, this was accomplished through 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*.

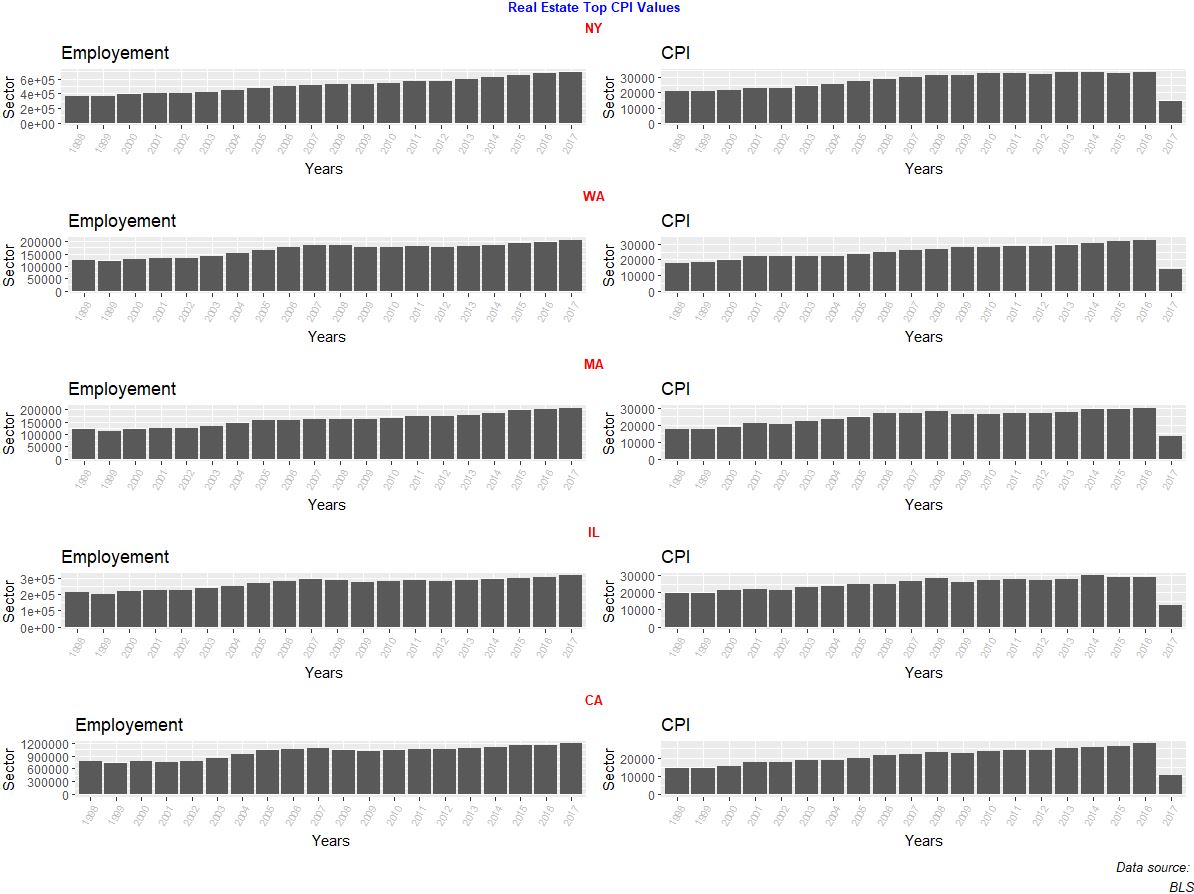
Initially analyzing the five industries sampled from the BLS, the goal was to identify the industry with the highest impact on the impact on the Consumer Price Index. According to Table 2, of the five industries sampled, Real Estate reporting a CPI value of *333,754.24* in 2014, Table 3. As shown in 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.

Upon the discovery that the *Real Estate* markethas 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.

### The Figures and Graphs

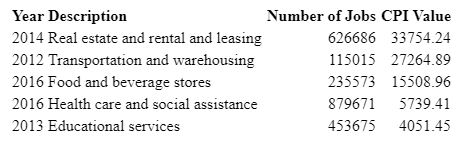
Should illustrate the important features of the methods and results. Should allow the reader to understand the figure or graph without having to refer to the text of the manuscript. Common mistakes made by inexperienced authors are failing to include figures that best depict their findings, writing unclear figure legends, and making poor use of arrows.

**Fig. 4**

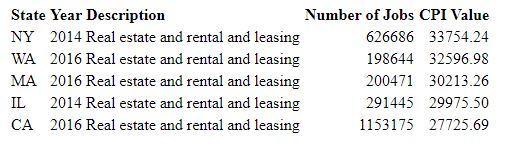
**Fig. 5**

### Tables

Should summarize the data, make the data more easily understandable, and point out important comparisons. Description of the data in the text, if possible, is preferable to the use of a space consuming table.

 **Table 2**

 **Table 3**

**Table 4**

 **Table 5**

# Discussion

### Discussion Section

Should state whether their hypotheses were verified or proven untrue or, if no hypotheses were given, whether their research questions were answered. The authors should also comment on their results in light of previous studies and explain what differences (if any) exist between their findings and those reported by others and attempt to provide an explanation for the discrepancies

# Discussion

### Hypothesis Result

### Discrepancy if any

### 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

|  |
| --- |
|  |
| Above graph shows some pattern and correlation with No. of Jobs and CPI value (Y-axis) and Year. |

|  |  |
| --- | --- |
| **lm\_cpi\_item** lm(Jobs ~ CPI+C\_Item,ALL\_train) | Multiple R-squared: 0.4133, Adjusted R-squared: 0.4114 F-statistic: 215.4 on 5 and 1529 DF, p-value: < 2.2e-16 |
| **lm\_cpi\_item\_name**  lm(Jobs ~ (CPI\*GeoName)+CPI+GeoName+  C\_Item+Year,ALL\_train) | Multiple R-squared: 0.781, Adjusted R-squared: 0.772 F-statistic: 87.59 on 60 and 1474 DF, p-value: < 2.2e-16 |
| **lm\_cpi\_item\_name2**  lm(Jobs ~ (CPI\*GeoName\*C\_Item)+  CPI+GeoName+C\_Item+Year,ALL\_train) | Multiple R-squared: 0.9797, Adjusted R-squared: 0.9765 F-statistic: 310.7 on 206 and 1328 DF, p-value: < 2.2e-16 |
| **lm\_cpi\_item\_name3**  lm(Jobs ~ (CPI\*GeoName) + (GeoName\*C\_Item)+CPI+GeoName  +C\_Item+Year,ALL\_train) | Multiple R-squared: 0.9739, Adjusted R-squared: 0.9715 F-statistic: 400.1 on 131 and 1403 DF, p-value: < 2.2e-16 |
|  |  |

**AIC check on multiple linear Model, shows ho**

dAICc df weight

lm\_cpi\_item\_name2 0.0 208 1

lm\_cpi\_item\_name3 191.5 133 <0.001

lm\_cpi\_item\_name 3296.7 62 <0.001

lm\_cpi\_item 4693.8 7 <0.001

lm\_cpi2\_item 4693.8 7 <0.001

After this we used Model *lm\_cpi\_item\_name2* , 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 the support for our Alternate Hypotheses where we see some relation with Consumer Price Index (CPI) on the Job market.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Names** | **RMSE** | **Rsquared** | **MAE** | **MAPE** |
| **lm\_cpi\_item\_name** | 1.37E+05 | 7.95E-01 | 1.01E+05 | 1.368536 |
| **lm\_cpi\_item\_name2** | 6.14E+04 | 9.63E-01 | 2.67E+04 | 0.2057522 |
| **lm\_cpi\_item\_name3** | 6.07E+04 | 9.62E-01 | 3.16E+04 | 0.3190908 |
| **Simple LM** | 1.40E+05 | 7.84E-01 | 1.03E+05 | 1.304226 |
| **lm\_cpi2\_item** | 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 |

**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.**

**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.**

**The performance of the model on the test set was satisfactory, as we were able to project**

|  |  |
| --- | --- |
| **Training Model Performance** | **Test Model Performance** |

|  |
| --- |
| Below graph shows how our model can predict most of the data points. |
|  |
|  |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

### Hypothesis Result

### Discrepancy if any

### Model /Prediction :

### Limitation:

Data: Big size, availability

Multicollinearity:

Missing Value:

# 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 include *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*.*

# 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 &amp; 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].