

DataFest 2018 Write Up

Scott Cohn, Chester Moses, Dan Brickell, Dennis Hofmann, Gurudeep Machupalli

3/25/2018

Introduction

In macroeconomic theory, the Phillips Curve represents an inverse relationship between rising wage rates and corresponding rates of inflation. This tradeoff is often visible in the short-run. In the medium to long-run, this relationship is not as evident. In the long-run of the economy, the inflation rate is represented graphically as a vertical line over the natural rate of inflation indicating no effect. Typically, this relationship is used by central banking agencies to dictate how to adjust inflation rates given expectations and unemployment levels. We tested how the variables, including several other economic indicators, affected the job market.

Data

The data was given by career website Indeed.com. Indeed is the biggest job search engine. The provided set had 4.5 million rows and 27 variables. Some of these jobs are aggregated from third-parties, while others are input directly. Each row represents the information about a job on a particular day. Each column was a piece of information relevant to each job. Some jobs were posted multiple times under the same job identifier value.

Methods

The data was imported and a random sample of 75000 rows was selected. From this sample, it was filtered by industry and saved into a smaller sample. The industry specific dataframe trimmed to contain only dates and number of job applicants. This was binded with log-based monthly economic indicator data from the Federal Reserve Economic Data (FRED). Outliers were pulled in through our log-log model. However, we did test for outliers using the minimum value ellipsoid and robust residual tests. Given our large sample size, the few outliers had no significant affect on our analysis. We then regressed these variables on $\ln(\text{Applies})$.

```
# lm(lnapply ~ UNRATE + INFLATION + PAYEMS + PCEPILFE + IceCreamSales, data = x)
```

We ran multiple regressions using different sample sizes across different industries. We then utilized ANOVA tests to remove insignificant variables. We later got excited by the prospect of novel correlations and tested the change in applies against Industry Sales — specifically against the change of sales of hard ice cream nationally.

Conclusion

Our explained variance, measured by R^2 was less than one percent. The unemployment rate was seen as a significant variable with a p-value less and 0.05. This was likely due to overfitting our model given the degrees of freedom we had. These economic indicators that suggest movement in the labor market did not have any measurable affect on the number of job applications in our dataset. There is no effect in the short run on Indeed job application numbers given the indicators we tested against.