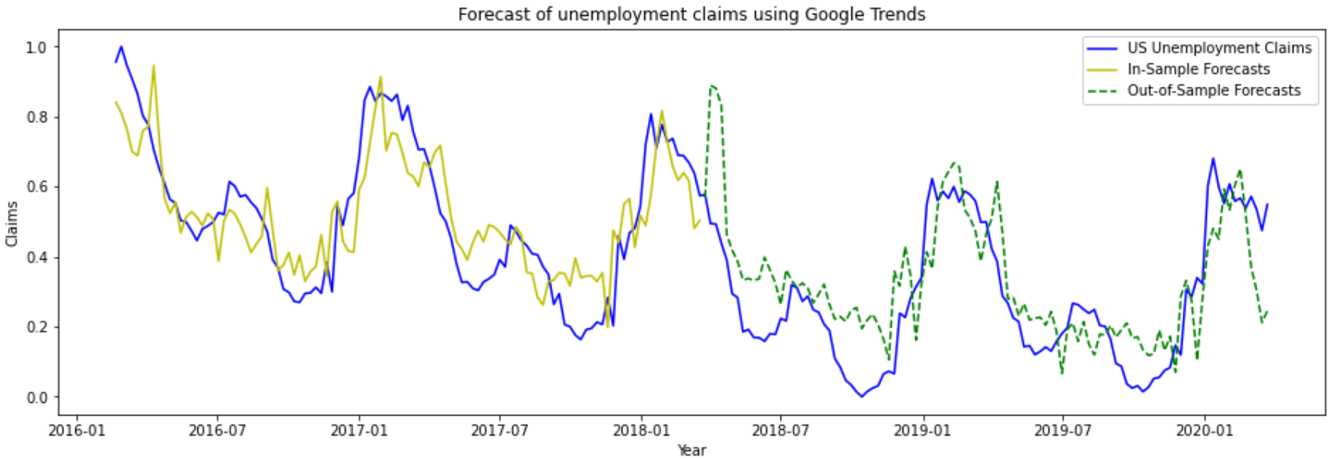
# Nowcasting Using Google Trend

*Group Members: Yiran Sun, Yuxi Ma, and Cristian Martinez*

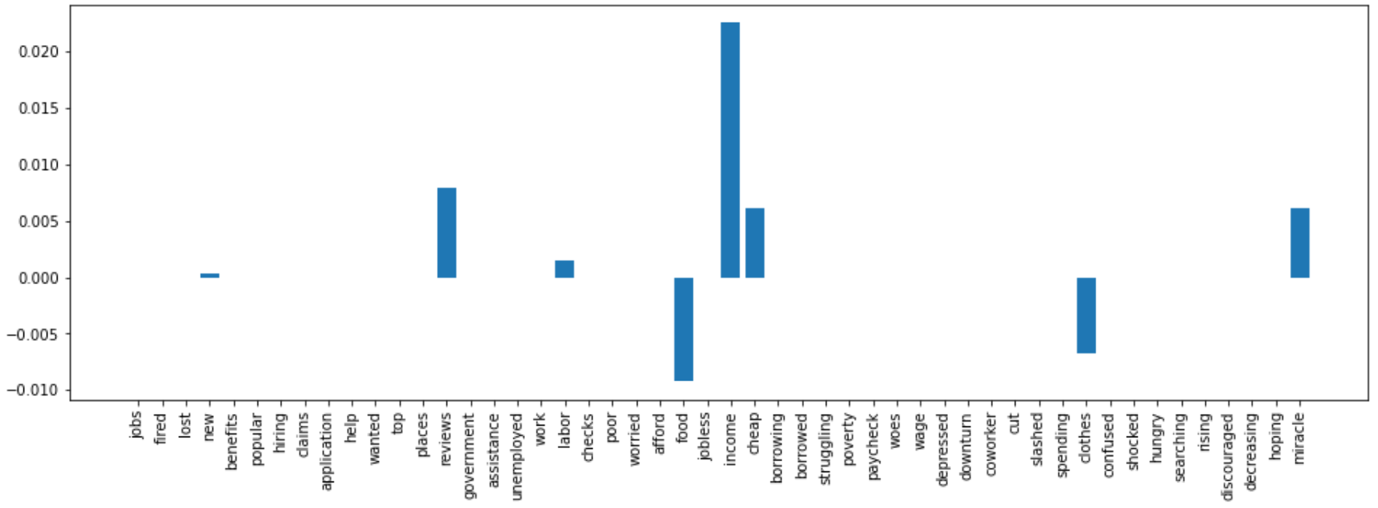
**Overview**: In this report, we focus on nowcasting the weekly US unemployment claims during Feb 2018 and Mar 2020, with the complete dataset covering period from Feb 2016 to Mar 2020. In order to achieve the nowcast, we first adopt both the *GloVe Word Vectors* and *Twitter* *API* to identify the most relevant keywords. Then, we leverage *Google Trend* to get the corresponding search data. Finally, we fit the cleaned data to a *Lasso Model* and evaluate its performance both in and out of sample.

**Model Evaluation**

**1. Graph our Lasso nowcast and the series**

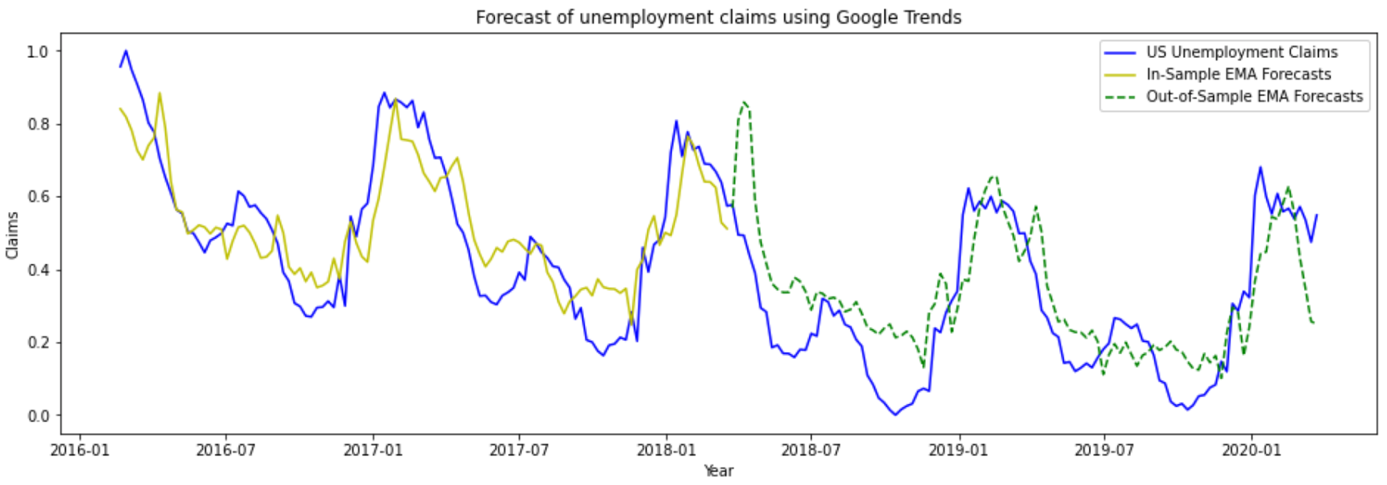


**2. The performance of keywords**



As the graph suggests, words like *reviews, food, income, cheap, clothes* and *miracle* outperform others, which may suggest that people’s desire for essentials like food and clothes (*cheap* may also count) contributes more to the depiction and reality of being unemployed.

**3. Fit an EMA to our nowcast**



**4. State whether or not the series reverts to the nowcast (Challenge: prove analytically)**

From the two nowcast graphs we can see that the prediction does not converge with the true value, which may suggest it does not revert to the nowcast.

**5. In sample MSE and Out-of Sample MSE**

|  |  |  |
| --- | --- | --- |
|  | **In-Sample MSE** | **Out-of-Sample MSE** |
| **Lasso** | 0.013286 | 0.020809 |
| **EMA** | 0.011852 | 0.019592 |

As mentioned previously, we can see from the table above that overall, EMA performs better than Lasso as it has lower both in-sample and out-of-sample MSE.

**Conclusion**

Based on the results we derived, it is apparent that the keywords the Lasso Model identified seemed to be words that describe the physical toll that individuals experience when they are fired, laid off, or their employer ceased operations. Our Lasso Model performed well in forecasting what the weekly unemployment claims would be based on the Google Trend data collected, however, it did not forecast the external shock of the Covid-19 pandemic as it was expected (refer to the figure below). As stated before, this analysis should be revisited once the Covid-19 pandemic is over to observe how the model would perform as time goes on. Perhaps exploring other machine learning models are warranted to derive stronger results such as using SuperLearners to determine which model is the best based on the results of other models used.

