

University of Connecticut

MS in Business Analytics and Project Management

OPIM-5671- Data Mining and Business Intelligence

Professor Cruz

Forecasting Unemployment Rate

Group 4:

Abhinav Dubey

Jiaxuan Wang

Marina Suberlyak

Shaista Usman

Table of Contents

Problem Statement	2
Methodology	2
Sample	2
Explore	3
Modify	4
Model 1	4
Model 2	6
Assess	7
Results	Error! Bookmark not defined.
Conclusions and Recommendations	8
References	10

Problem Statement

Unemployment rate is a metric used by governments, economists and researchers to gauge the health of the labor market and the economy. For clarity, unemployment rate is calculated by expressing the number of unemployed persons as a percentage of the total number of persons in the labor force.

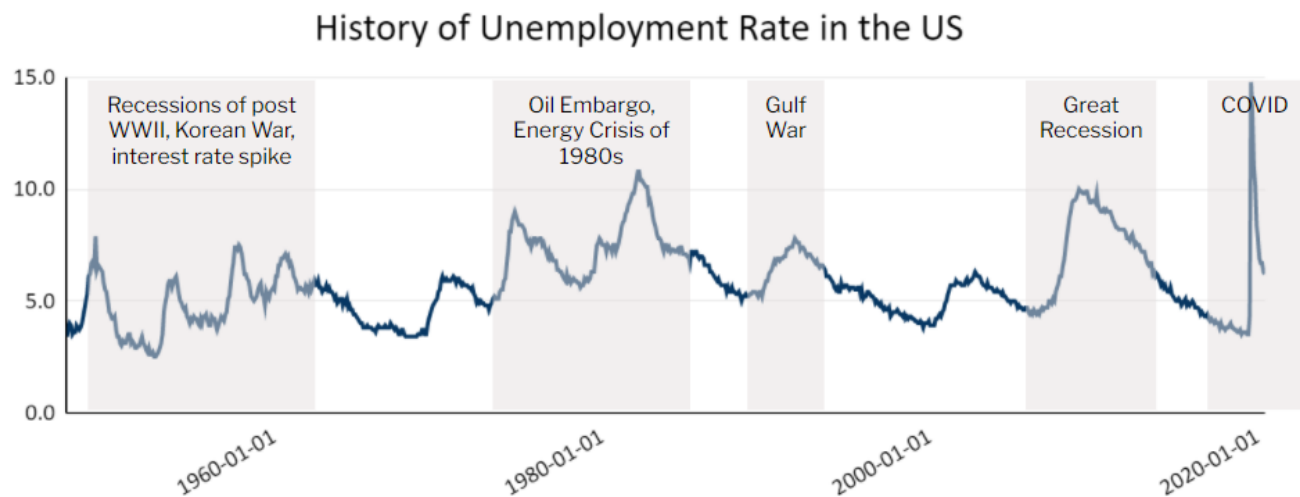
In this analysis, we will seek to model the unemployment rate in the United States by exploring trends, seasonality, events and several regressors from the historical economic data. We will then forecast near-term unemployment rate to provide an output from our model.

We expect this analysis will arm the decision-makers with unemployment rate forecasts as well as thorough understanding of its patterns and correlated variables, and equip them to devise effective fiscal policies to remedy the causes of unemployment. This is particularly critical during economic shocks such as the economic disruption during COVID-19 pandemic, where labor market participants rely on government intervention to provide a safety net, stabilize the economy and devise incentives for the markets to bounce back.

Methodology

Sample

The time series used is the official U.S. unemployment statistics¹ that are produced by the U.S. Bureau of Labor Statistics and are released on the first Friday of every month. The timeline for this data spans from 1948 to February 2021, and showcases how the unemployment rate has oscillated during a variety of historical events.

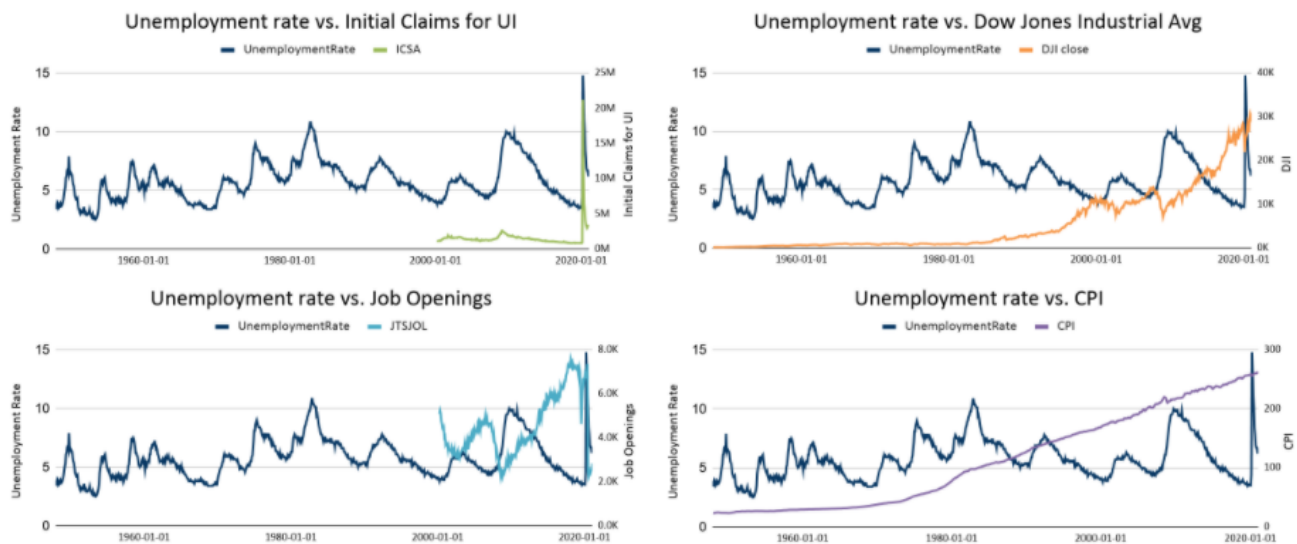


Additional related variables were also considered as regressor variables:

- 1) Job Openings and Labor Turnover Survey² - absolute job openings per month; from here on, referred as JTSJOL
- 2) Initial Claims for Unemployment Insurance³ - absolute jobless claims per week; from here on, referred as ICSA
- 3) Dow Jones Industrial average⁴ as measure of stock market performance - close level on last day of month; from here on, referred as DJI

- 4) Consumer Price Index⁵ as measure of inflation - average price of a basket of goods indexed against same basket price of 1982-84; from here on, referred as CPI

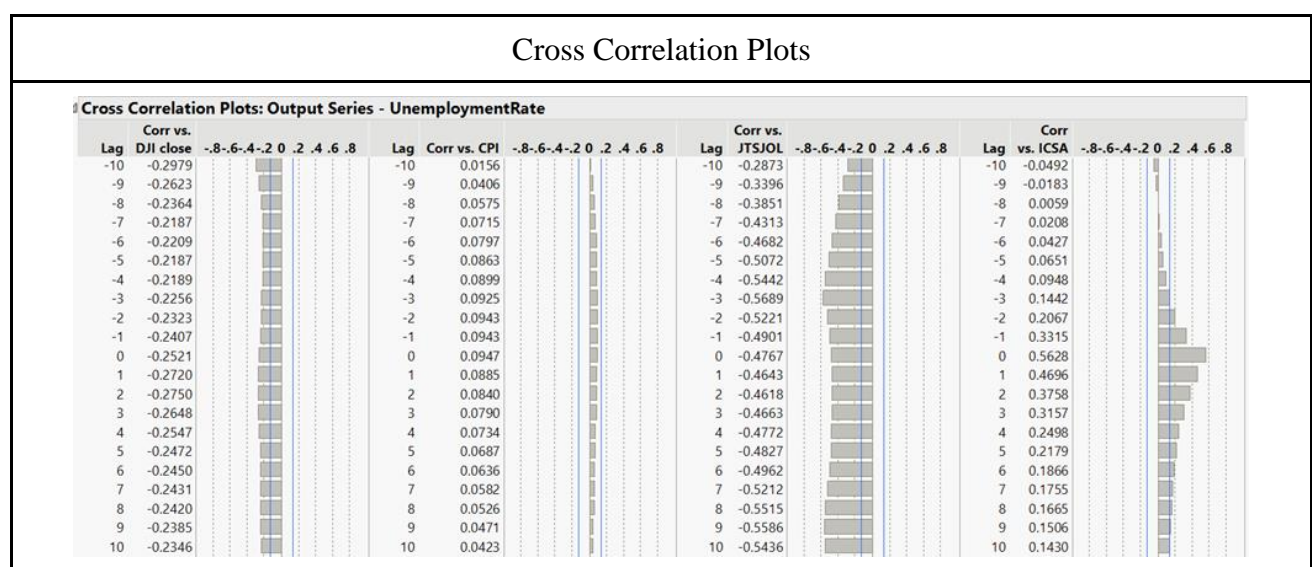
A visual representation of the variables vs. unemployment rate is below:



Explore

During the sampling and research phase, eleven unemployment rate spikes and subsequent drops were identified and explained by historical events such as wars, interest rate mistakes, and crises. Using significance tests on the parameters, only the spike of 2020 driven by COVID-19 pandemic was worthy to include as an *event intervention* in the modeling.

During this stage, *regressor variables* were also evaluated in cross correlation to unemployment rate. Visual inspection and cross-correlation plots indicated only JTSJOL and ICSA were promising as regressors due to significance and spikes around lag 0 and will be attempted during modeling.



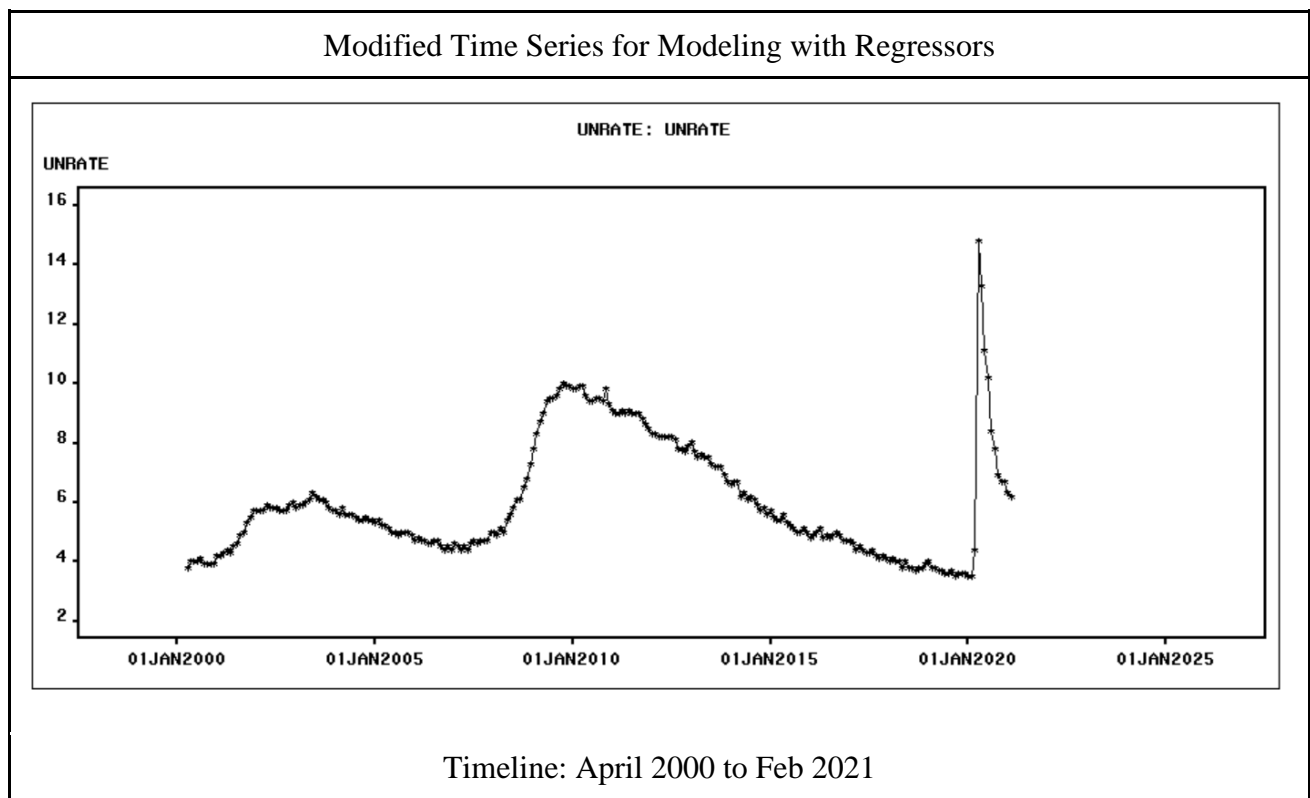
Importantly, the two significant regressors introduce the necessity to shorten the unemployment time series. As a consequence, we chose to break our analysis as follows:

- (1) we consider the entire time series and model without the regressors; event interventions, trend, seasonality, and irregularity are all explored. This is called Model 1
- (2) we modify the unemployment time series to align with the period where the regressors are available. Specifically, only years 2000-2021 are modeled and in addition to trend, seasonality and irregularity, we include the regressors. This is called Model 2

We compare models for performance and recommend the better approach in our Results section.

Modify

To accommodate analysis with regressors, we perform the following modifications to our dataset: (1) the historical data on the time series is pruned to align with the regressor variables where year 2000 onwards is available; (2) the weekly data for Initial Claims for Unemployment Insurance (ICSA) is aggregated over a period of 4 weeks to align with the available monthly unemployment data.

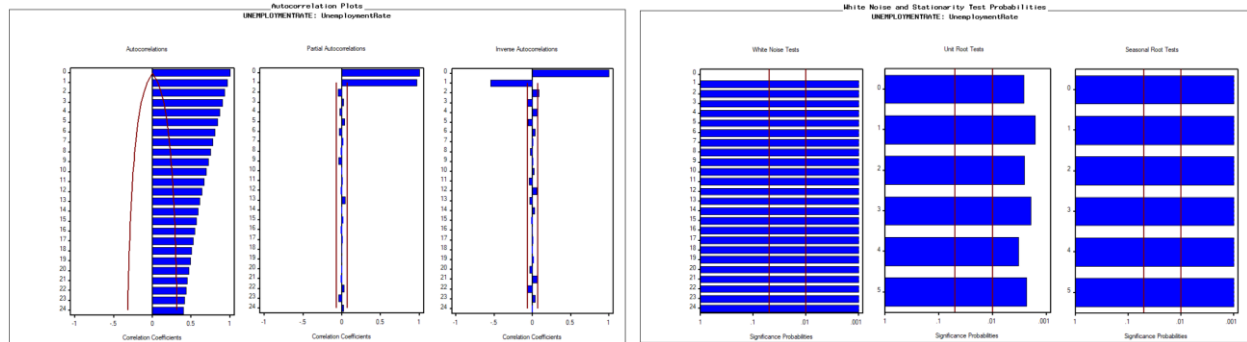


Model 1

Model 1 includes all the events and the timeline from January 1948 to February 2021.

Examining the autocorrelation plots, we observe decaying significance post lag 1 in ACF, and no repetitive behavior over S time units in PACF and IACF. Unit Root test shows significance, denying presence of trend. We remain conscious that we have multiple events, which may be breaking the trend. We conclude the series is non-stationary, with possible trend, irregularity and no seasonality .

Stationary and White Noise Tests



After capturing events as interventions, trialing several trend equations and modeling irregularity with autoregressive (AR) and moving average (MA) parameters, we narrow down the models to two where all parameters are significant: COVID + AR(2) and Linear Trend + COVID + AR(2). We choose **COVID + AR(2)** as our final Model 1 because it yields slightly better RMSE and SBC, accounts for all irregularly as supported by White Noise test, autocorrelation plots show no significant lags and residual plot has random peaks. It also has less complexity than all others and is therefore more operationally and intellectually appealing.

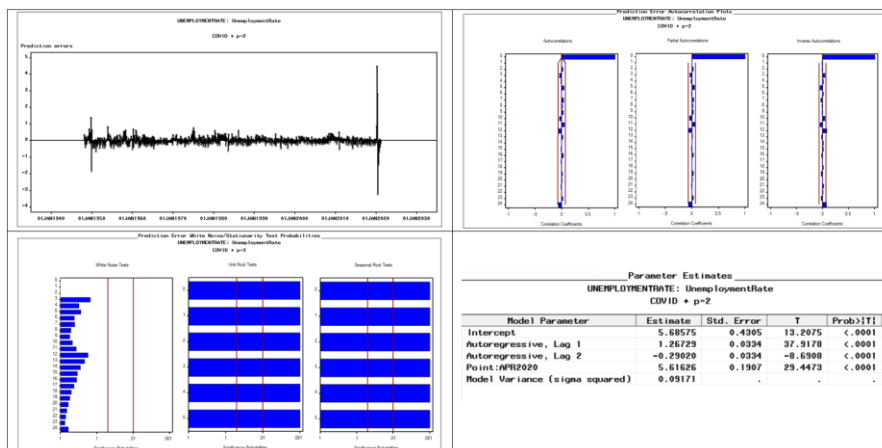
Modeling Trend, Events and Irregularity

Forecast Model	Model Title	Root Mean Square Error
<input type="checkbox"/>	Linear Trend + Historical events X 1950-60s + AR(2)	0.29736
<input checked="" type="checkbox"/>	COVID + AR(2)	0.30202
<input type="checkbox"/>	Linear Trend + COVID + AR(2)	0.30567
<input type="checkbox"/>	Linear Trend + Every historical event captured as intervention	1.18238
<input type="checkbox"/>	Linear Trend + Historical events X 1950-60s captured	1.21678
<input type="checkbox"/>	Linear Trend + Historical events X 1950-70s captured	1.26928

Forecast Model	Model Title	Schwarz Bayesian Information Criterion
<input type="checkbox"/>	Linear Trend + Historical events X 1950-60s + AR(2)	-2030.5
<input checked="" type="checkbox"/>	COVID + AR(2)	-2075.3
<input type="checkbox"/>	Linear Trend + COVID + AR(2)	-2049.7
<input type="checkbox"/>	Linear Trend + Every historical event captured as intervention	423.31277
<input type="checkbox"/>	Linear Trend + Historical events X 1950-60s captured	433.05214
<input type="checkbox"/>	Linear Trend + Historical events X 1950-70s captured	493.75793

Comparison of Statistics of Fit			
Model 1: COVID + AR(2)			
Model 2: Linear Trend + COVID + AR(2)			
NAME	LABEL	MODEL1	MODEL2
MSE	Mean Square Error	0.091214	0.093437
RMSE	Root Mean Square Error	0.302017	0.305675
MAPE	Mean Absolute Percent Error	2.903763	2.961561
MAE	Mean Absolute Error	0.165624	0.167593
RSQUARE	R-Square	0.968523	0.967790
AIC	Akaike Information Criterion	-2034.411458	-2073.640309
SBC	Schwarz Bayesian Information Criterion	-2075.300870	-2049.746384

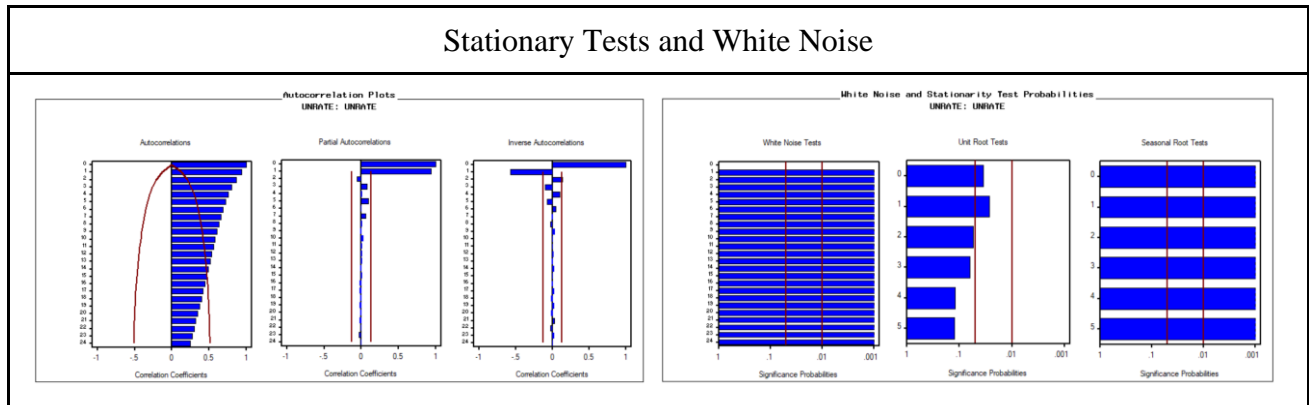
Best Model: COVID + AR(2)



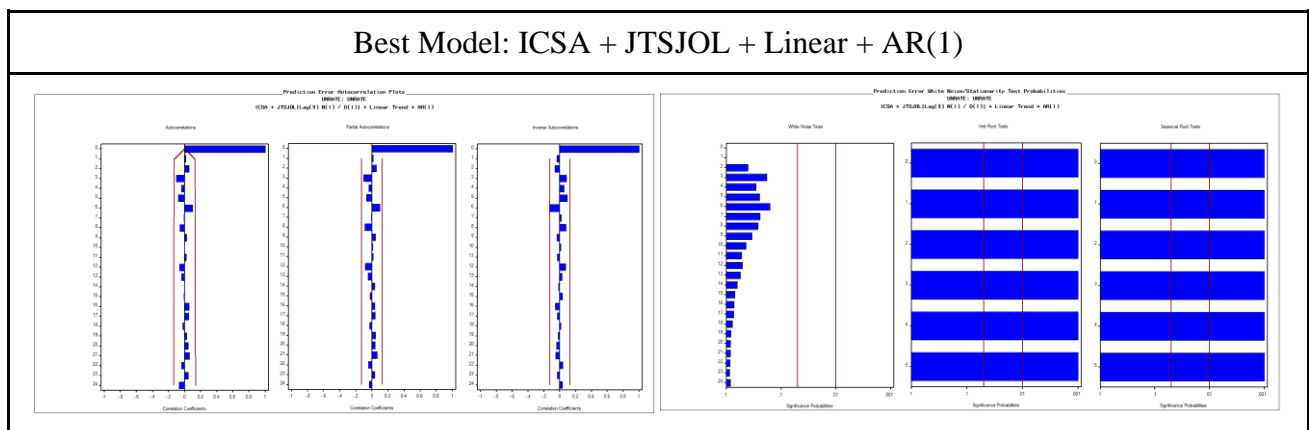
Model 2

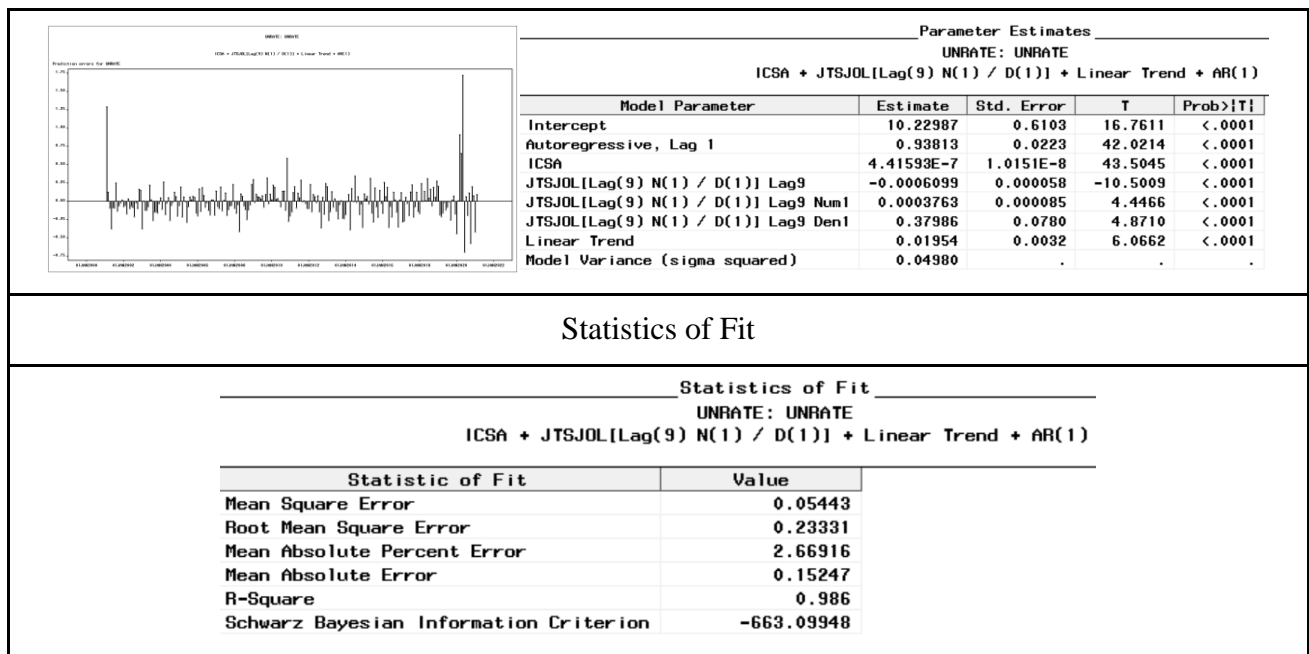
In Model 2, we are analyzing a shorter time series of unemployment rate but adding regressors to improve our model accuracy. We also choose to omit event intervention as our series has a limited horizon; we will rely on regressors and irregularity modeling to capture the patterns.

Performing stationary diagnosis, ACF shows decaying significance after lag 1 and Unit Root test confirms presence of trend. PACF and IACF show no repetitive patterns at S time periods; White Noise provides evidence for irregularity. We can conclude that the series is non-stationary.



For our second modeling approach, we included the regressor variables that had promising levels of cross correlation with the unemployment rate, specifically Job Openings and Initial Claims for unemployment insurance. Including trend and the stationary components, we settle on our optimal model: **Linear trend + regressor variables + AR(1)**. The model's parameters are all significant, it passes the white noise test, unit root test and seasonal root test. The autocorrelation plots no longer have any significant lags and the residual plot has random peaks and fluctuations, further indicating that the time series is close to being white noise.

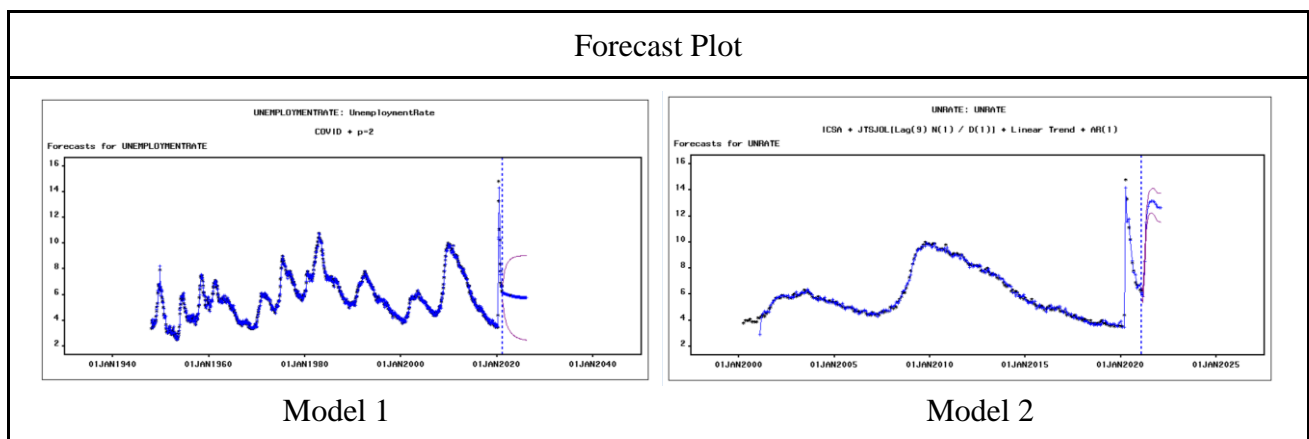




Assess

We can now assess the results of the two best models. We have already established that both models are sound, with significant parameters, captured trend and irregularity. Therefore, we focus on error, complexity, data horizon and 5-year forecasts to identify the better performing model.

Fit Statistics			
	Model 1 — Covid + AR(2)	Model 2 — ICSA + JTSJOL +Liner Trend + AR(1)	
RMSE	0.302	0.2333	
SBC	-2075.3	-663.1	
AIC	-2094.41	-687.5	
Timeline	1948-2021	2000-2021	



Penalizing for complexity, SBC and AIC indicate Model 1 is a better choice. Interestingly, forecasts for both models show very different outcomes: Model 1 has a more moderate prediction for post-COVID unemployment rate, while Model 2 is offering another spike.

Results

Evaluating Model 1 and Model 2 side by side, we believe Model 1 delivers better results and will assist decision-makers in a more comprehensive, encompassing manner than Model 2.

Firstly, Model 1 has less parameters and is therefore less complex to achieve lower SBC and AIC. It mainly relies on irregularity modeling to capture the unemployment rate time series across decades and events.

Secondly, Model 1 incorporates a longer time series and is therefore more informative for discerning the unemployment rate as a partner to and consequence of external events or fiscal actions.

Finally, the forecast of the unemployment rate in the next five years is more reasonable in Model 1 than Model 2. COVID-19 is an abrupt event and the upward slope is much steeper and extreme here than in nearly any previous event; however, history shows that the unemployment rate returns to more moderate levels gradually once the event is complete.

Conclusions and Recommendations

The unemployment rate is accepted to be a derivative of many factors. Therefore, to improve model accuracy, we believe that in addition to the trend, seasonality and critical events, other considerations, especially leading indicators for unemployment and the time horizon for additional variables are among the keys to forecasting unemployment rate.

Specifically:

1. **Appropriate interventions.** Events typically disrupt any trend or seasonality underlying the series, and the inclination is to include them as interventions to “fix” the series. However, attempting to include all events does not always add to the model, as was the case with historical events across the unemployment rate.
2. **Regressor variables.** Although in our analysis adding regressors did not produce a better model, we believe seeking out strong, informative and complete regressors will improve the accuracy of the model and future projections. This is especially true for variables which are accepted to be impacted by other variables, such as unemployment rate.
3. **Domain experts.** Domain knowledge and intuition can assist both with identifying appropriate intuition and regressor variables, as well as validate the model’s prediction for reasonableness. For example, when evaluating the forecast for post-COVID event, domain expert may be able to compile leading signals from the markets to hypothesize an economic downturn and resulting jump rather than moderately falling unemployment rate, which would make our Model 2 more appropriate.

To create conditions for an improved labor market and manage unemployment more effectively, especially during downturns, we make the following suggestions to policymakers:

1. **Change fiscal policy:** The most direct and powerful way to create jobs quickly is to change fiscal policy. In a recession, resources (both capital and labor) are idle. Therefore, Government can change the level of taxes or government spending to boost aggregate demand and create more jobs.
2. **Develop an expansionary monetary policy:** By reducing the federal funds rate and buying US Treasury and mortgage-backed securities on the open market, the Federal Reserve increases the supply of money in the economy. These tactics can spur businesses to borrow money to buy capital equipment and hire more workers.
3. **Provide education and training:** Funding education is also an effective unemployment solution. Higher levels of education increase the chance an unemployed person will emerge with a comparable wage and reduce the time required to find new employment. By providing professional vocational training and education, the unemployed individual can successfully navigate the re-employment market.

References

- ¹ U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATENSA], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/UNRATENSA>, March 19, 2021.
- ² U.S. Bureau of Labor Statistics, Job Openings: Total Nonfarm [JTSJOL], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/JTSJOL>, March 19, 2021.
- ³ U.S. Employment and Training Administration, Initial Claims [ICSA], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/ICSA>, March 19, 2021.
- ⁴ Dow Jones Industrial Index [DJI], historical data scraped and available for download at: <https://stoq.com/q/?s=%5edji>
- ⁵ U.S. Inflation Calculator, Consumer Price Index [CPI], released by U.S. Department of Labor Bureau of Labor Statistic and compiled by US Inflation Calculator resource for ease of use; <https://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/>, April 13, 2021.