Forecasting

Unemployment Rate

OPIM-5671
Data Mining and Business Intelligence

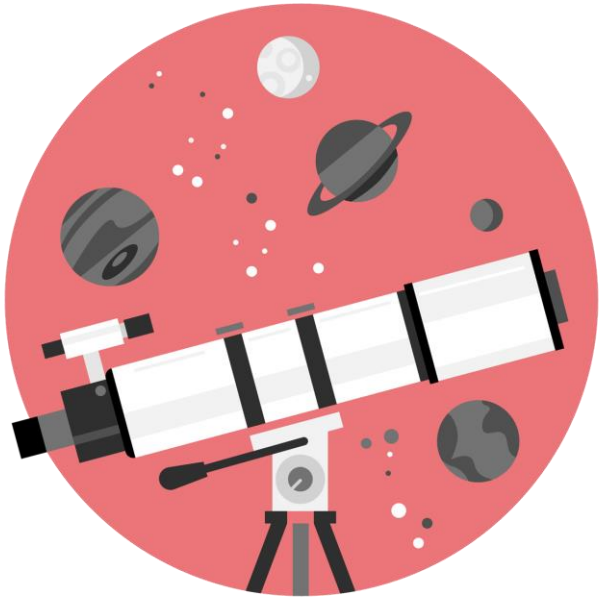
GROUP 4
Abhinav Dubey
Jiaxuan Wang
Marina Suberlyak
Shaista Usman

Table Of Contents

Background

Unemployment rate is a proportion calculated by

$$\frac{\text{number of unemployed persons}}{\text{total number of persons in the labour force}}$$



People are classified as unemployed if they fulfill the following three criteria:

- Do not have a job
- Have actively looked for work in the prior four weeks
- Are currently available for work

Problem Statement

Unemployment rate allows policy makers to examine health of the labor market and the economy. Understanding and forecasting patterns in unemployment are key to:

- inform government unemployment insurance budgets and incentives
- devise fiscal policies to remedy the causes of unemployment, particularly during economic downturns
- identify when governments should increase asset purchases and investments

Goal

We will model historical **unemployment rate** using available time series for unemployment in the United States.

The analysis will seek to model **trend, seasonality, irregularity and any useful events** from the historical unemployment rate, plus explore any regressors, and **then forecast** near-term unemployment rate to inform policymakers in devising government and economic response.

Data Description - sources

- The Official U.S. unemployment statistics [1] are produced by the U.S. Bureau of Labor Statistics, and released on the first Friday of every month
- Additional time series considered:
    - **Job Openings** and Labor Turnover Survey [2]
    - **Initial Claims** for Unemployment Insurance [3]
    - **Dow Jones Industrial average** [4] as measure of stock market performance
    - **Consumer Price Index** [5] as measure of inflation

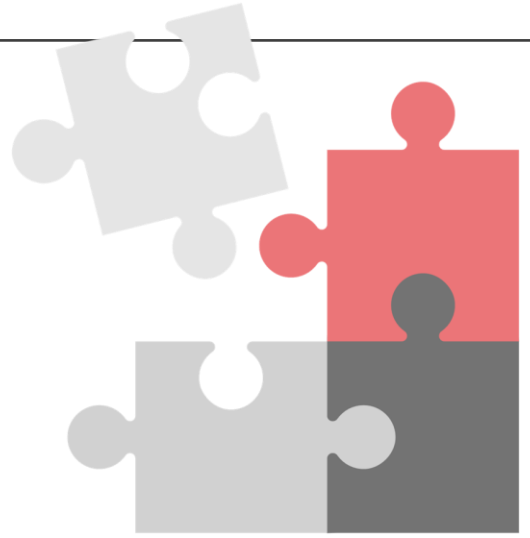
[1] Federal Reserve Economic Data | FRED | St. Louis Fed

[2] Job Openings, [3] Initial Claims

[4] Stooq, historical stock market data available for download

[5] US Inflation Calculator

Data Description - data elements

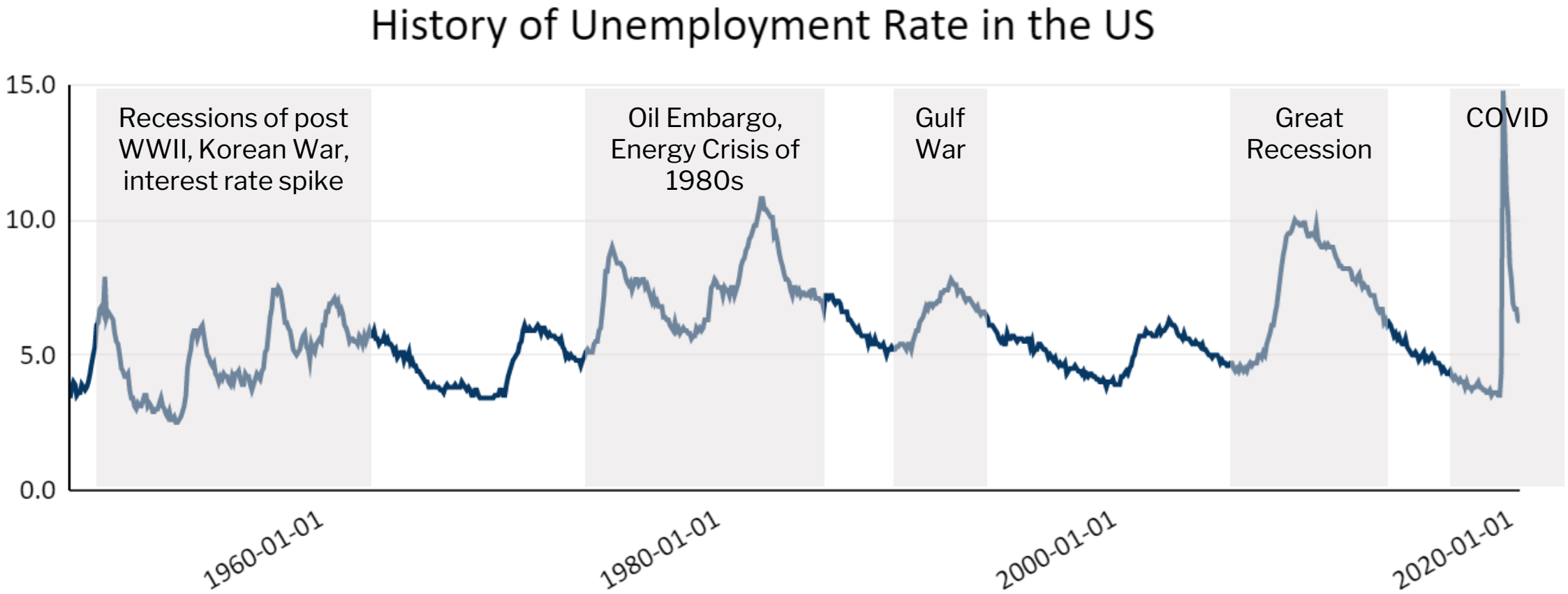


- Observation date, monthly
- Unemployment rate, as percentage of total labour force, monthly
- ICSA, as absolute jobless claims per month
- Job openings, as absolute job openings per month
- DJI, as close level on last day of month
- CPI, as average price of a basket of goods indexed against same basket price of 1982-84

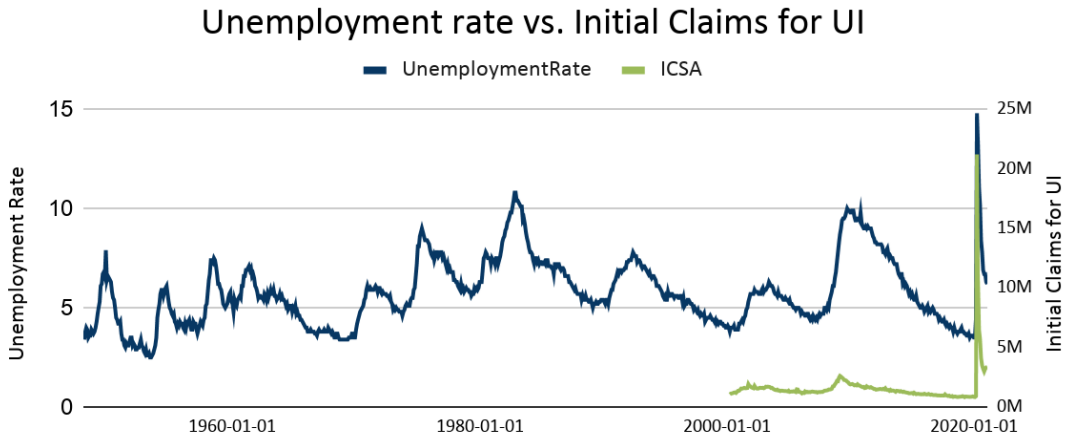
Methodology

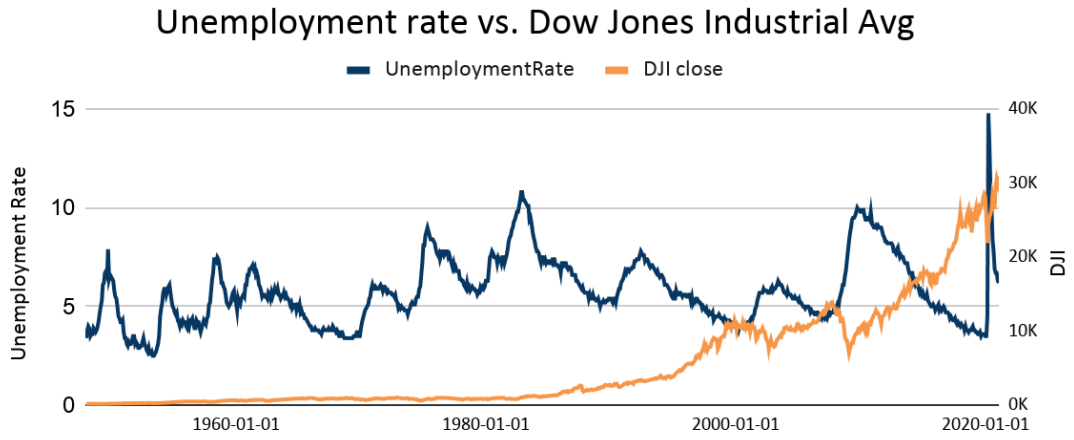
1. Sample the time series
   - visual inspection & research
2. Explore and diagnose
   - interventions, stationary analysis
3. Modify
   - define time periods to model, holdout and any missing values
4. Model and examine the results
   - attempt multiple approaches
5. Assess and conclude
   - compare models and discuss recommendations

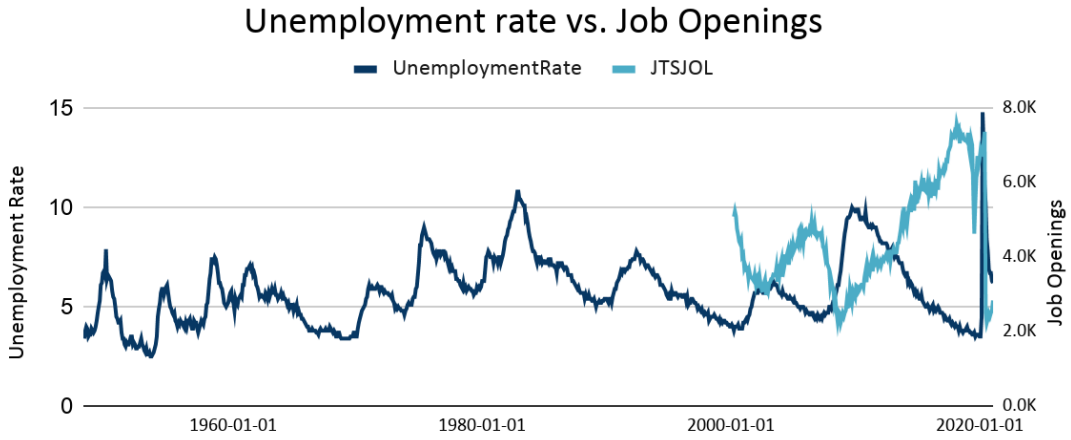
Sample - many events, or none

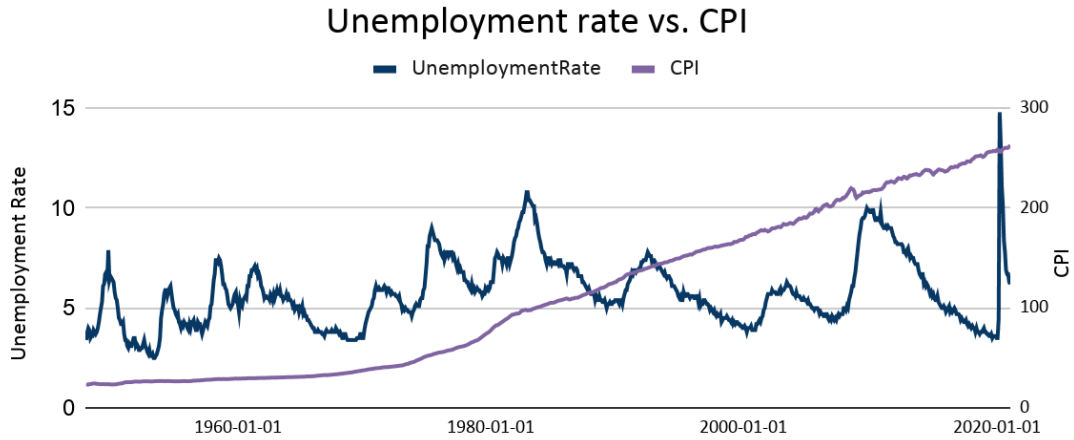


Explore - regressors









Explore - cross correlation

Cross Correlation Plots: Output Series - UnemploymentRate

Lag	Corr vs. DJI close	Lag	Corr vs. CPI	Lag	Corr vs. JTSJOL	Lag	Corr vs. ICSA
-10	-0.2979	-10	0.0156	-10	-0.2873	-10	-0.0492
-9	-0.2623	-9	0.0406	-9	-0.3396	-9	-0.0183
-8	-0.2364	-8	0.0575	-8	-0.3851	-8	0.0059
-7	-0.2187	-7	0.0715	-7	-0.4313	-7	0.0208
-6	-0.2209	-6	0.0797	-6	-0.4682	-6	0.0427
-5	-0.2187	-5	0.0863	-5	-0.5072	-5	0.0651
-4	-0.2189	-4	0.0899	-4	-0.5442	-4	0.0948
-3	-0.2256	-3	0.0925	-3	-0.5689	-3	0.1442
-2	-0.2323	-2	0.0943	-2	-0.5221	-2	0.2067
-1	-0.2407	-1	0.0943	-1	-0.4901	-1	0.3315
0	-0.2521	0	0.0947	0	-0.4767	0	0.5628
1	-0.2720	1	0.0885	1	-0.4643	1	0.4696
2	-0.2750	2	0.0840	2	-0.4618	2	0.3758
3	-0.2648	3	0.0790	3	-0.4663	3	0.3157
4	-0.2547	4	0.0734	4	-0.4772	4	0.2498
5	-0.2472	5	0.0687	5	-0.4827	5	0.2179
6	-0.2450	6	0.0636	6	-0.4962	6	0.1866
7	-0.2431	7	0.0582	7	-0.5212	7	0.1755
8	-0.2420	8	0.0526	8	-0.5515	8	0.1665
9	-0.2385	9	0.0471	9	-0.5586	9	0.1506
10	-0.2346	10	0.0423	10	-0.5436	10	0.1430

will attempt Job Openings and Initial Claims as standard regressors; DJI showed no significance during analysis

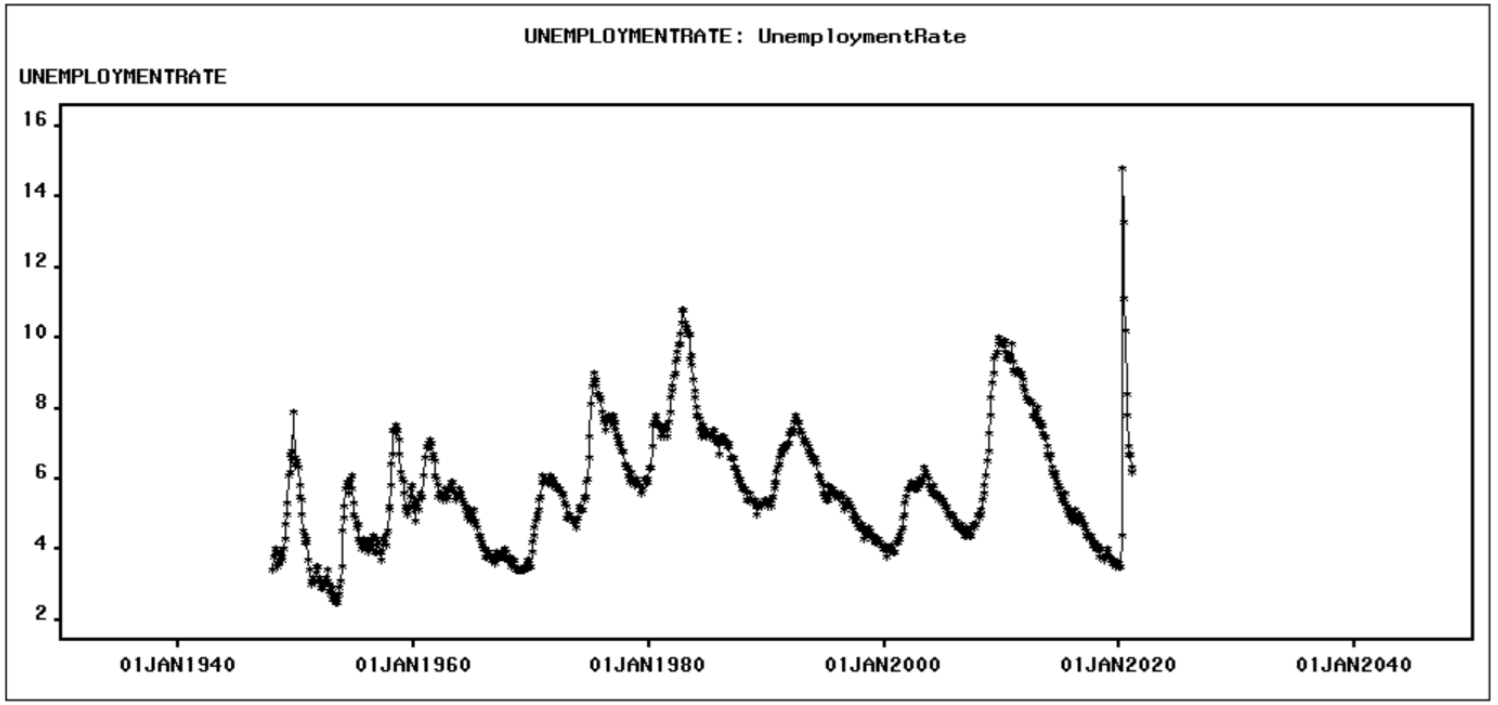
Model 1 - with events

Timeline:

January 1948

to

February 2021

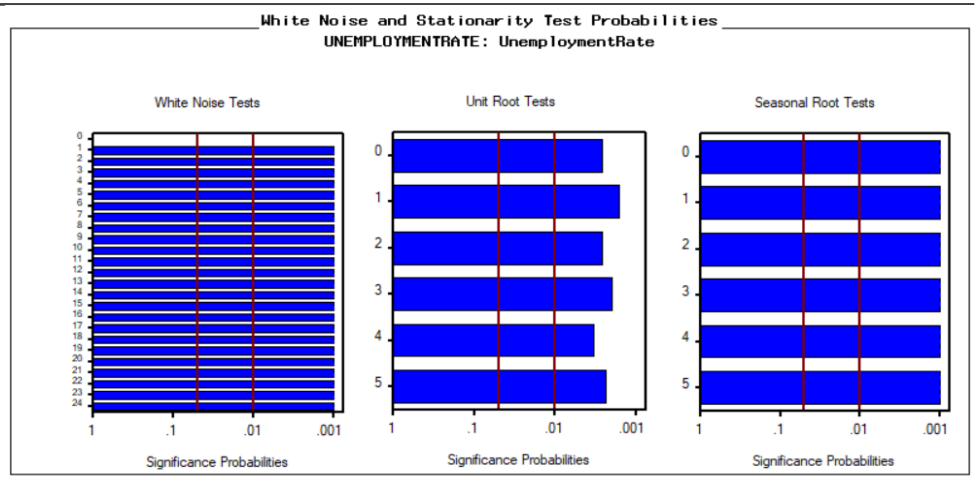


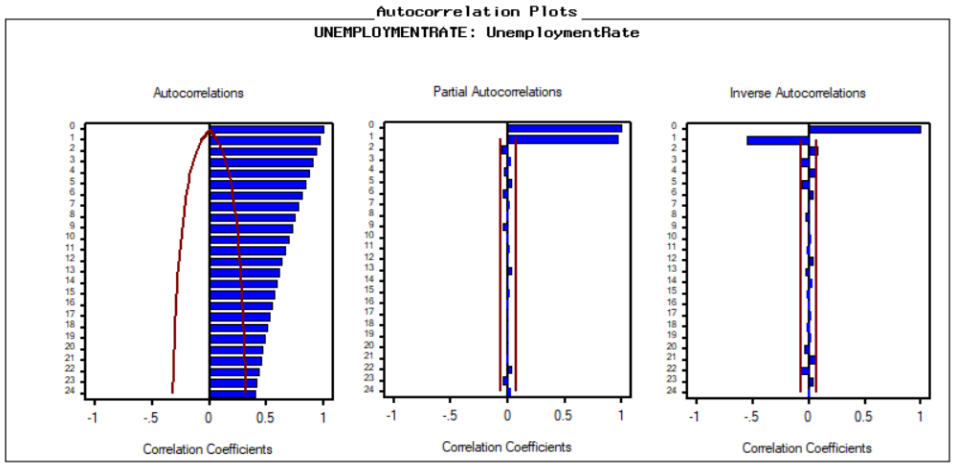
Stationary or Non-Stationary

The Autocorrelation Plots, White Noise tests and Unit Root tests indicate:

- **Presence of Trend.**
- **Absence of Seasonality.**
- The series is **not White Noise**.

Therefore, the Time Series is Non-Stationary.





Modeling Trend, Events and Irregularity

Forecast Model	Model Title	Root Mean Square Error
☐	Linear Trend + Historical events X 1950-60s + AR(2)	0.29736
☑	COVID + AR(2)	0.30202
☐	Linear Trend + COVID + AR(2)	0.30567
☐	Linear Trend + Every historical event captured as interventio	1.18238
☐	Linear Trend + Historical events X 1950-60s captured	1.21678
☐	Linear Trend + Historical events X 1950-70s captured	1.26928

Forecast Model	Model Title	Schwarz Bayesian Information Criterion
☐	Linear Trend + Historical events X 1950-60s + AR(2)	-2030.5
☑	COVID + AR(2)	-2075.3
☐	Linear Trend + COVID + AR(2)	-2049.7
☐	Linear Trend + Every historical event captured as interventio	423.31277
☐	Linear Trend + Historical events X 1950-60s captured	433.05214
☐	Linear Trend + Historical events X 1950-70s captured	493.75793

We modelled all the events from 1948 to 2020 and trend without including regressors. The best two models with all parameters significant and in limits are

1. Linear Trend + COVID + AR(2)
2. COVID + AR(2)

Model Evaluation

Conclusion: COVID + AR(2) is the best model based on RMSE, AIC and SBC.

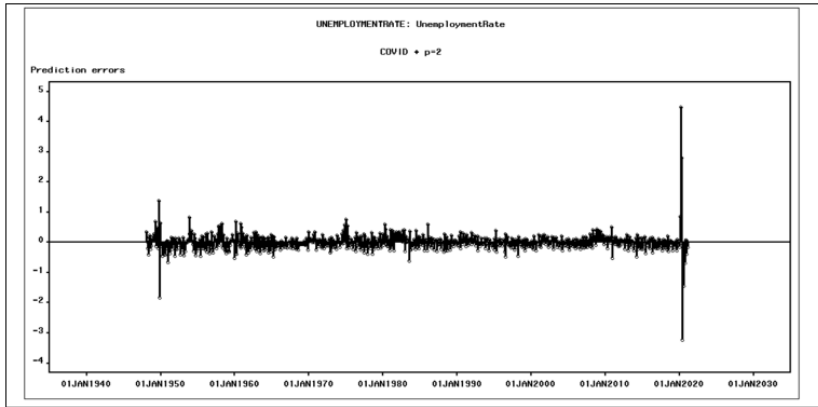
Comparison of Statistics of Fit

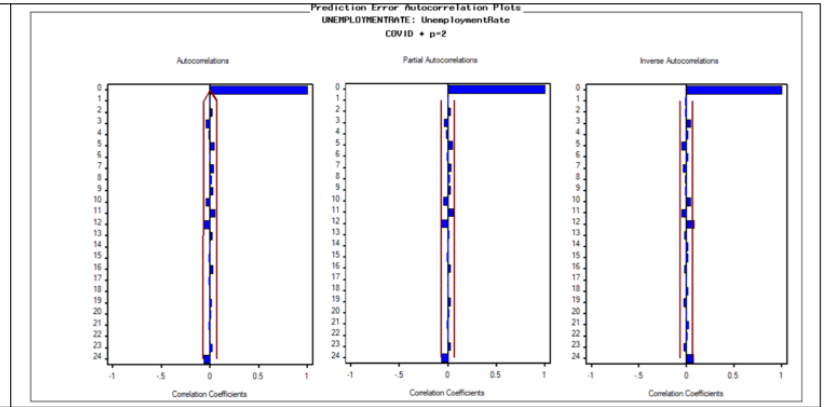
Model 1: COVID + AR(2)
Model 2: Linear Trend + COVID + AR(2)

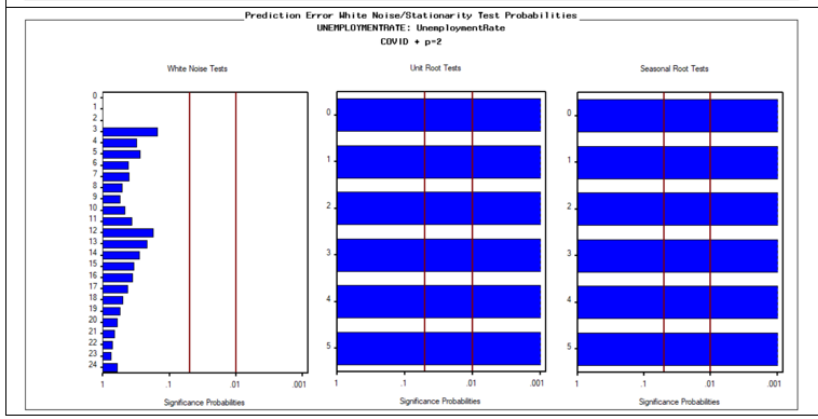
NAME	LABEL	MODEL1	MODEL2
MSE	Mean Square Error	0.091214	0.093437
RMSE	Root Mean Square Error	0.302017	0.305675
MAPE	Mean Absolute Percent Error	2.903763	2.961561
MAE	Mean Absolute Error	0.165624	0.167593
RSQUARE	R-Square	0.968523	0.967790
AIC	Akaike Information Criterion	-2094.411456	-2073.640309
SBC	Schwarz Bayesian Information Criterion	-2075.300870	-2049.746384

Model Evaluation (Cont'd)

Reasoning: model achieves random residuals, clean Autocorrelation plots and White Noise. All parameters are significant.







Parameter Estimates

UNEMPLOYMENTRATE: UnemploymentRate

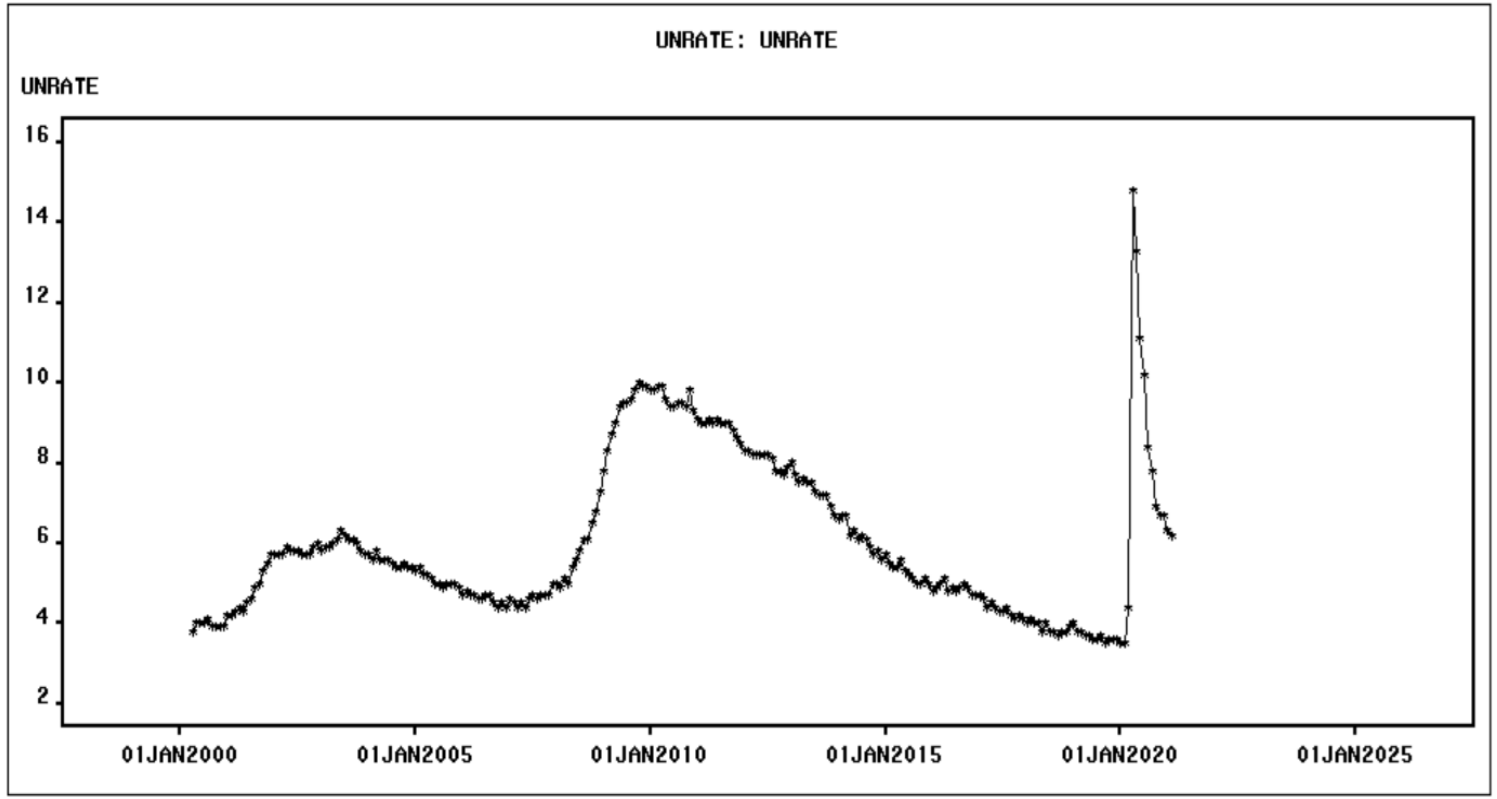
COVID + p=2

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	5.68575	0.4305	13.2075	<.0001
Autoregressive, Lag 1	1.26729	0.0334	37.9178	<.0001
Autoregressive, Lag 2	-0.29020	0.0334	-8.6908	<.0001
Point:APR2020	5.61626	0.1907	29.4473	<.0001
Model Variance (sigma squared)	0.09171	.	.	.

Model 2 - with Regressors

Timeline:

April 2000 to Feb 2021



Modify

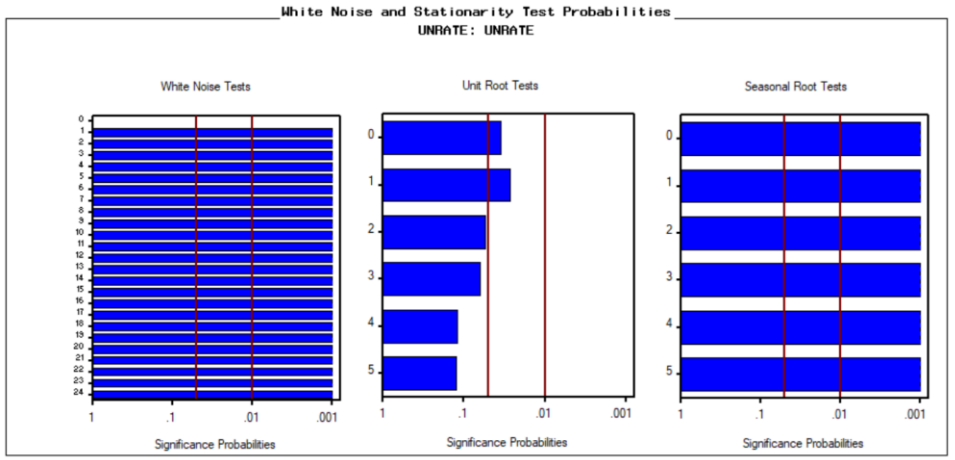
- The **historical data on the Time series was pruned**, to align it with the data for Regressors that were available from 2000 onwards.
- The weekly data for Initial Claims for Unemployment Insurance (ICSA) was **aggregated over a period of 4 weeks** (in the same way as UNRATE data is collected) to align it with the available monthly UNRATE data.

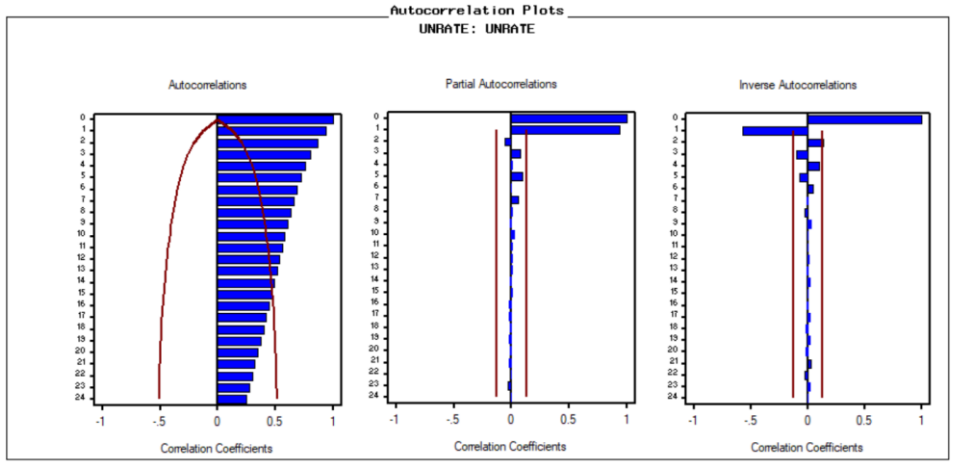
Stationary or Non-Stationary

The Autocorrelation Plots, White Noise tests and Unit Root tests indicate:

- **Presence of Trend.**
- **Absence of Seasonality.**
- The series is **not White Noise**.

Therefore, the Time Series is Non-Stationary.

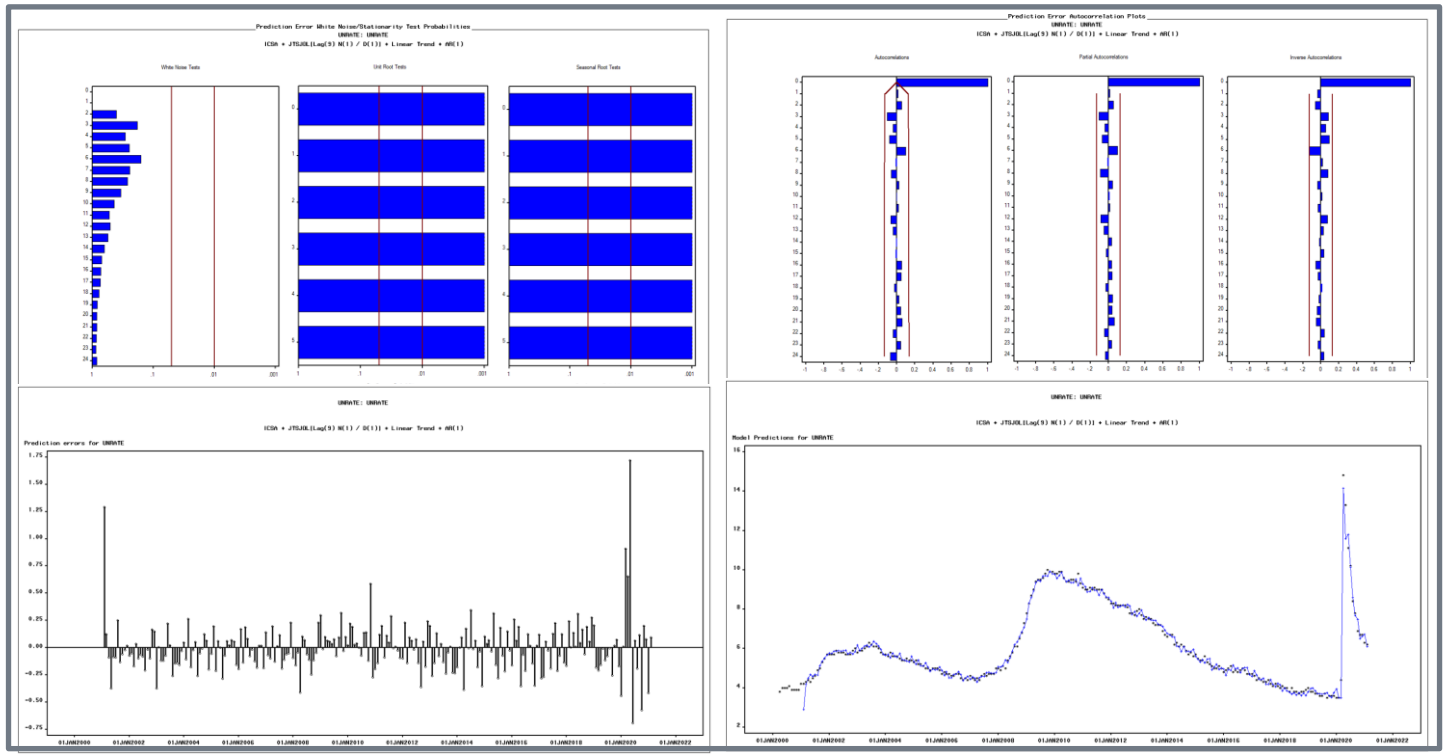




Modeling Regressors, Trend and Irregularity

Model includes:

1. **Regressors**: Initial Claims for Unemployment Insurance (ICSA) and Job Openings and Labor Turnover (JTSJOL)
2. **Trend**
3. **Irregularity** via autoregressive model



Model Evaluation

Parameter Estimates

UNRATE: UNRATE

ICSA + JTSJOL[Lag(9) N(1) / D(1)] + Linear Trend + AR(1)

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	10.22987	0.6103	16.7611	<.0001
Autoregressive, Lag 1	0.93813	0.0223	42.0214	<.0001
ICSA	4.41593E-7	1.0151E-8	43.5045	<.0001
JTSJOL[Lag(9) N(1) / D(1)] Lag9	-0.0006099	0.000058	-10.5009	<.0001
JTSJOL[Lag(9) N(1) / D(1)] Lag9 Num1	0.0003763	0.000085	4.4466	<.0001
JTSJOL[Lag(9) N(1) / D(1)] Lag9 Den1	0.37986	0.0780	4.8710	<.0001
Linear Trend	0.01954	0.0032	6.0662	<.0001
Model Variance (sigma squared)	0.04980	.	.	.

Statistics of Fit

UNRATE: UNRATE

ICSA + JTSJOL[Lag(9) N(1) / D(1)] + Linear Trend + AR(1)

Statistic of Fit	Value
Mean Square Error	0.05443
Root Mean Square Error	0.23331
Mean Absolute Percent Error	2.66916
Mean Absolute Error	0.15247
R-Square	0.986
Schwarz Bayesian Information Criterion	-663.09948

- The model has all significant parameters and a low RMSE and SBC.
- The model captures enough variance from the data while still being less complex and not overfitting the data.

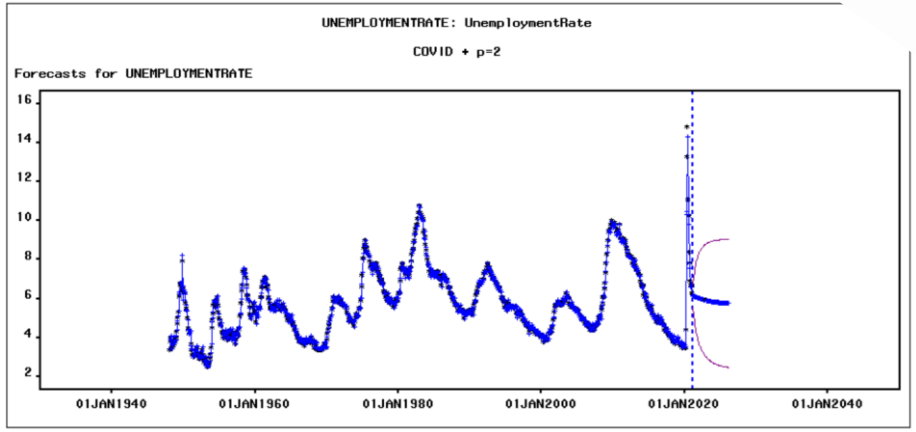
Assess

We will evaluate our 2 models on:

- RMSE
- SBC & AIC
- Complexity
- Forecast reasonableness
- Inclusion of data available

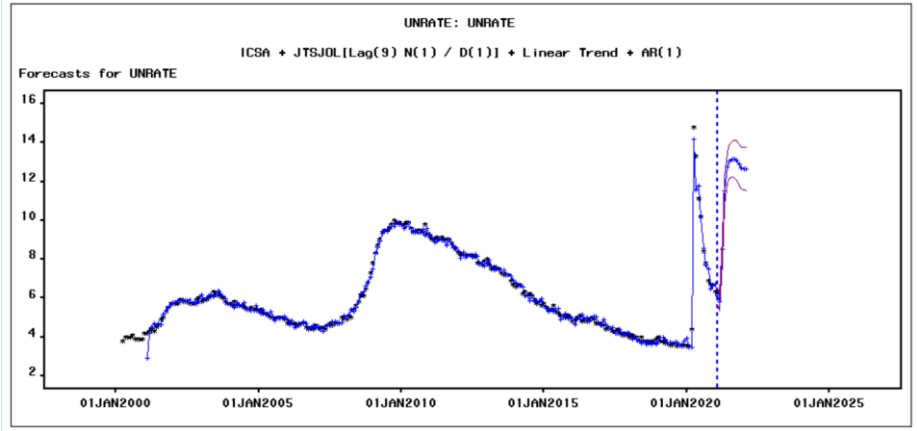
Results

MODEL 1



RMSE:	0.3020
SBC:	-2075.3
AIC:	-2094.41
Forecast:	believable
Period:	more comprehensive

MODEL 2



RMSE:	0.2333
SBC:	-663.1
AIC:	-687.5
Forecast:	unreasonable
Period:	only recent decades

Recommendations

Improve accuracy:

1. Add interventions appropriately
2. Regressors worthwhile to explore if more historical data is available
3. Validate against domain expert knowledge and intuition, e.g. post COVID forecast

Improve labor market:

1. Change fiscal policy and develop an expansionary monetary policy
2. Provide education and training to the labor force

Conclusion

Lastly, it is **critical to model unemployment rate in the context of:**

- Other economic factors, especially leading indicators for unemployment
- Ample, complete data for additional variables


THANK
YOU !