FORECASTING UNEMPLOYMENT IN THE US

Linh Nguyen, Silvia Chalkou, Taiwo Bada, Veena Iyer
MSBA, MSMA, MSDI, MSBA
Bentley University, Massachusetts

MA 611 - Spring 2021 - Dr. Moinak Bhaduri

ltnguyen@falcon.bentley.edu, schalkou@falcon.bentley.edu, bada_taiw@bentley.edu, iyer_veen@bentley.edu

Agenda

INTRODUCTION

THE DATASET

EXPLORATORY DATA ANALYSIS

MODELLING

CONCLUSION AND FUTURE WORK

Introduction

Source: Financial Times

$$\text{Unemployment Rate} = \left(\frac{\text{Number of Unemployed}}{\text{Labor Force}}\right) * 100\%$$

- ❑ Unemployment: key data to determine a nation's economic health
- ❑ Unemployment peaked to record high since Covid-19 pandemic
- ❑ Project focus:
  - ✓ Forecast unemployment in the US
  - ✓ Unemployment trend across states and gender

The Dataset

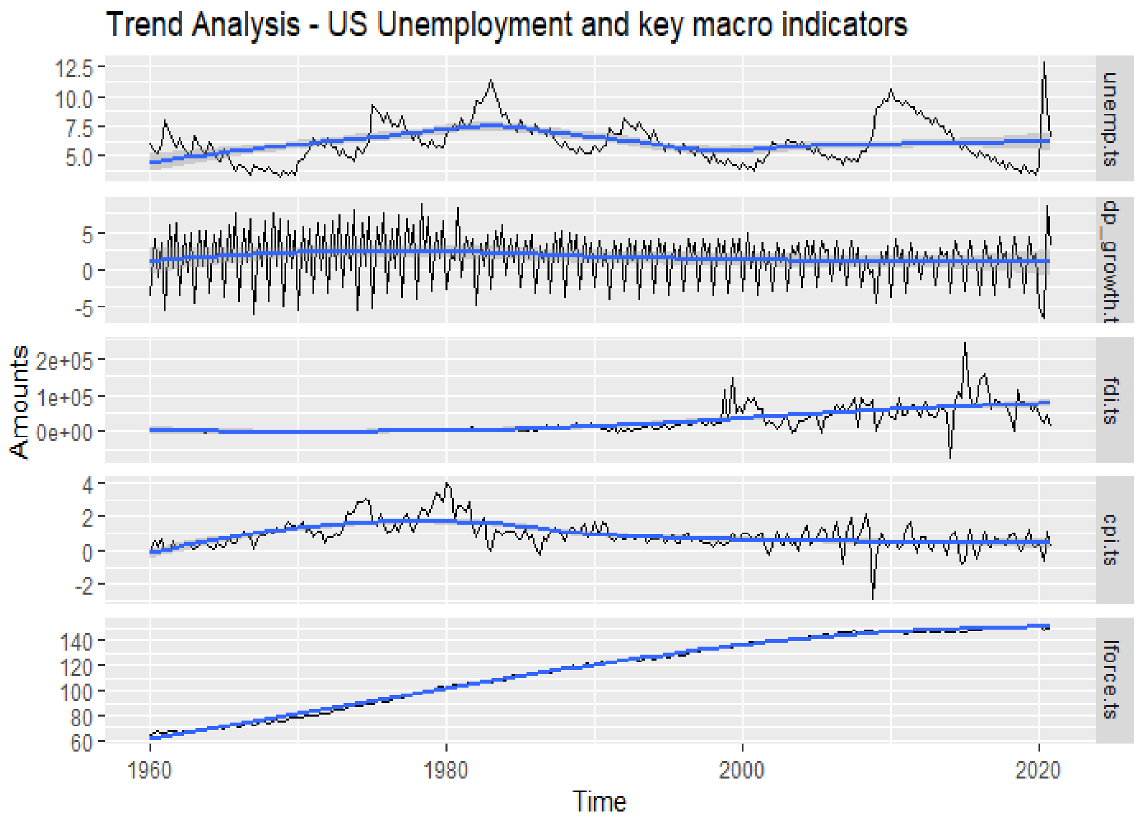
Data Source: [Federal Reserve Bank of St. Louis (FRED) Economic Data]

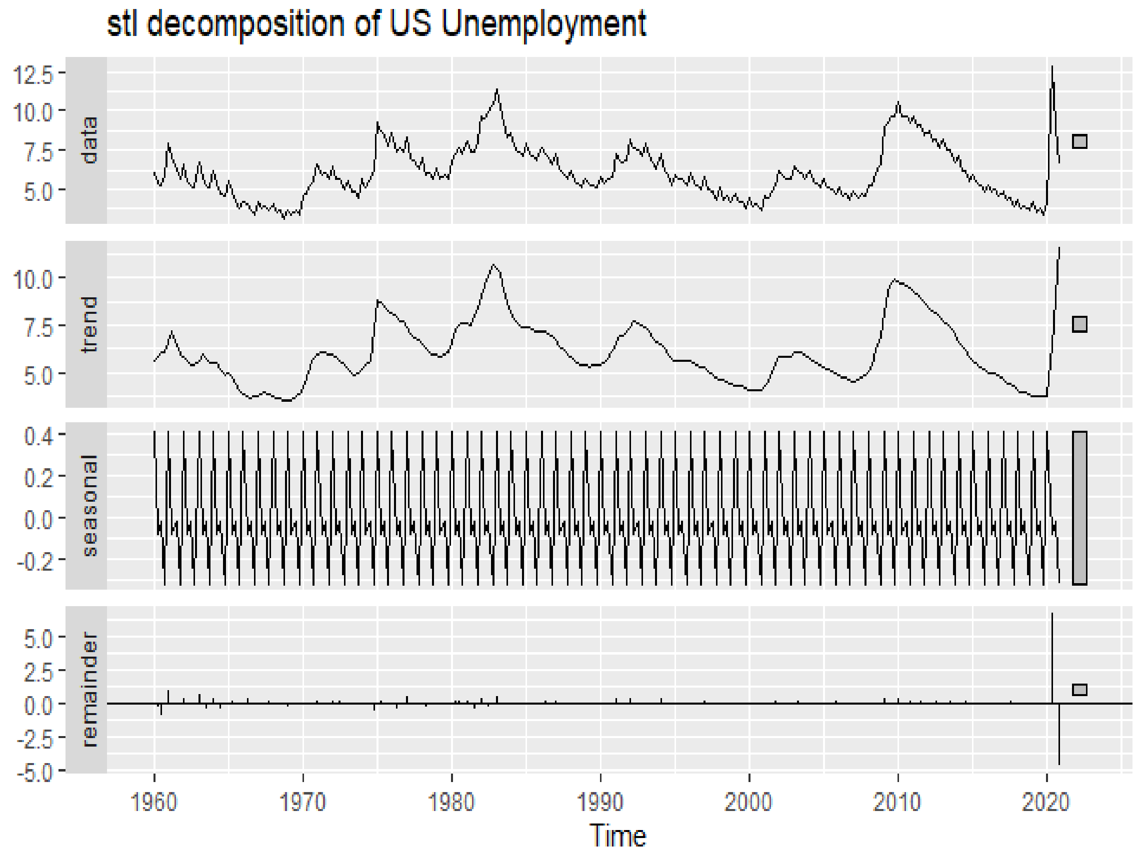
Name	Size	Duration	Frequency
Total unemployment rate in the US	244 Entropy = 1.02	1960 - 2020	4
Total unemployment rate in the US by state	69	2003 - 2020	4

Exploratory Data Analysis

- ❑ Trend Analysis
- ❑ Seasonality Check
- ❑ Data Clustering
- ❑ Similarity Check

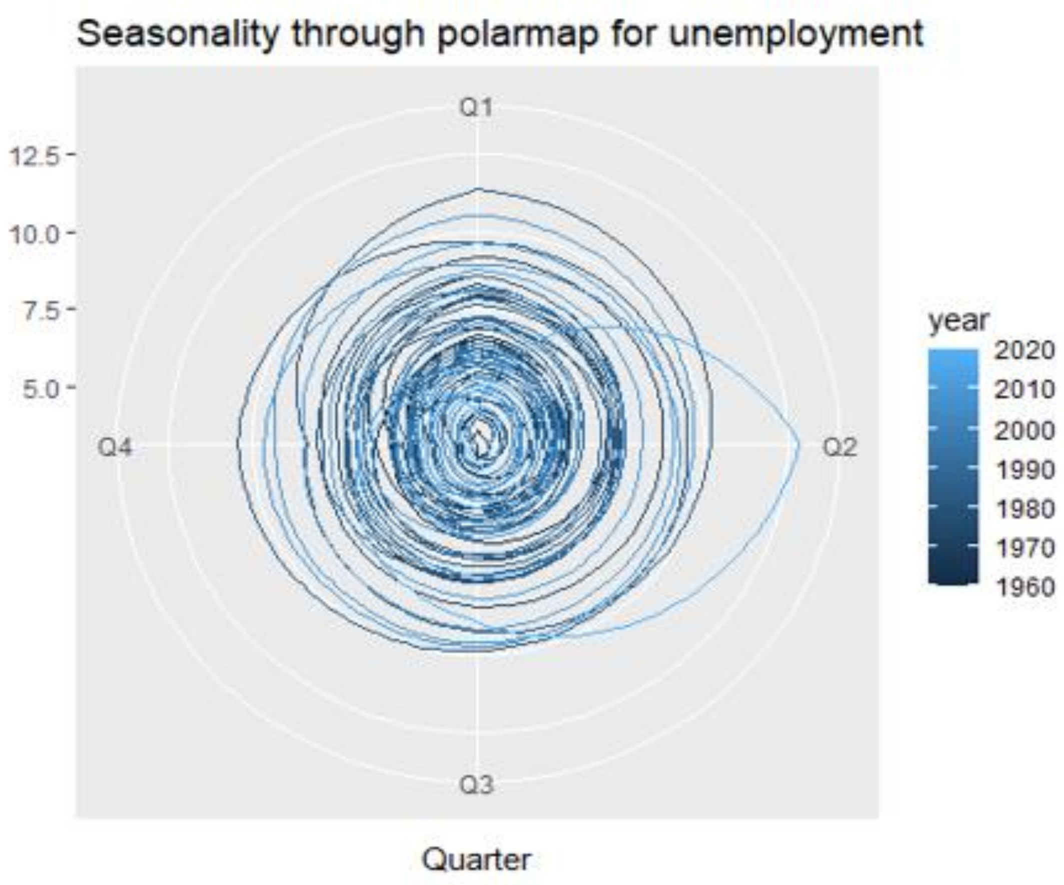
Trend Analysis & Stl decomposition

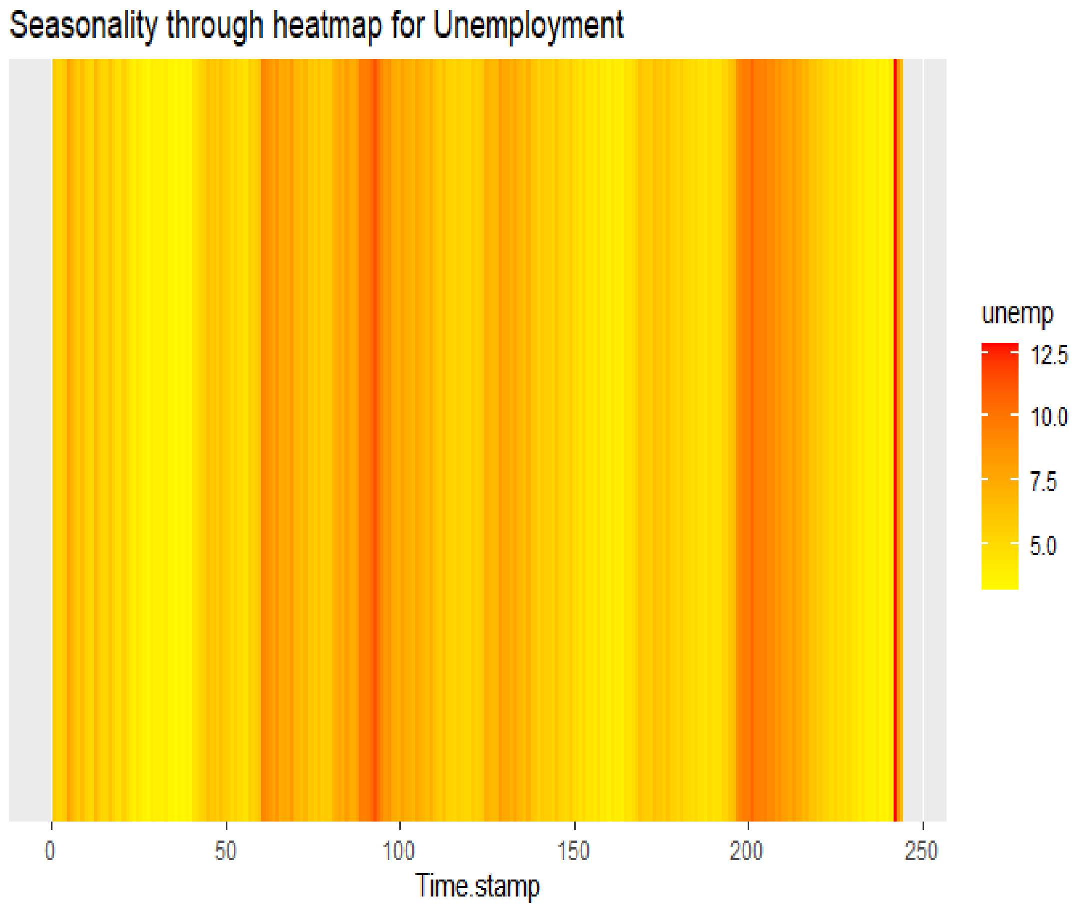




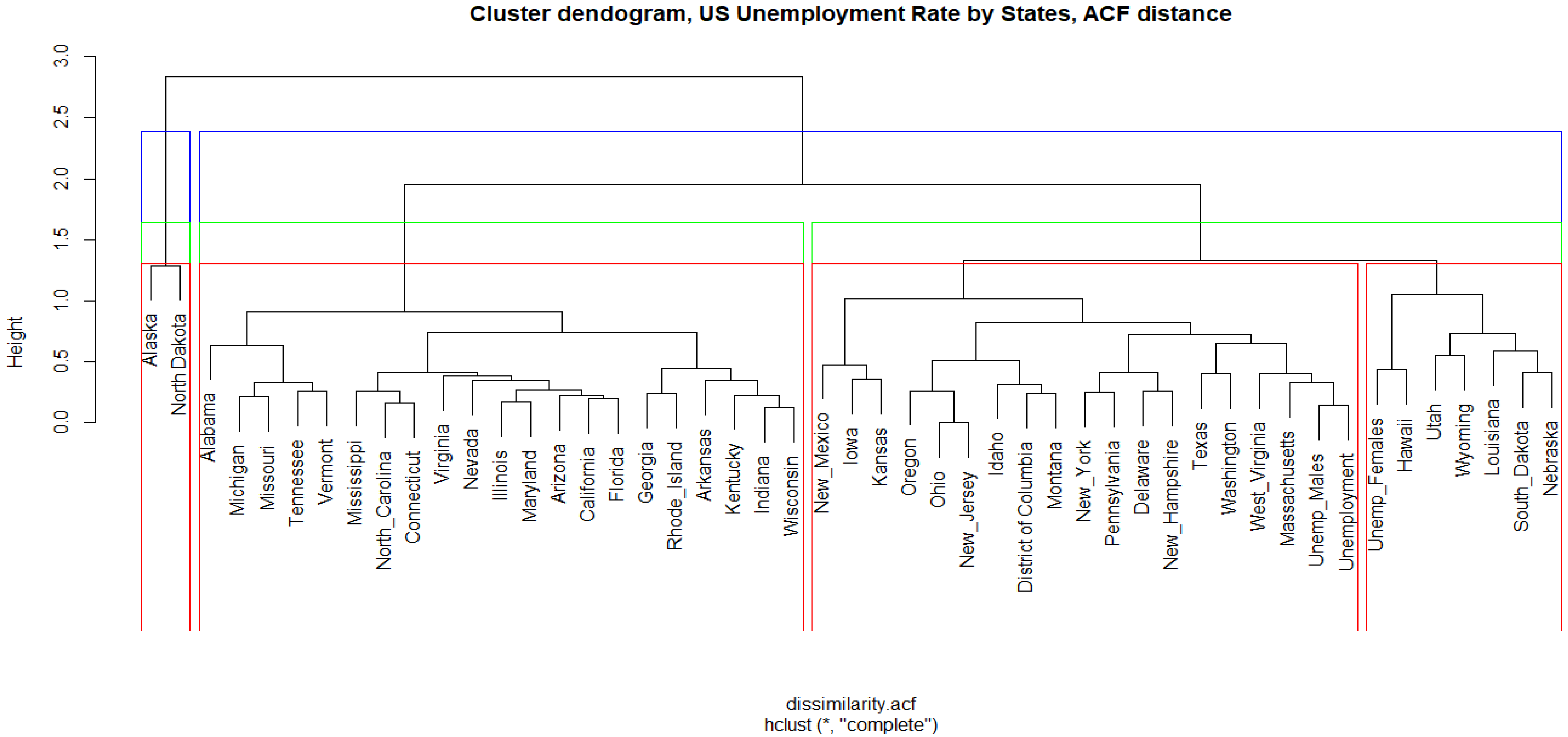
$F_t = 0.94$; $F_s = 0.31$

Seasonality Check





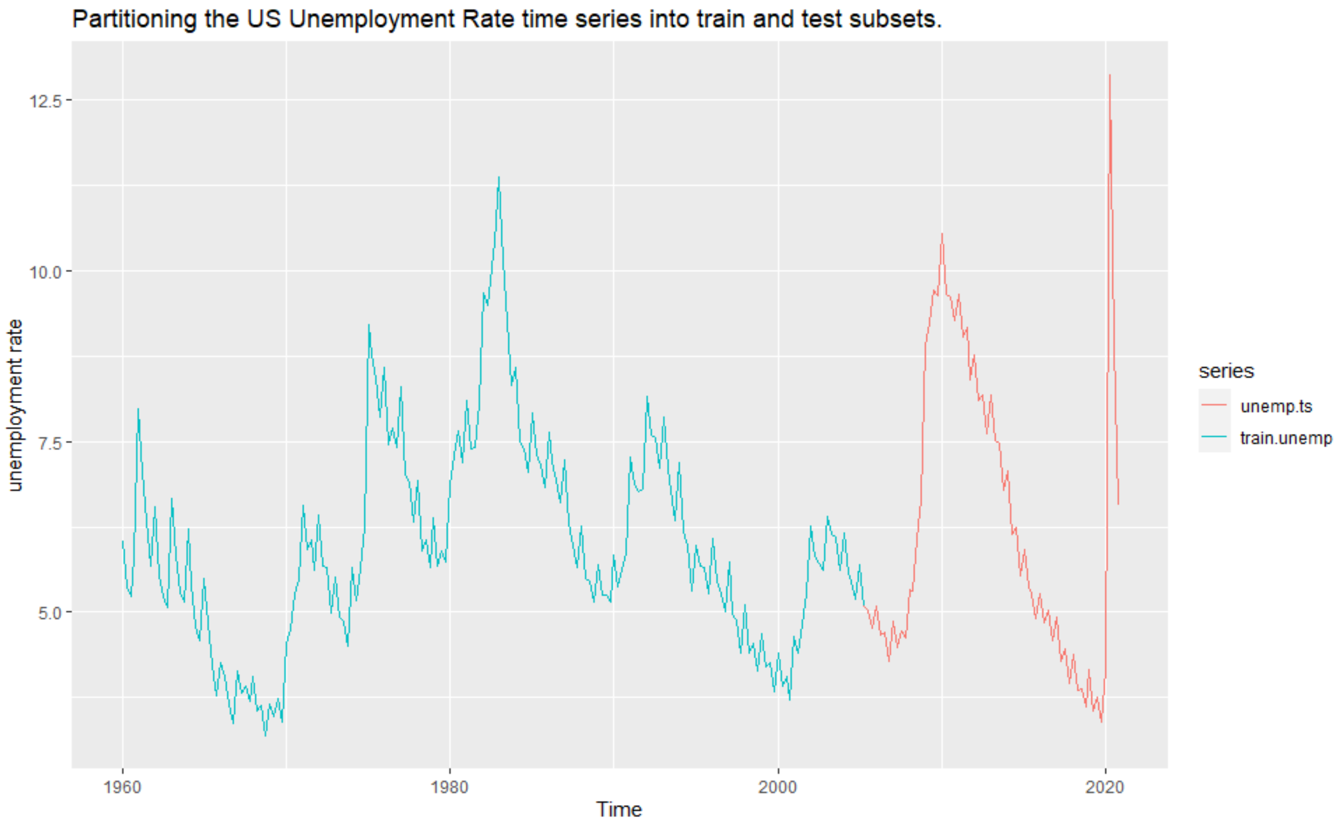
Similarity Check



Modeling

- Data Cleaning
- Data Partitioning
- Non-linearity Test
- Models
- Rolling Window Cross Validation
- Retrospective Analysis
- Diebold-Mariano Test
- Model Evaluation

Data Partitioning & Non-linearity Tests



75:25 Data Partition – Train data (183) & Test data (61 points)

- ❑ Non-linearity tests:
  - ✓ p-values – mostly more than 0.05 at 5% significance level
  - ✓ Fail to reject the null hypotheses
- ❑ To stationarize the data we have to go through:
  - Ordinary difference = 1
  - Seasonal difference = 1

Models, Accuracy Metrics & Residual Check

Fitted Model	Subset	AIC	RMSE	MAE	MAPE	MASE	Residual check
Naive	Training Testing		0.6694 0.4326	0.5399 0.3690	8.9968 8.0665	0.7187 0.4912	2.2 e-16
Seasonal Naive	Training Testing		1.0039 0.6513	0.7512 0.6262	11.9427 13.3562	1.0000 0.8337	2.2 e-16
STL Random Walk	Training Testing		0.3008 0.2454	0.2112 0.2145	3.4563 4.5867	0.2812 0.2856	0.0007153
Best ETS (A,Ad,A)	Training Testing	-352.33	0.3957 0.3082	0.2745 0.2627	4.4321 5.6238	0.3655 0.3497	7.183 e-11
ARIMA Brute Force ARIMA(2,0,1)(1,1,1)[4]	Training Testing	-801.78	0.3322 0.3509	0.2289 0.2942	3.7798 6.3380	0.3047 0.3916	0.09763
ARIMA Stepwise ARIMA(3,0,2)(1,1,1)[4]	Training Testing	-800.05	0.3228 0.3339	0.2269 0.2741	3.7541 5.9170	0.3022 0.3649	0.1894

Forecast Intervals

89.898

```
> forecast(naive.un, h = 12)
         Point Forecast        Lo 80    Hi 80      Lo 95    Hi 95
Jan 2018        152.311 121.53480 179.5653  103.07621 192.9749
Feb 2018        152.311 107.26041 190.1090   77.73341 208.2873
Mar 2018        152.311  95.43744 197.9552   53.96962 219.5702
Apr 2018        152.311  84.69952 204.4204   26.54856 228.8049
May 2018        152.311  74.47053 210.0111  -25.25114 236.7500
Jun 2018        152.311  64.39035 214.9859  -48.34970 243.7908
Jul 2018        152.311  54.14613 219.4973  -64.51731 250.1539
Aug 2018        152.311  43.35805 223.6445  -77.52688 255.9861
Sep 2018        152.311  31.36650 227.4959  -88.59427 261.3884
Oct 2018        152.311  16.24797 231.1011  -98.31087 266.4340
Nov 2018        152.311 -14.49270 234.4975 -107.01986 271.1777
Dec 2018        152.311 -29.03857 237.7140 -114.94188 275.6617
```

27.537

```
> forecast(stl.rw.un, h = 12)
         Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
Jan 2018       160.0078 150.85444 168.8418 145.86454 173.4018
Feb 2018       136.0183 123.36993 147.9635 116.33273 154.0457
Mar 2018       130.7847 115.65135 144.8838 107.11379 152.0055
Apr 2018       115.0818  96.40984 132.0057  85.54279 140.4279
May 2018       133.8848 115.42658 150.8569 104.86285 159.3674
Jun 2018       163.5742 146.17557 179.8744 136.41690 188.1346
Jul 2018       192.8223 176.15666 208.6237 166.92016 216.6888
Aug 2018       189.2140 171.26589 206.1453 161.26897 214.7604
Sep 2018       156.6961 135.06390 176.5844 122.68217 186.5566
Oct 2018       137.1878 112.11963 159.6385  97.30910 170.7443
Nov 2018       131.9597 104.90639 155.8762  88.64204 167.6342
Dec 2018       151.7597 126.31091 174.7524 111.44180 186.1763
```

30.215

```
> forecast(arima.un)
         Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
Jan 2018       157.8840 147.82320 167.5553 142.31891 172.5345
Feb 2018       131.2087 117.76047 143.8372 110.23534 150.2454
Mar 2018       124.6071 109.93370 138.2610 101.64495 145.1530
Apr 2018       108.2729  91.76075 123.3252  82.21627 130.8382
May 2018       126.3280 111.42760 140.1915 103.00958 147.1890
Jun 2018       156.5205 143.62932 168.7732 136.50163 175.0365
Jul 2018       185.3032 173.78466 196.3864 167.48826 202.0961
Aug 2018       182.9159 171.30322 194.0806 164.95071 199.8293
Sep 2018       146.9437 133.48381 159.6659 126.00130 166.1470
Oct 2018       127.8463 113.02928 141.6490 104.66923 148.6207
Nov 2018       120.7393 105.31560 135.0064  96.54711 142.1841
Dec 2018       142.6075 128.86800 155.5570 121.20842 162.1426
```

56.458

```
> forecast(snaive.un, h = 12)
         Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
Jan 2018        154.255 135.19045 171.9335 124.38111 180.8396
Feb 2018        120.576  97.79512 140.9045  84.27208 150.9429
Mar 2018        127.915 106.10996 147.5837  93.34410 157.3480
Apr 2018        116.751  93.41118 137.4457  79.43672 147.6339
May 2018        133.270 112.10829 152.4895  99.82240 162.0642
Jun 2018        155.979 137.06257 173.5443 126.35309 182.4003
Jul 2018        188.467 171.86560 204.1919 162.65576 212.2129
Aug 2018        177.863 160.59184 194.1338 150.95900 202.4062
Sep 2018        148.666 129.09884 166.7244 117.94662 175.7978
Oct 2018        135.383 114.46116 154.4322 102.35069 163.9343
Nov 2018        131.357 109.97150 150.7340  97.52026 160.3755
Dec 2018        152.311 133.07562 170.1194 122.15044 179.0829
```

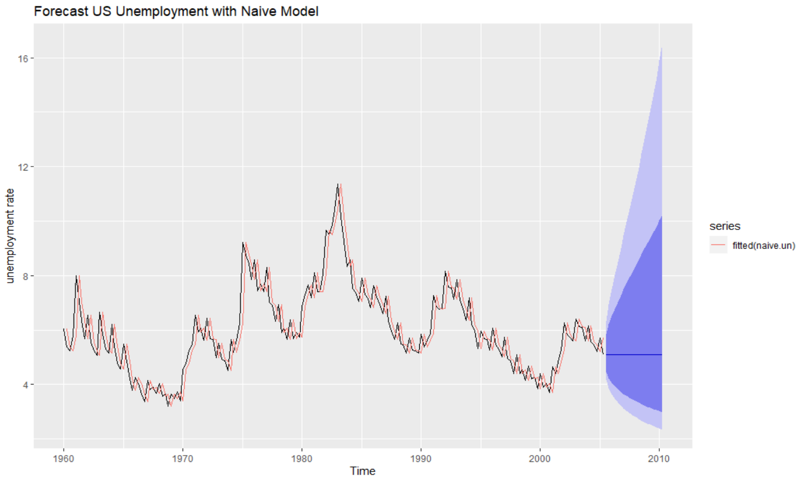
30.317

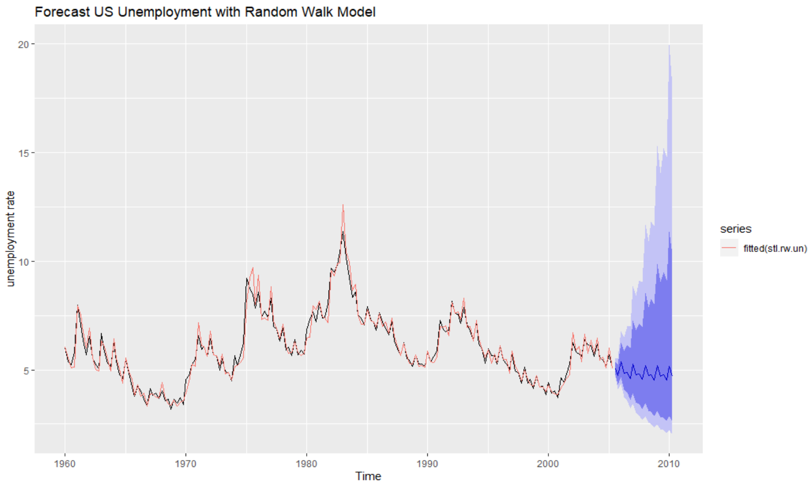
```
> forecast(ets.un)
         Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
Jan 2018       159.7160 149.62251 169.4219 144.10193 174.4199
Feb 2018       134.2417 120.64704 147.0171 113.04562 153.5026
Mar 2018       128.9375 112.93942 143.7682 103.86589 151.2385
Apr 2018       113.6778  94.16195 131.2691  82.72616 139.9991
May 2018       131.8847 112.70938 149.4380 101.67950 158.2189
Jun 2018       161.8642 143.95423 178.6005 133.88217 187.0691
Jul 2018       190.0906 172.96489 206.2926 163.45293 214.5508
Aug 2018       189.8894 171.72459 207.0169 161.60210 215.7293
Sep 2018       154.4085 132.28112 174.6891 119.57123 184.8409
Oct 2018       135.6988 110.16628 158.4932  95.01886 169.7517
Nov 2018       129.6122 101.95091 153.9473  85.20603 165.8843
Dec 2018       152.3555 126.80429 175.4405 111.87516 186.9102
```

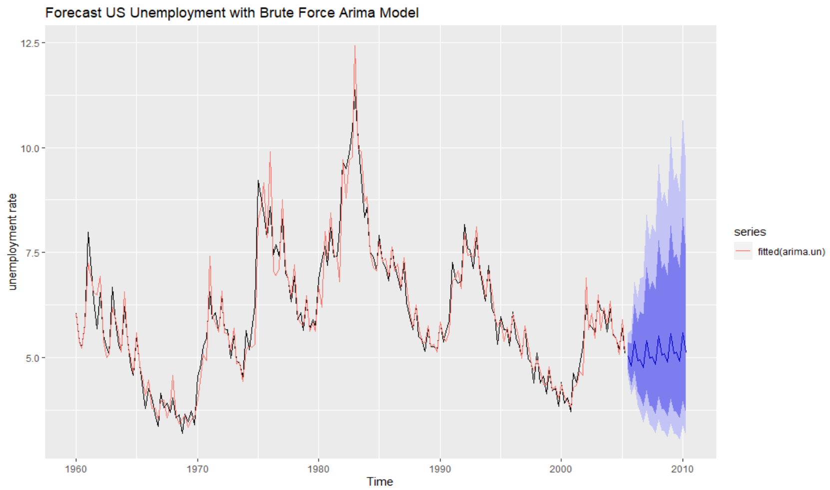
30.190

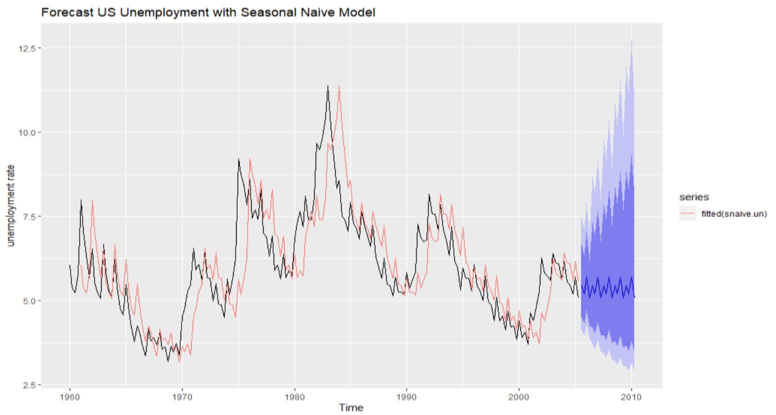
```
> forecast(arima.un.sw)
         Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
Jan 2018       156.0875 146.03343 165.7483 140.53069 170.7208
Feb 2018       130.2924 117.03933 142.7433 109.62695 149.0631
Mar 2018       127.6219 113.55998 140.7661 105.65419 147.4181
Apr 2018       109.4607  93.55630 124.0206  84.40850 131.3043
May 2018       128.2191 113.92063 141.5738 105.87521 148.3292
Jun 2018       155.1413 142.57119 167.0986 135.62637 173.2139
Jul 2018       183.9959 172.77837 194.7972 166.65059 200.3643
Aug 2018       179.8818 168.49833 190.8276 162.27189 196.4639
Sep 2018       145.5209 132.39070 157.9418 125.09735 164.2726
Oct 2018       127.5456 113.16777 140.9646 105.07161 147.7497
Nov 2018       122.8530 108.09401 136.5672  99.74430 143.4841
Dec 2018       141.6583 128.28418 154.2782 120.83686 160.7006
```

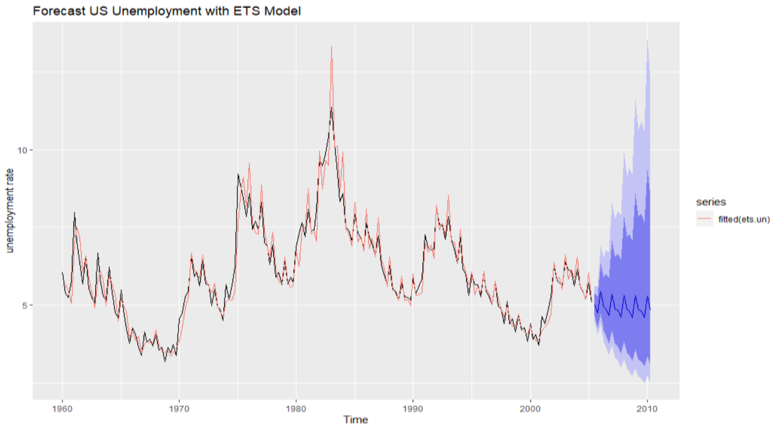
Graphical representation of fitted models

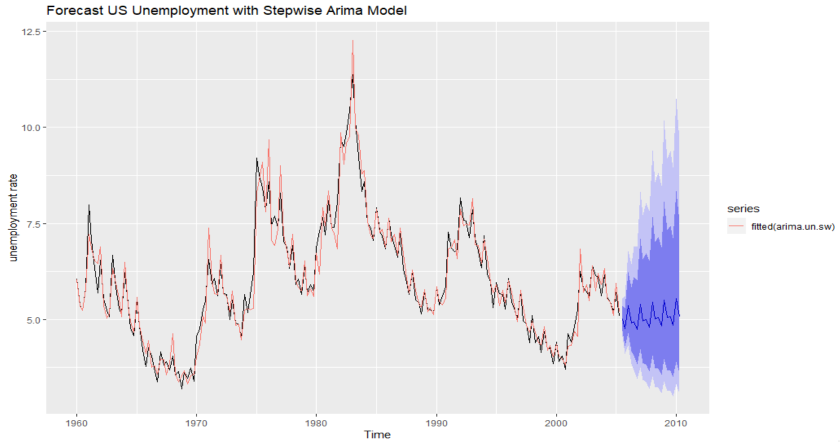




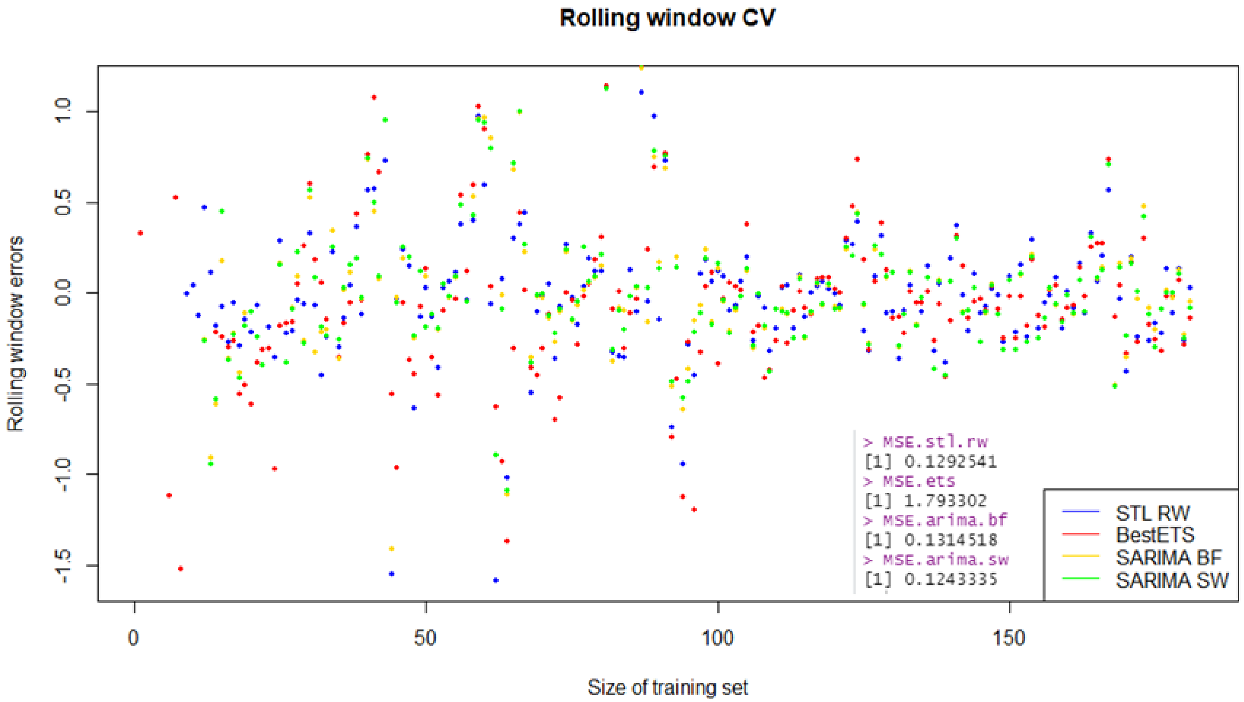


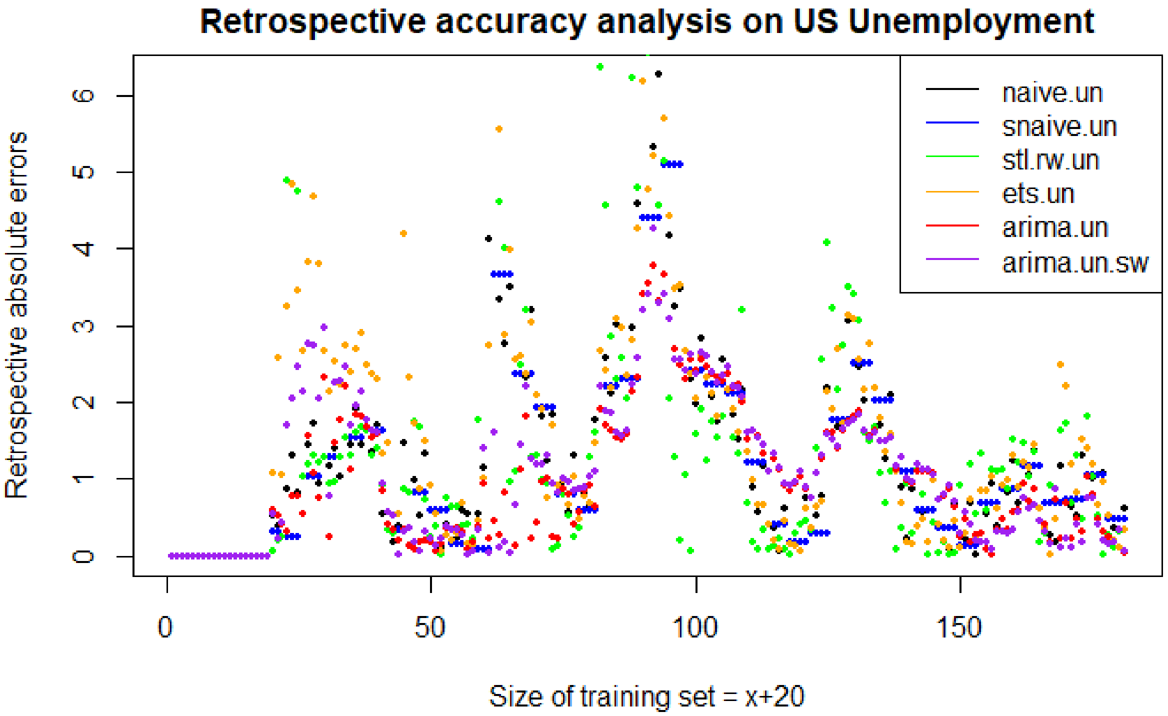






Rolling Window CV & Retrospective Analysis



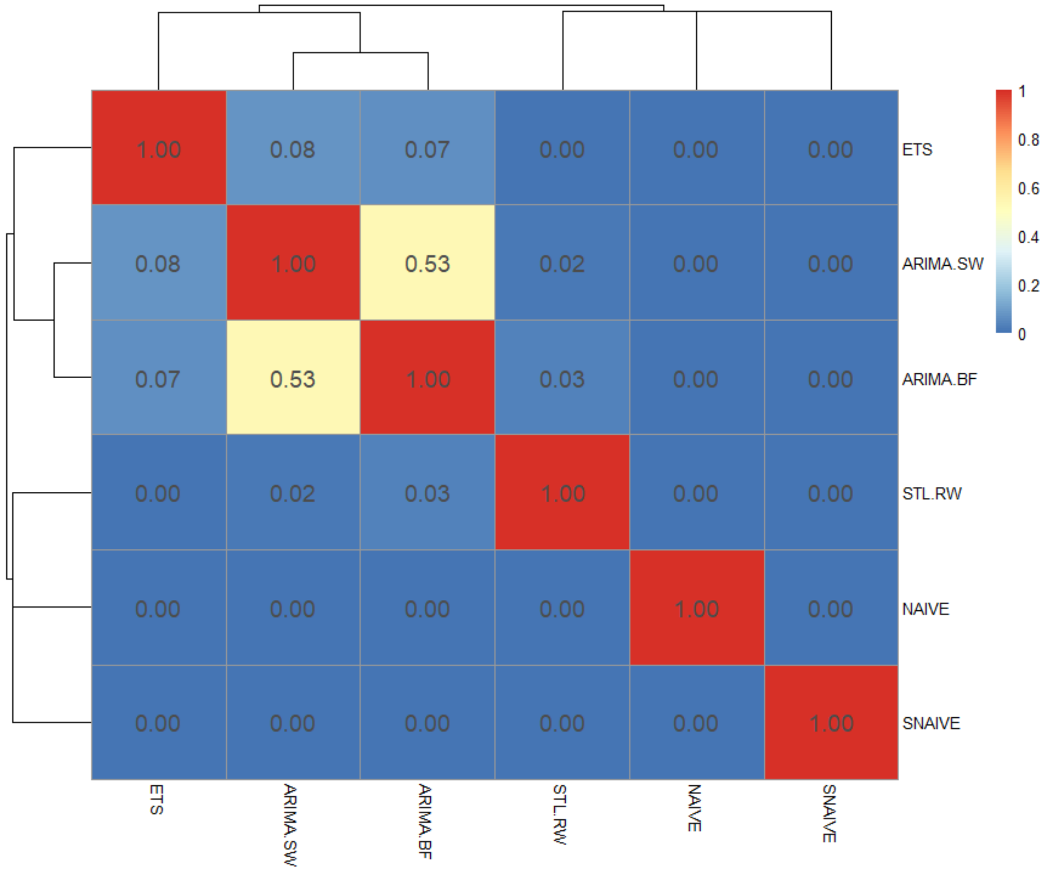


Winner: ARIMA Stepwise
ARIMA(3,0,2)(1,1,1)[4])

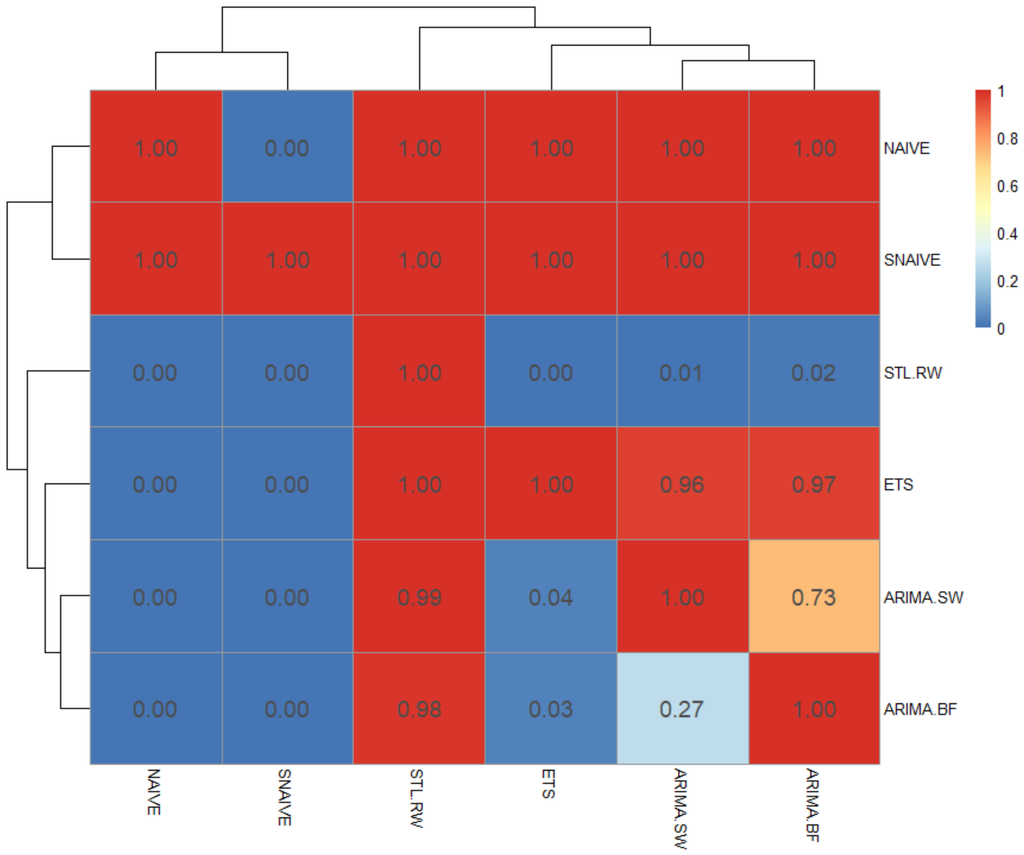
Winner: ARIMA Stepwise
ARIMA(3,0,2)(1,1,1)[4])

Diebold-Mariano (DM) test for in-sample accuracies

P-values from two tailed type test



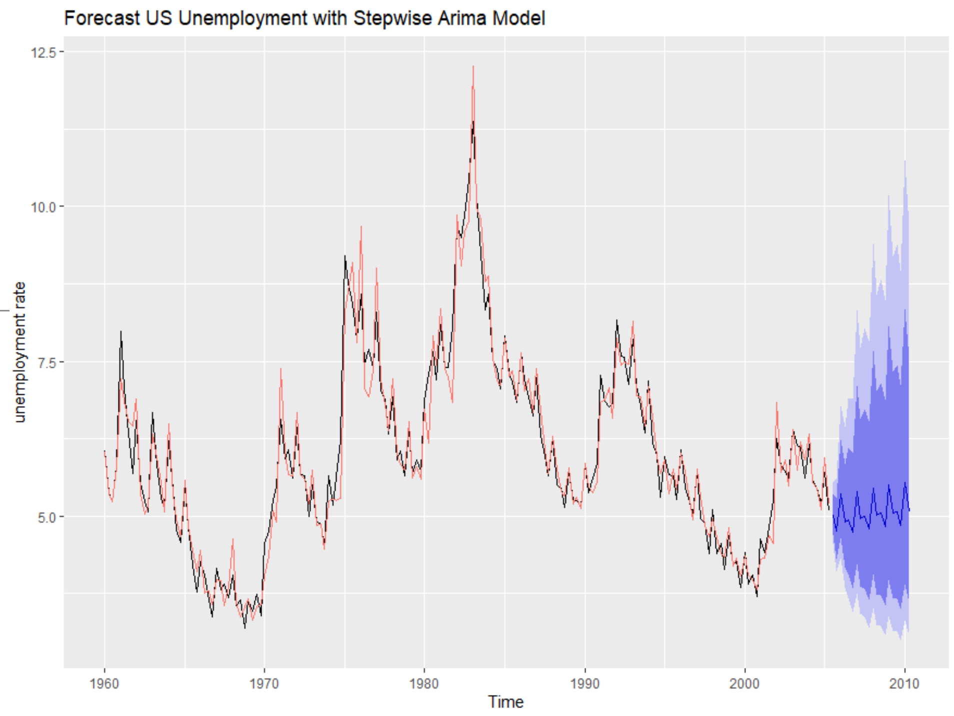
P-values from less-than type test



Model Comparison

	Properties of Good Models	ARIMA Brute Force ARIMA(2,0,1)(1,1,1)[4]	ARIMA Stepwise ARIMA(3,0,2)(1,1,1)[4]	ETS $A\,A_d\,A$	STL RW
1	Small MAPE / MASE	* (as per DM test)	*	**	***
2	Compact Forecast Interval	*	**		***
3	Small AIC	*		-	-
4	Significant Parameters	***	***		
5	Good Residual Properties	***	***		
6	Good Retro Analysis	**	***	*	
7	Rolling Window Analysis	*	***		**
8	No Model Violation	***	***	***	***
9	Logical	***	***	***	***
10	Parsimonious	*		**	***

Selected model



ARIMA(3,0,2)(1,1,1)[4] with no drift
p = 3, d = 0, q = 2 for the ordinary part
P = 1, D = 1, Q = 1 for the seasonal part
with seasonal period 4

$$(1-\Phi_1 B^4)(1-B^4)(1-\phi_1 B-\ \phi_2 B^2-\phi_3 B^3)X_t=(1+\theta_1 B+\theta_2 B^2)(1+\Theta_1 B^4)\ \epsilon_t$$

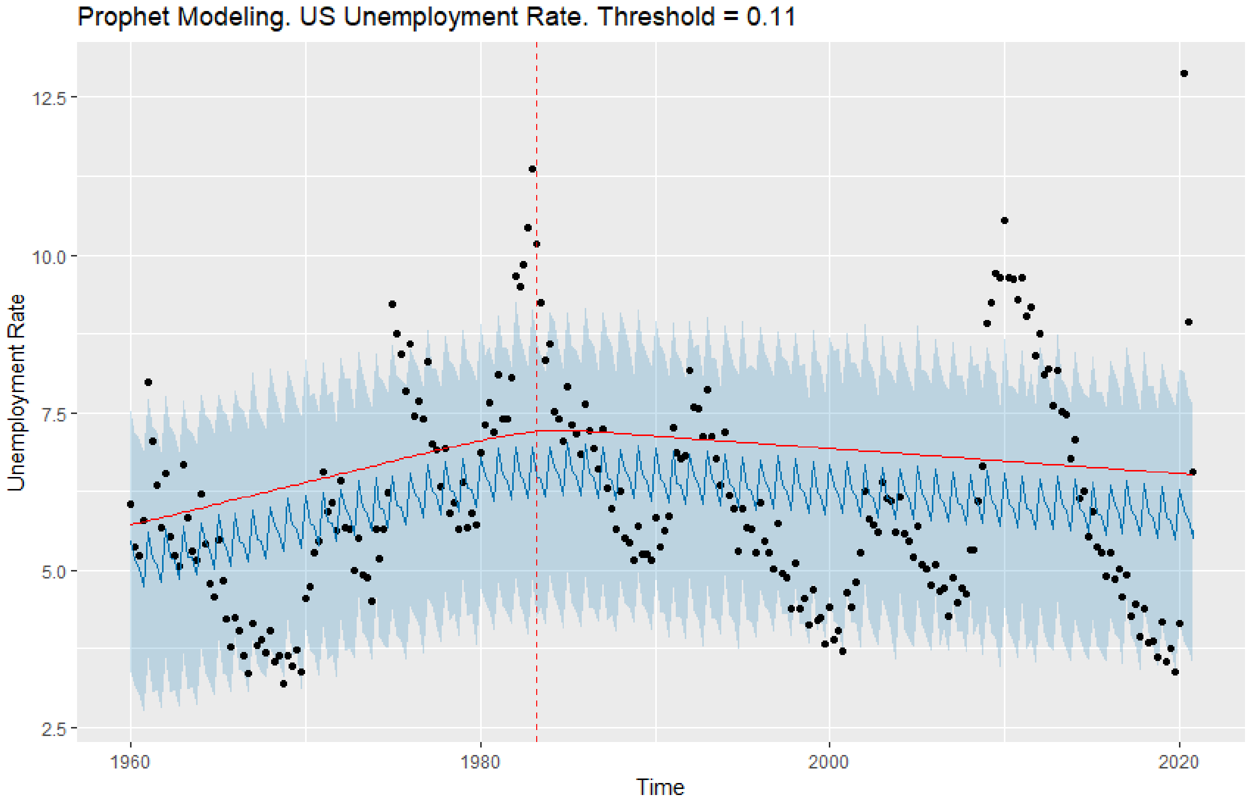
where $\phi_1 = 1.7665,\ \ \phi_2 = \ -1.1822,\ \ \phi_3 = \ 0.3325,\ \ \theta_1 = -0.4806;\ \ \theta_2 = 0.4055;$

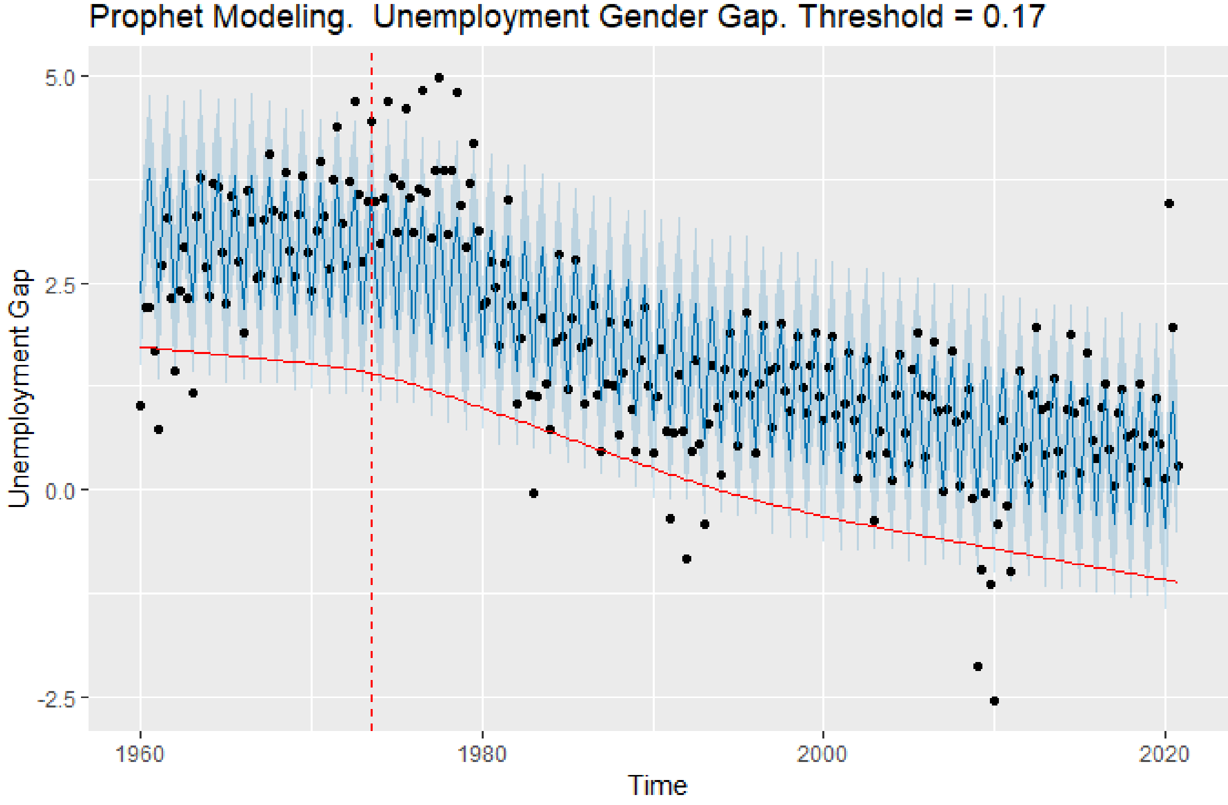
$\Phi_1 = 0.3661,\ \ \ \Theta_1 = \ -0.8188$

Additional Option: Ensemble Model with Weighted Average:
40% ARIMA.SW + **40%** ARIMA.BF + **10%** STL.RW + **10%** ETS

Detecting Change Points

Unemployment Gap = Female Unempl Rate – Male Unempl Rate





Conclusion

- Linear models proved to be the best for forecasting unemployment rate in the US.
  - Best linear model: ARIMA(3,0,2)(1,1,1)[4] - Stepwise
- Clustering for state unemployment into 4 large clusters which can be used to transfer features between states within the same cluster
- Prophet model showed that the change point in gender unemployment gap was in year 1975

Future Work

- Unemployment rate peak in 2020 Q2: the next change point in the future, relevant for further studies upon similar events in the future
- Study for better forecast accuracy: ensemble weighted averaging the four best linear models (Arima BF – 40%, Arima SW – 40%, ETS – 10%, STL – 10%).

Q&A

References

- Albanesi, S., & Sahin, A. (2017). The Gender Unemployment Gap .
- Amadeo, K. (2020, September 27). The Balance. Retrieved from www.thebalance.com: https://www.thebalance.com/labor-force-definition-how-it-affects-the-economy-4045035
- Fred Economic Data. (2021, April 2). Fred Economic Data. Retrieved from www.fred.stlouisfed.org: https://fred.stlouisfed.org/series/UNRATE
- Simpson, S. (2020, September 24). Investopedia. Retrieved from investopedia.com: https://www.investopedia.com/financial-edge/0811/the-cost-of-unemployment-to-the-economy.aspx
- U.S. Bureau of Labour Statistics . (2020, May 13). U.S. Bureau of Labour Statistics. Retrieved from www.bls.gov: https://www.bls.gov/opub/ted/2020/unemployment-rate-rises-to-record-high-14-point-7-percent-in-april-2020.htm?view_full
- https://www.ft.com/content/bc513b58-8ef2-11ea-af59-5283fc4c0cb0