ECON 5P04: Unemployment Rate Forecast

OKONKWO IFEANYI (5770219) and ZACH MESIC(5820105)

March 23, 2020

Contents

1 Introduction

The civilian unemployment rate represents the number of unemployed as a percentage of the labor force. It contains monthly, not seasonally adjusted U.S data from January 2010 to March 1, 2020. Data was retrieved from FRED and originates from the Current Population Survey (Household Survey).

2 Trend and Seasonality

A time series and seasonal plot were generated for the years 2010 to 2020 and reported in figures 1 and 2. A steady decreasing trend can be observed in civilian unemployment rates from the time series plot, this has been the case since the end of the 2007-2009 recession. The decrease in unemployment rate can be attributed to the decreasing trend in the number of civilian in the Labor force of the United states. The seasonality plot displays monthly(seasonal) effects within the data. It is clear that Unemployment rate is on a steady decline between January and April and starts to increase again between May and July. This slight increase can be attributed to teens and young adults who have recently graduated and are in search for employment opportunities to match their skillset. In the Fall and Winter periods the unemployment rate stabilizes as recent graduates are expected to have been matched with employers and teenagers return to school concentrating on full time studies.
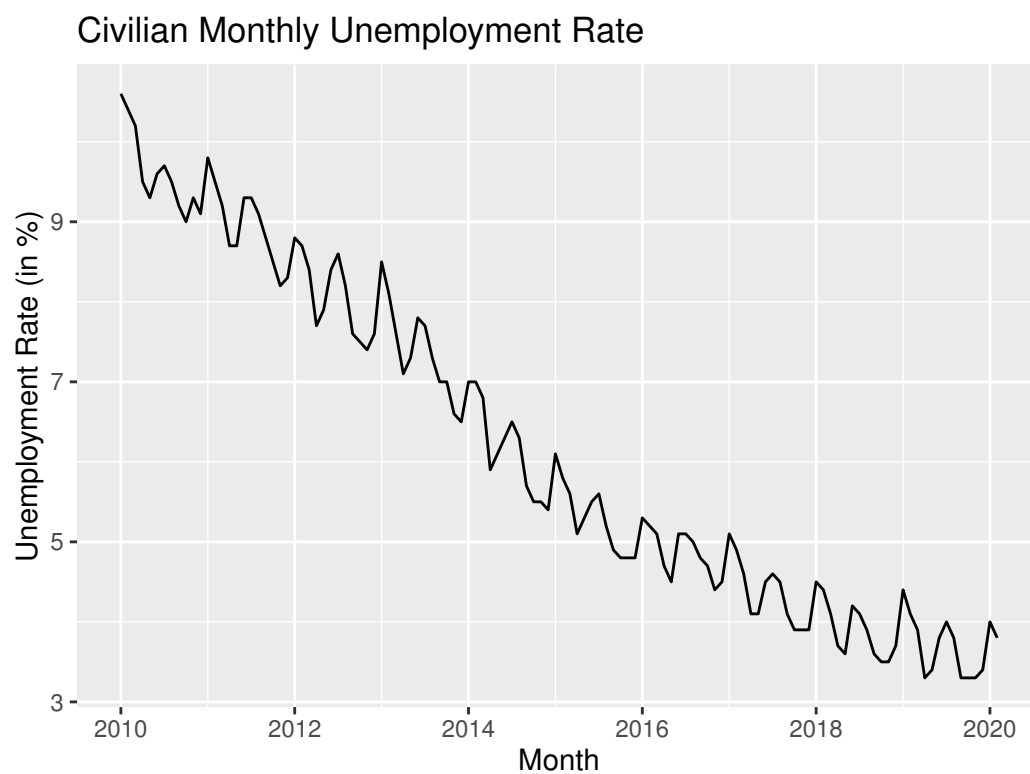
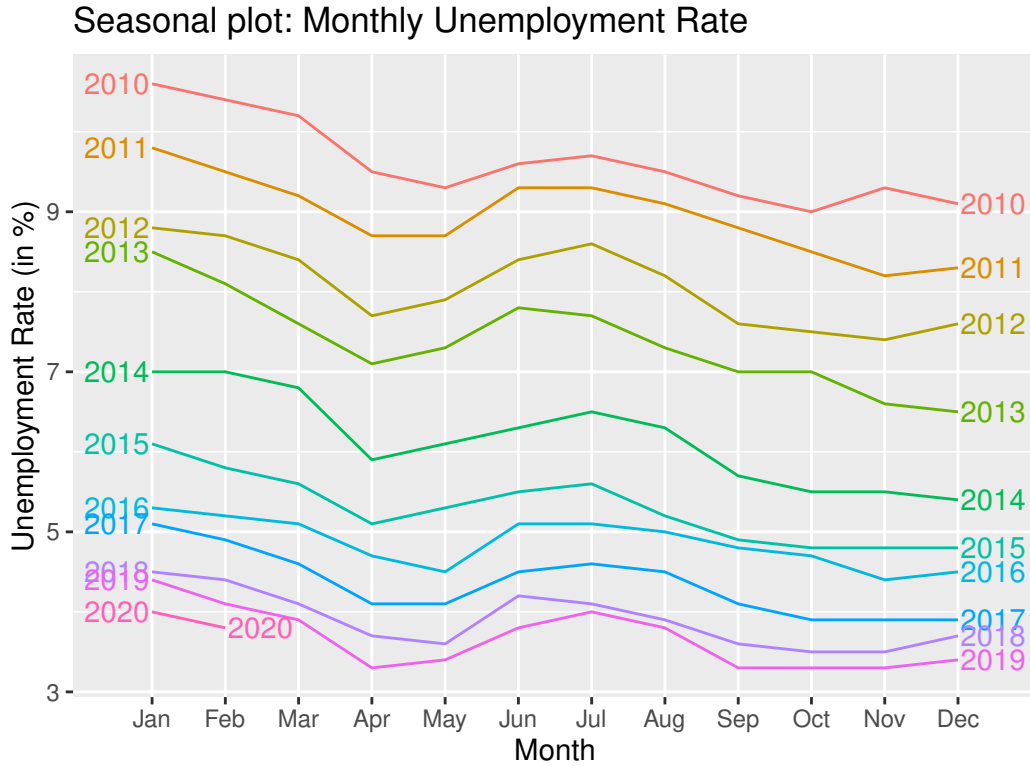Figure 1: Unemployment rate time series plot

Figure 2: Unemployment rate seasonality plot

3 Autocorrelation

As observed in previous plots the data consist of seasonal components as well as a declining trend. Therefore, as a result of the trend component, large and positive values that slowly decay as lags increase can be observed from the correlogram and table 1. This means that individual periods have positive correlation with one another and decrease as period lags increase (Observations nearby in time are also nearby in size).

Table 1: Correlogram and Autocorrelation values.

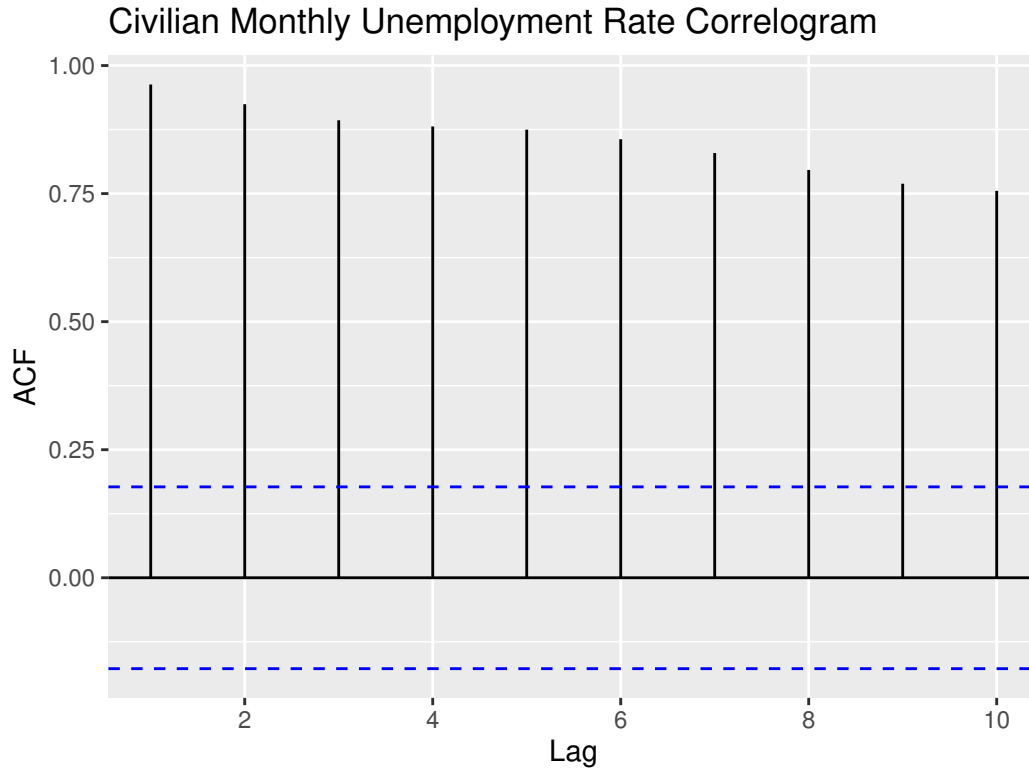| Lag	1	2	3	4	5	6	7	8	9	10
Correlation	0.96	0.93	0.89	0.88	0.88	0.87	0.83	0.80	0.77	0.76

Figure 3: Correlogram

4 Forecasting

A visualization of Accuracy for forecast generated using the Average, Naive and Seasonal Naive forecasting methods is provided in figure 4. To determine the accuracy of forecasting methods , the data is split into a training set (Jan 2010-Dec 2017) and a test set (Jan 2018- Feb 2020), Afterwards the afore-

mentioned forecasting methods are used over the training set and measured to determine which method provides the best fitting result in comparison to the test set.

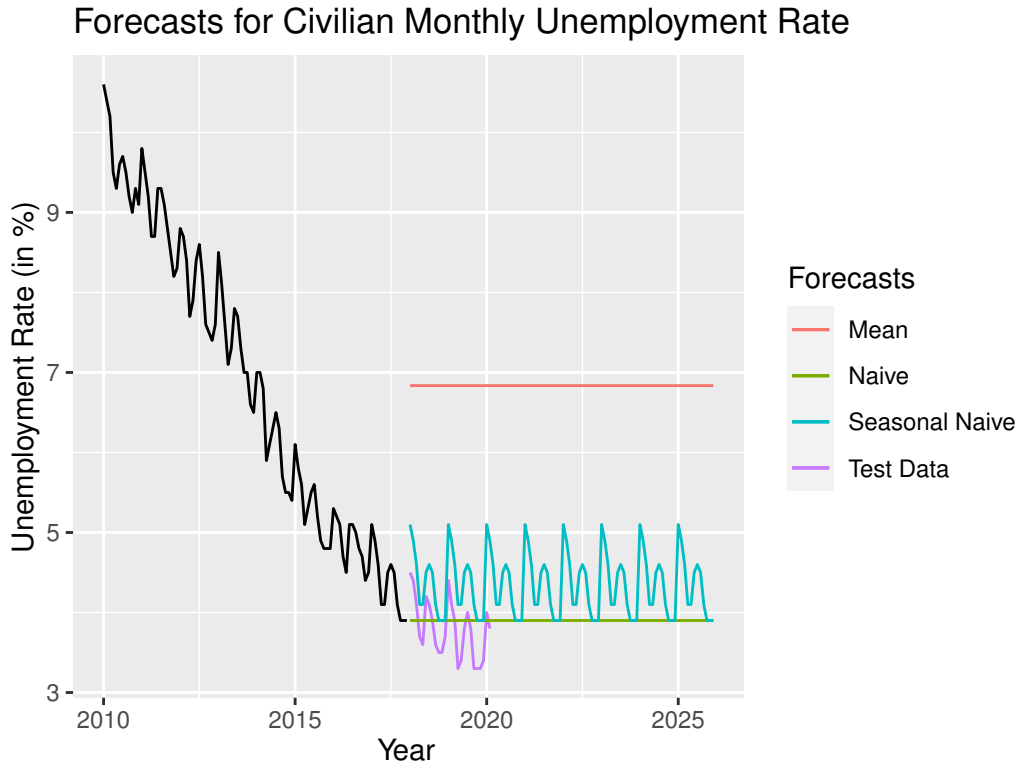## Forecasts for Civilian Monthly Unemployment Rate

Figure 4: Forecasting Methods

Accuracy is also determined numerically through error measurement and a graphical representation of each method alongside the test set. According to the tabulated error measurement values in table 2, the naive method is determined to be the best fit based on the fact that it minimizes the error in forecasting better than the average and seasonal naive method, Although based on graphical observation, the Seasonal Naive method best fits the Test set. The Naive method provides a constant forecast across all months and does not take into consideration the seasonality component of the data similar to the Average method. Therefore, we conclude that the seasonal naive

method is the best forecasting method as does an excellent job capturing the seasonality component of the data compared to the other two methods.

Table 2: Measures of accuracy.

		RMSE	MAE	MAPE	MASE
Average	Training set	1.892	1.685	27.145	2.240
	Test set	3.064	3.043	81.842	4.045
Naive	Training set	0.349	0.279	4.242	0.371
	Test set	0.374	0.315	8.585	0.419
Seasonal Naive	Training set	0.816	0.752	12.172	1
	Test set	0.642	0.608	16.164	0.808

Table 3: Mean forecasts.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84
2019	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84	6.84
2020	6.84	6.84	-	-	-	-	-	-	-	-	-	-

Table 4: Naive forecasts.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9
2019	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9
2020	3.9	3.9	-	-	-	-	-	-	-	-	-	-

Table 5: Seasonal naive forecasts.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	5.1	4.9	4.6	4.1	4.1	4.5	4.6	4.5	4.1	3.9	3.9	3.9
2019	5.1	4.9	4.6	4.1	4.1	4.5	4.6	4.5	4.1	3.9	3.9	3.9
2020	5.1	4.9	-	-	-	-	-	-	-	-	-	-

5 Regression based forecasting analysis

The following tables reports regression values for trend, seasonal dummies and Trend with seasonal dummies forecasting models. The best forecasting model is determined by assessing the R squared and Adjusted R squared to determine the goodness of fit of each model. The Trend with seasonal dummies model is concluded to be the best and most accurate forecast model. This decision is supported by our Time series plot as we already analyzed the Unemployment rate variable to exhibit a declining trend. Furthermore, the Check residual function in R programming is used to validate our decision in selecting the Trend with Seasonal dummies forecast as the best model compared to Trend and seasonal Models. The graphics generated using the Check Residual function by R program are reported in figures 5,6 and 7. From the visual representation we can observe seasonal patterns in the residuals time plot obtained from the Trend only forecasting model and the seasonal forecasting model residuals exhibit a declining trend. The Trend with seasonal dummies model Residual time plot contains no observable Trend or seasonal Patterns, and it?s mean is closely centered around Zero, Thus validating our decision to select this model as the most accurate model between the three.

Table 6: Trend-only regression results.

Call: tslm(formula = unrtrain ~ trend)

Residuals:				
Min	1Q	Median	3Q	Max
-0.70106	-0.30502	0.00026	0.26722	0.89454

Coefficients:				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)	10.082982	0.077165	130.67	¡2e-16 ***
trend	-0.066960	0.001381	-48.47	¡2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.3751 on 94 degrees of freedom
Multiple R-squared: 0.9615, Adjusted R-squared: 0.9611
F-statistic: 2349 on 1 and 94 DF, p-value: ¡ 2.2e-16

Table 7: Seasonal dummies-only regression results.

Call:
tslm(formula =
unrtrain ∼ season)

Residuals:

	Min	1Q	Median	3Q	Max
	-2.587	-1.669	-0.075	1.581	3.038

Coefficients:

	Estimate	Std. Error	t value	Pr($¿$—t—)	
(Intercept)	7.6500	0.6951	11.006	¡2e-16	***
season2	-0.2000	0.9830	-0.203	0.839	
season3	-0.4625	0.9830	-0.470	0.639	
season4	-1.0500	0.9830	-1.068	0.289	
season5	-1.0000	0.9830	-1.017	0.312	
season6	-0.5875	0.9830	-0.598	0.552	
season7	-0.5125	0.9830	-0.521	0.603	
season8	-0.7625	0.9830	-0.776	0.440	
season9	-1.1375	0.9830	-1.157	0.250	
season10	-1.2875	0.9830	-1.310	0.194	
season11	-1.3875	0.9830	-1.411	0.162	
season12	-1.3875	0.9830	-1.411	0.162	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.966 on 84 degrees of freedom
Multiple R-squared: 0.0555, Adjusted R-squared: -0.06818
F-statistic: 0.4487 on 11 and 84 DF, p-value: 0.9288

Table 8: Trend with seasonal dummies regression results.

Call:
tslm(formula =
unrtrain $\sim$ trend + season)

Residuals:

	Min	1Q	Median	3Q	Max
	0.4944	-0.1761	0.0375	0.2221	0.4523

Coefficients:

	Estimate	Std. Error	t value	Pr($¿$—t—)	
(Intercept)	10.5003142	0.1024510	102.491	¡ 2e-16	***
trend	-0.0662864	0.0009804	-67.614	¡ 2e-16	***
season2	-0.1337136	0.1320577	-1.013	0.3142	
season3	-0.3299272	0.1320686	-2.498	0.0145	*
season4	-0.8511409	0.1320868	-6.444	7.22e-09	***
season5	-0.7348545	0.1321122	-5.562	3.17e-07	***
season6	-0.2560681	0.1321450	-1.938	0.0561	.
season7	-0.1147817	0.1321850	-0.868	0.3877	
season8	-0.2984954	0.1322322	-2.257	0.0266	*
season9	-0.6072090	0.1322867	-4.590	1.56e-05	***
season10	-0.6909226	0.1323485	-5.220	1.30e-06	***
season11	-0.7246362	0.1324174	-5.472	4.61e-07	***
season12	-0.6583499	0.1324936	-4.969	3.56e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2641 on 83 degrees of freedom
Multiple R-squared: 0.9832, Adjusted R-squared: 0.9807
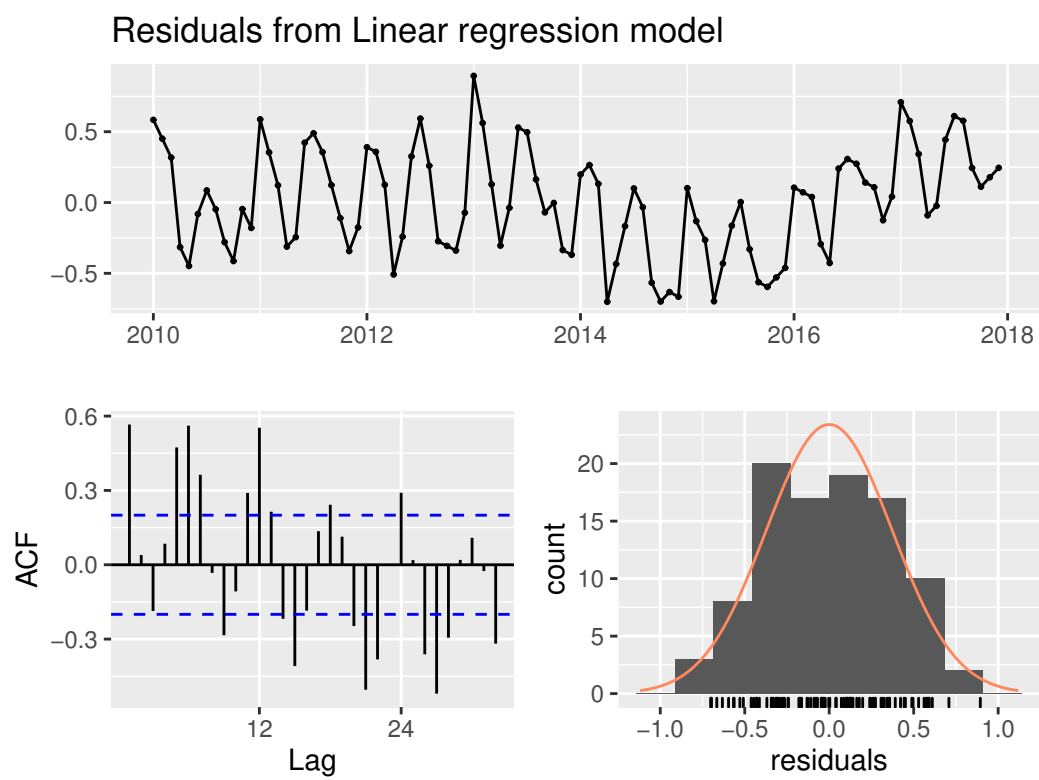F-statistic: 403.8 on 12 and 83 DF, p-value: ¡ 2.2e-16

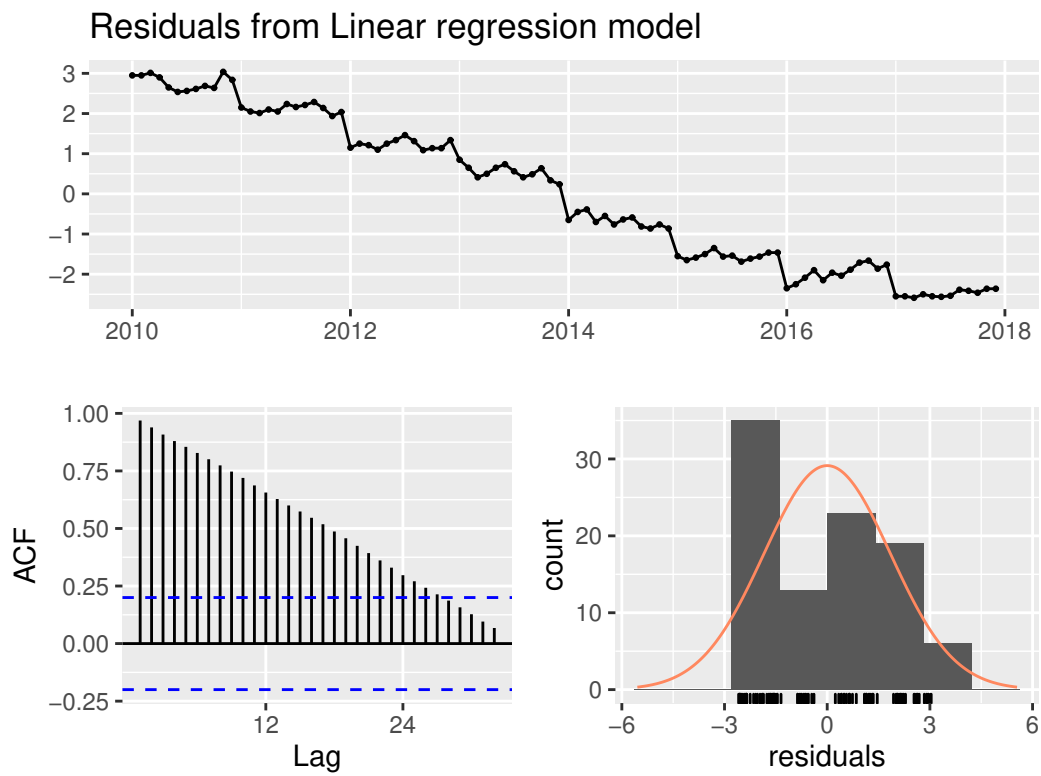Figure 5: Regression residuals from Trend Forecast
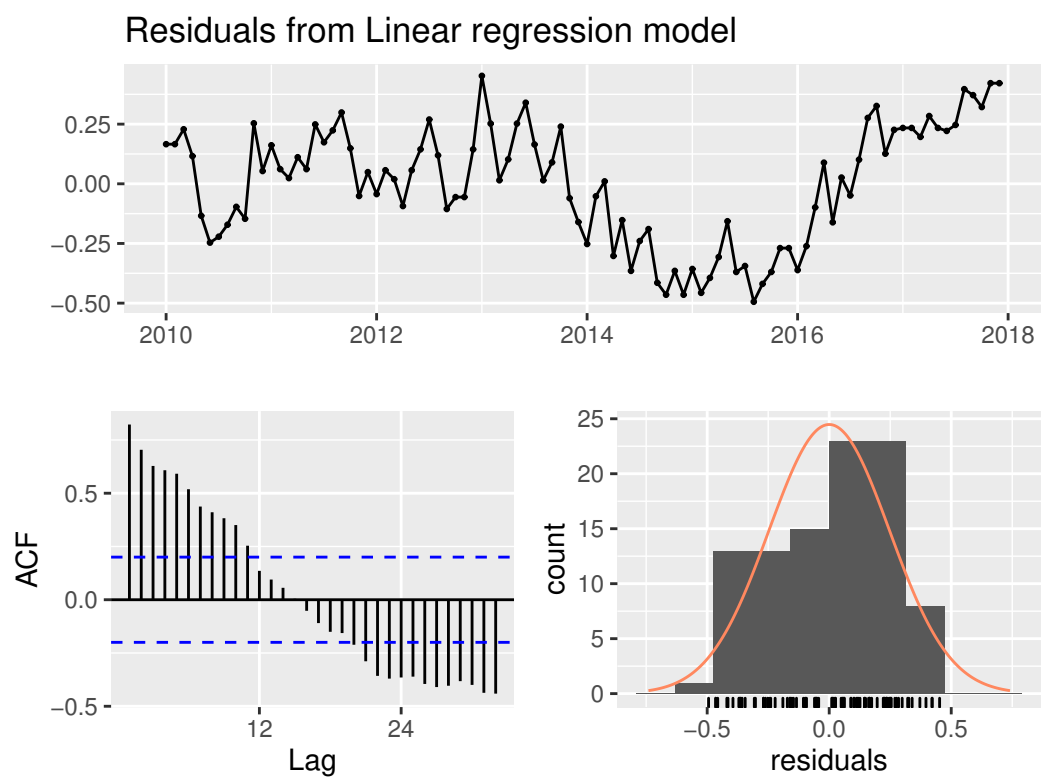
Figure 6: Regression residuals from Seasonal Forecast

Figure 7: Regression residuals from Trend with Seasonal and Forecast

6   Coefficient of Variation(CV)

The Coefficient of Variation (CV) is the ratio of the Standard deviation to the mean. For the purpose of this assignment, we use it to compare dispersion of our model values to the actual values. The decision rule is based on the lower CV, which signifies that our model is a close representation of the actual data. The values reported in table below represents the CV for each model. The CV of the Trend and Seasonal model has a value of 0.08 and is the lowest amongst the three model further supporting our decision in selecting the Trend with Seasonal dummies model as the best model.

Table 9: Coefficients of variation.

	Trend	**Season**	**Trend and Season**
CV	0.143	4.417	0.081

7   Forecasting Over The Test Set

The three regression model forecast over the test set is visually reported in figures 8,9, 10. Graphically the seasonal model does a good job accounting for seasonality patterns in the data although we can see from the data that this forecast is way off from the observed declining trend in the data. The trend model fixes this problem but misses the seasonal fluctuations in the data. Once again, the Trend and seasonal model comes out top amongst the three models. Measures of accuracy (RMSE and MAPE) are also reported in a table and based on this values we can clearly Justify our selection of the Trend with Seasonal dummies forecasting model.
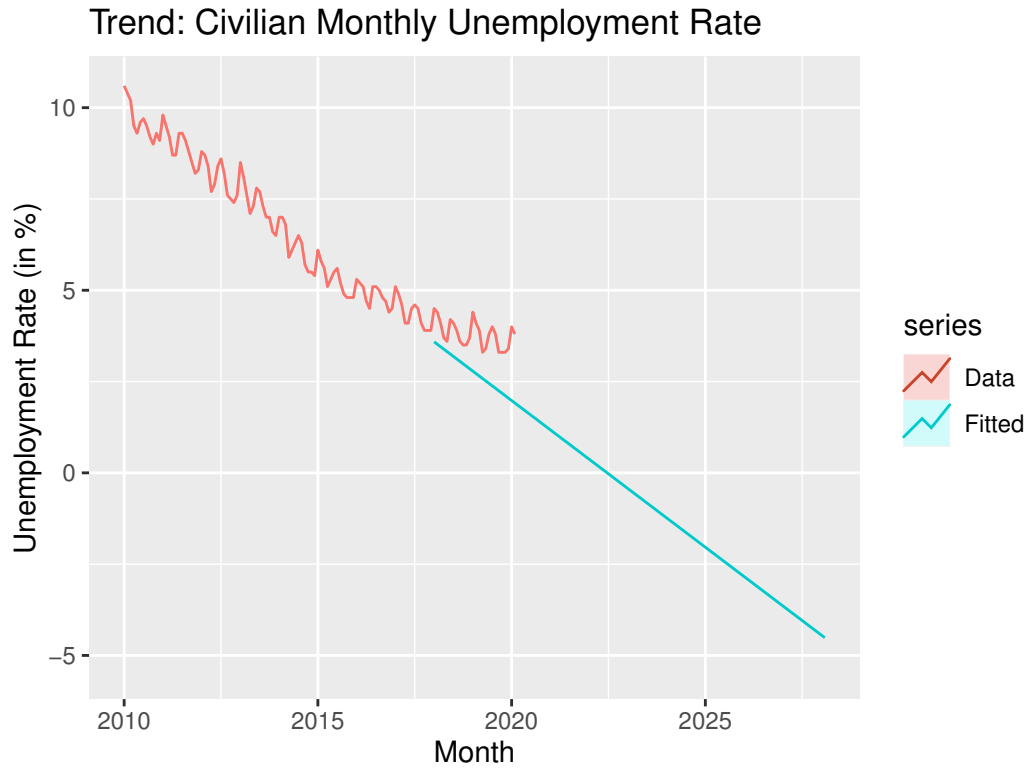
Figure 8: Trend Regression Forecast

Table 10: Measures of forecast accuracy.

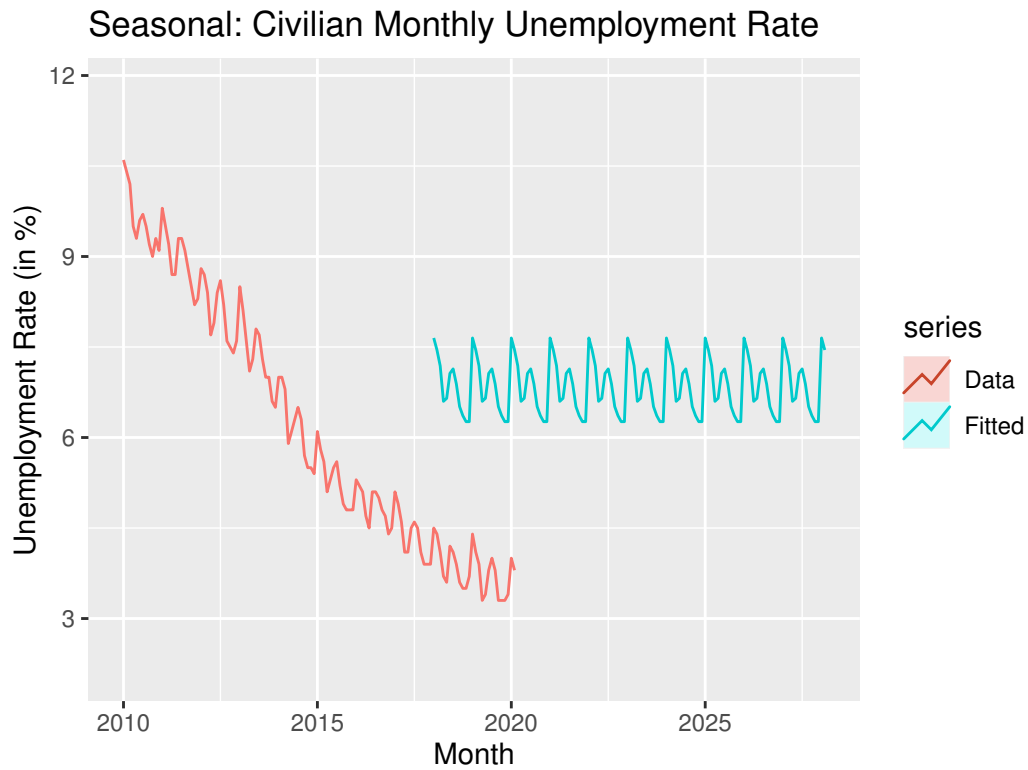|		RMSE	MAPE
Trend	Training set	0.371	4.910
	Test set	0.715	16.703
Seasonal	Training set	1.839	26.686
	Test set	2.992	75.925
Trend and seasonal	Training set	0.246	3.584
	Test set	0.597	15.113

Figure 9: Seasonal Regression Forecast

The trend with Seasonal forecast model proves to be the most accurate model when working with the Unemployment rate data. This is due to the strong declining trend observable from figure 1 as well as seasonal fluctuations noticeable in the data. The Trend with seasonal dummies forecast model does a good job accounting for each characteristic of the Unemployment rate data. The Forecast values generated using each Regression are reported in the tables below.

Table 11: Trend only forecasts.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	3.59	3.52	3.45	3.39	3.32	3.25	3.19	3.12	3.05	2.99	2.92	2.85
2019	2.78	2.72	2.65	2.58	2.52	2.45	2.38	2.32	2.25	2.18	2.11	2.05
2020	1.98	1.91	-	-	-	-	-	-	-	-	-	-

Table 12: Seasonal only forecasts.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	7.65	7.45	7.19	6.60	6.65	7.06	7.14	6.89	6.51	6.36	6.26	6.26
2019	7.65	7.45	7.19	6.60	6.65	7.06	7.14	6.89	6.51	6.36	6.26	6.26
2020	7.65	7.45	-	-	-	-	-	-	-	-	-	-

Table 13: Trend and seasonal forecasts.

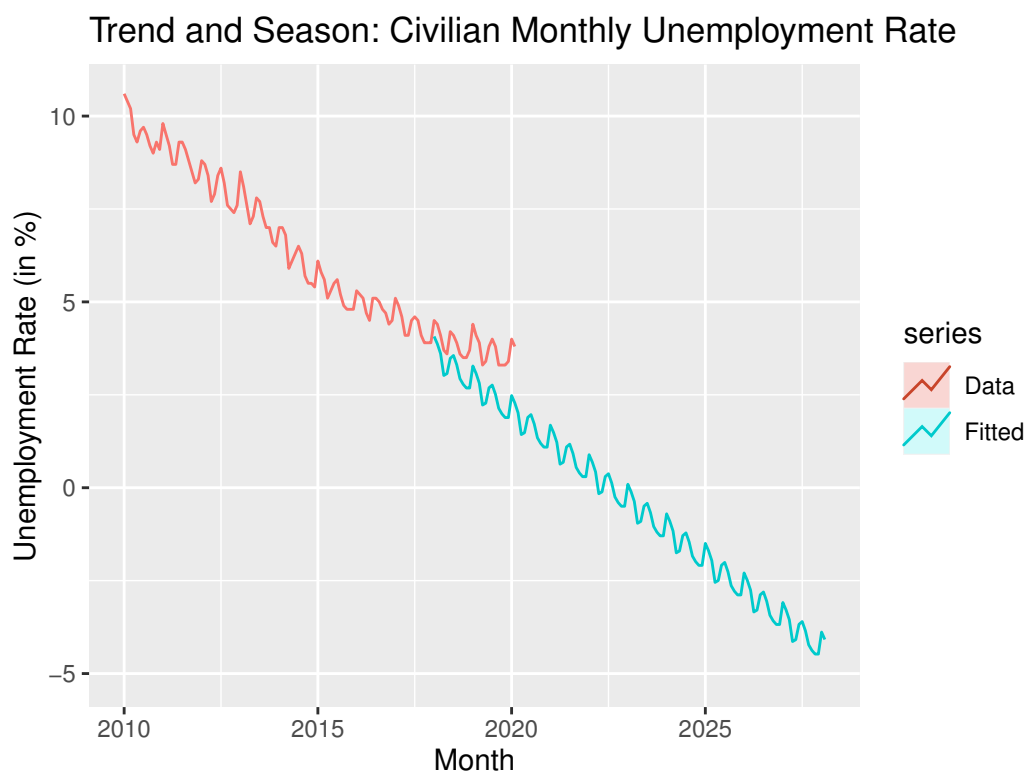|	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	4.07	3.87	3.61	3.02	3.07	3.48	3.56	3.31	2.93	2.78	2.68	2.68
2019	3.28	3.08	2.81	2.23	2.28	2.69	2.76	2.51	2.14	1.99	1.89	1.89
2020	2.48	2.28	-	-	-	-	-	-	-	-	-	-

Figure 10: Trend with Seasonal Regression Forecast

8 Future Forecast

The trend with seasonal dummies model is chosen as our favorite model and used to create a forecast over the next 12 month period. The forecasted values are reported in the table below and a visual representation is displayed in figure 11.

Table 14: Trend and seasonal model forecasts.

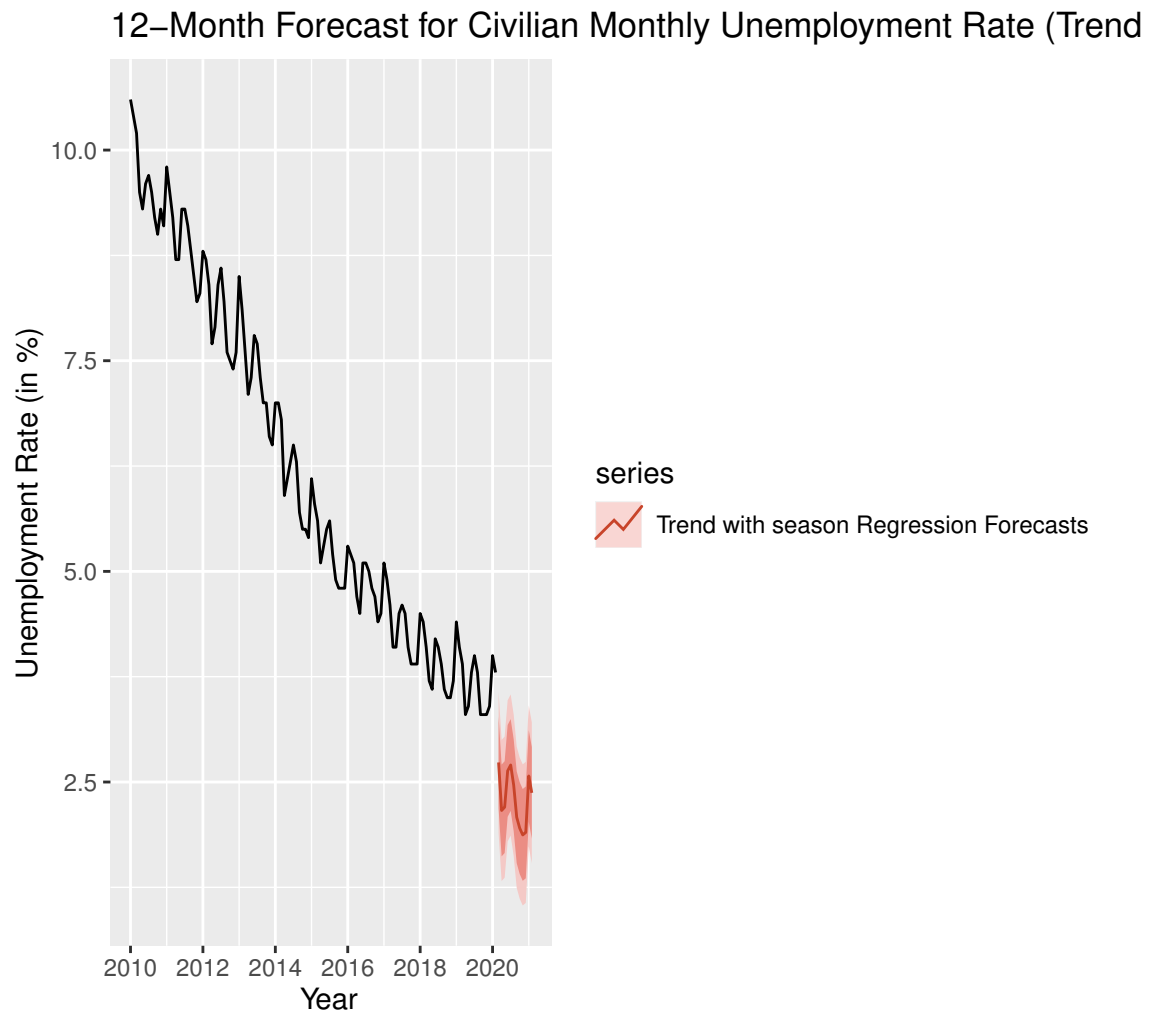|	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2020	-	-	2.73	2.16	2.20	2.63	2.70	2.46	2.08	1.95	1.87	1.90
2021	2.57	2.37	-	-	-	-	-	-	-	-	-	-

Figure 11: Trend with Seasonal Regression Forecast

The forecast predicts the declining trend to generally continue over the next year with seasonal spikes in the unemployment rate occurring in the middle of the year (summer time) due to an increase in the Labour Force Participation rate around this period. The Unemployment rate stabilizes at the end of the year and continues declining in 2021.

9 Appendices

9.1 R Code

```
setwd("X:/")

install.packages("ggplot2")
install.packages("forecast")
install.packages("fpp2")
library(ggplot2)
library(forecast)
library(fpp2)

df<-read.csv("Unemployment Rate Data.csv", header = TRUE)
unrate<-df$UNRATENSA

#create time series
ts_unrate<- ts(unrate,start=c(2010,1),end=c(2020,2), frequency=12)

#Q1
autoplot(ts_unrate) + ggtitle("Civilian Monthly Unemployment Rate") +
  xlab("Month") + ylab("Unemployment Rate (in %)")
ggseasonplot(ts_unrate, year.labels=TRUE, year.labels.left=TRUE) +
  ylab("Unemployment Rate (in %)") + ggtitle("Seasonal plot: Monthly Unemployment

#Q2
ggAcf(ts_unrate, lag=10) + ggtitle("Civilian Monthly Unemployment Rate Correlogram
  xlab("Lag") + ylab("ACF")
unrate_auto_corr<- ggAcf(ts_unrate, plot = FALSE)
unrate_auto_corr_10<-unrate_auto_corr[1:10,]

#Q3
unrtrain<-window(ts_unrate,start=c(2010,1), end=c(2017,12))
unrtest<-window(ts_unrate, start=c(2018,1), end=c(2020,2))

#a) Forecasts
meanf<-meanf(unrtrain,h=96)
meanf_list<-meanf$mean
```

```
naive<-naive(unrtrain, h=96)
naive_list<-naive$mean
snaive<-snaive(unrtrain, h=96)
snaive_list<-snaive$mean

#b) Plot forecasts
autoplot(ts_unrate) + ggtitle("Civilian Monthly Unemployment Rate") + xlab("Year")

autoplot(unrtrain) + autolayer(unrtest, series='Test Data') + autolayer(meanf$mean
  autolayer(naive$mean, series="Naive") + autolayer(snaive$mean, series="Seasonal
  ggtitle("Forecasts for Civilian Monthly Unemployment Rate") +
  xlab("Year") + ylab("Unemployment Rate (in %)") + guides(colour=guide_legend(tit

#c) Accuracy
unr_acc<-window(ts_unrate, start=2018)
accuracy(meanf, unr_acc)
accuracy(naive, unr_acc)
accuracy(snaive, unr_acc)

#Q4
#trend
trend_reg<-tslm(unrtrain ~ trend)
summary(trend_reg)
checkresiduals(trend_reg)
#seasonal
seas_reg<-tslm(unrtrain ~ season)
summary(seas_reg)
checkresiduals(seas_reg)
#trend and seas
trend_seas_reg<-tslm(unrtrain ~ trend + season)
summary(trend_seas_reg)
checkresiduals((trend_seas_reg))

summary(trend_reg)
summary(seas_reg)
summary(trend_seas_reg)

#Q5
```

```
bestmodel<-rbind(CV(trend_reg), CV(seas_reg), CV(trend_seas_reg))
bestmodel

#Q6
unr_forecast<-window(ts_unrate, start=2010, end=c(2020,2))
#trend
forecast_trend<-forecast(trend_reg, newdata = unr_forecast)
autoplot(ts_unrate, series="Data")+autolayer(forecast_trend, level = FALSE,PI = TF
  xlab("Month") +ylab("Unemployment Rate (in %)") +
  ggtitle("Trend: Civilian Monthly Unemployment Rate")
#season
forecast_season<-forecast(seas_reg, newdata = unr_forecast)
autoplot(ts_unrate, series="Data")+autolayer(forecast_season, level = FALSE,PI = T
  xlab("Month") +ylab("Unemployment Rate (in %)") + ggtitle("Seasonal: Civilian Mc
#trend and season
forecast_trend_season<-forecast(trend_seas_reg, newdata = unr_forecast)
autoplot(ts_unrate, series="Data") +
  autolayer(forecast_trend_season, level = FALSE ,PI = TRUE ,series="Fitted") +
  xlab("Month") + ylab("Unemployment Rate (in %)") +
  ggtitle("Trend and Season: Civilian Monthly Unemployment Rate")

#Q7
forecast_accuracy<-rbind(accuracy(forecast_trend, unr_forecast),
                          accuracy(forecast_season, unr_forecast), accuracy(forecast
forecast_accuracy


#Q8
finale_tslm<-tslm(ts_unrate~ trend + season)
finale_tslm_for<-forecast(finale_tslm, h=12)
autoplot(ts_unrate) + ggtitle("Civilian Monthly Unemployment Rate") + xlab("Year")
autoplot(ts_unrate) + autolayer(forecast(finale_tslm, h=12), series="Trend with se
  ggtitle("12-Month Forecast for Civilian Monthly Unemployment Rate (Trend with se
finale_tslm_for
```