ECON 5P04 Forecasting Assignment: Econometrics of Time Series

Madeline Arnott (5458195)
James Teasdale (5469523)
Professor JF Lamarche

April 29, 2019

Contents

1 Introduction

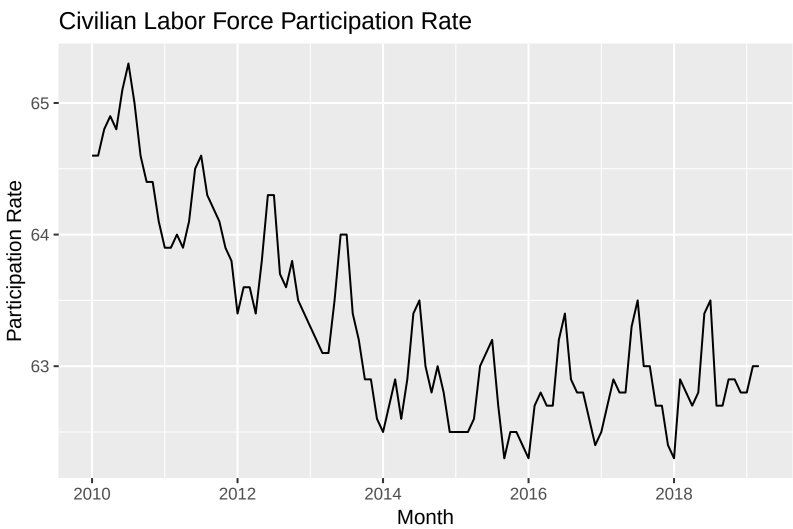
The civilian labor force participation rate is a measurement of an economy's active labor force. It is calculated by calculating the sum of all employed civilian workers and dividing by the working age civilian population. This is a measurement of people that are either looking for employment or currently employed, at or above the legal working age of 16.

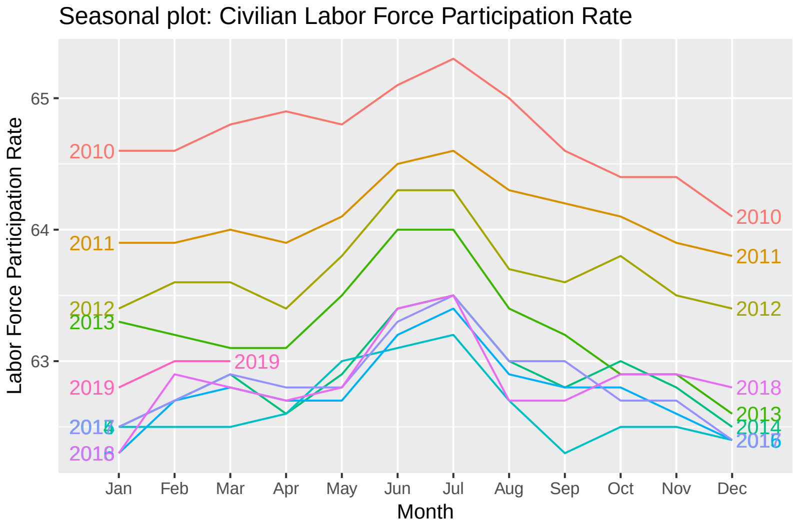
The data was retrieved from the St. Louis Federal Reserve Bank, Federal Reserve Economic Data (FRED). It consists of monthly, not seasonally adjusted U.S. data from January 1, 2010 to March 1, 2019. The data was extracted from the U.S. Bureau of Labor Statistics' "Current Population Survey (Household Survey)".

In order to determine the accuracy of our forecasting methods, we split the data into a training set (Jan. 2010 - Dec. 2015) in which we build the model, and a test set (Jan. 2016 - Mar. 2019) in which we test our model against actual observed data. We will then determine which forecasting method provides the most accurate results, and use that method to make forecasts over the next 12 month period (Apr. 2019 - Mar. 2020).

2 Trends and Seasonality

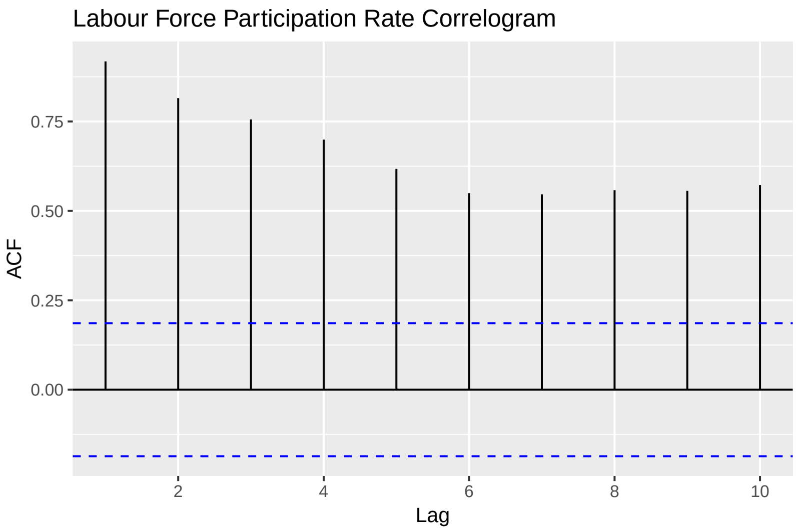
Upon creating time series and seasonal plots for the full range of the data set (2010 to 2019), it is evident that there is a steadily decreasing trend in the number of civilians in the labour force during the first half (2010 to 2014), stabilizing in the second half (2014 to 2019). This could be due to recessions making it harder to get back to work as unemployment is used as a lagging indicator, the types of jobs presently available, or that "Baby Boomers" are starting to retire, and due to modern medicine, are living longer.
We also observe a seasonal effect in the summer months of each year, beginning to increase in May and tapering off in July/August. This is likely explained by the observed influx of seasonal employment opportunities offered during the summer months as businesses increase production of goods and services during their busiest season. This also could also be explained by more people being inclined to participate as the temperature increases, or as people suffering from seasonal affective disorder begin to notice their symptoms decreasing and begin to look for work again.





3 Revealed Patterns and Autocorrelation

We can observe that our data is not White Noise as demonstrated in the previous two plots. There is clearly a seasonal component, as well as a declining linear trend over the first 6 years of the data. Each of the lags in the Correlogram is above the band, indicating positive correlation with the previous period observed value. This is intuitive since the job destruction rate is less than 1, and employers typically hire workers for more than one period to maximize profits, thus the proportion of workers in one period is related the proportion of workers in the previous periods, since unemployment and thus its counterpart, employment, are considered as lagging indicators.



4 Average, Naive and Seasonal Naive Forecasts: Training Set

Table One presents the resulting forecasts generated using the Average (Mean), Naive, and Seasonal Naive forecasting methods over the in-sample data beginning in January of 2010, and concluding December 2015. The measures of Forecast Accuracy are presented in Table 1 below:

Table 1: Measures of Forecast Accuracy

		Average	**Naive**	**Seasonal Naive**
RMSE	Training Set	0.7749502	0.2605520	0.4753946
	Test Set	0.7688055	0.5281802	0.3008535
MAE	Training Set	0.6584877	0.2056338	0.4200000
	Test Set	0.7117521	0.4512821	0.2589744
MAPE	Training Set	1.035100	0.3236017	0.6624247
	Test Set	1.134781	0.7161557	0.4117047
MASE	Training Set	1.567828	0.4896043	1.0000000
	Test Set	1.694648	1.0744811	0.6166056

Clearly, the best fitting method over the training set is the Naive Method, as each error measurement reported in the lowest of the 3 methods. This could be due to the downward trend where the beginning observations are skewing the average too high to best represent the period in entirety.

5 Determining Best Forecast Method

As displayed in the following figure, the mean and naive forecasts are constant over the test set. Clearly, the best fitting method over the training set is the Naive Method, as each error measurement reported in the lowest of the 3 methods. This could be due to the downward trend where the beginning observations are skewing the average too high to best represent the period in entirety.

However, we suspect neither the mean nor the naive methods would be a good choice since we have a strong seasonal component in our data that is excluded, and over the test set we do not observe any sort of downward trend. We find the Seasonal Naive to minimize all the measures of error over the test set. Since we notice a closer visual relation with the observed data over the test set with the Seasonal Naive Forecast, we conclude this is the best forecasting method to use. See the figures on the following page for a visual interpretation. More on this later...

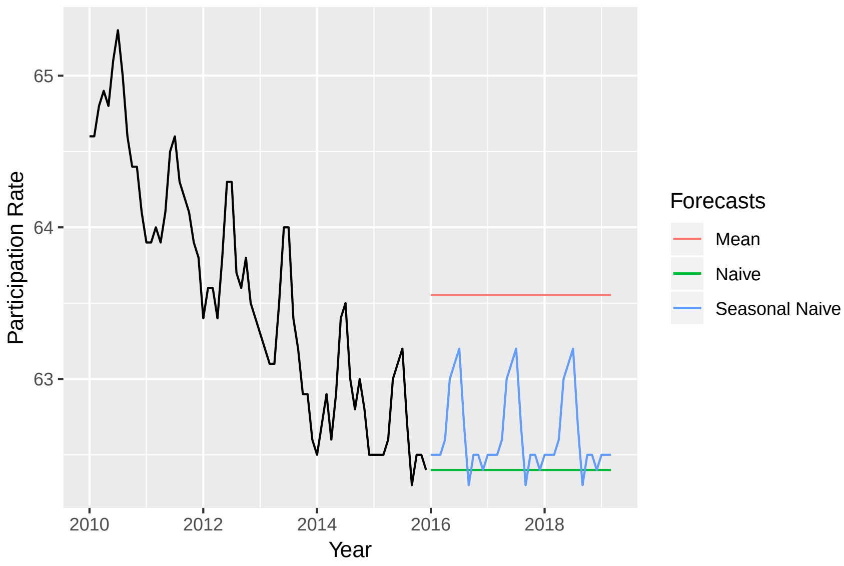


Figure 1: The Naive method is concluded to be the best method as the errors are minimized

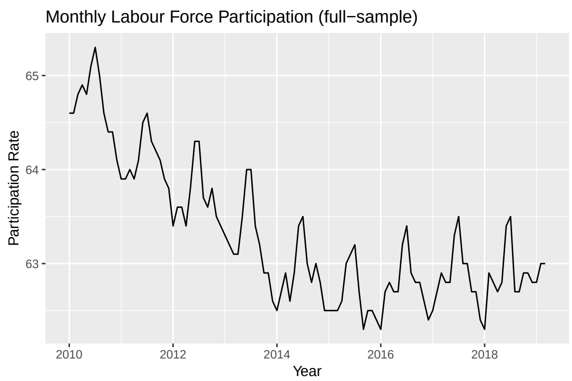


Figure 2: Compare Test Set to Seasonal Naive Forecasts in the previous figure, foreshadowing a further section within the assignment

6 Regression Based Forecasts

As displayed in table1, the R squared and Adjusted R squared are rather low for the Seasonal Dummies Model. The adjusted R squared is highest in the model using both the Trend and Seasonal Component; thus, we have selected to be the best and most accurate for our forecasts. We can also observe the This can be explained due the the rather steep downward sloping trend evident over the beginning sample. It would be naive of us to use this method solely based on its R squared measurements since the observed data in the out-sample does not reflect this sort of trend. In fact, it reflects somewhat of an increase or a rise in the Labour Force Participation.

Table 2: Forecasts

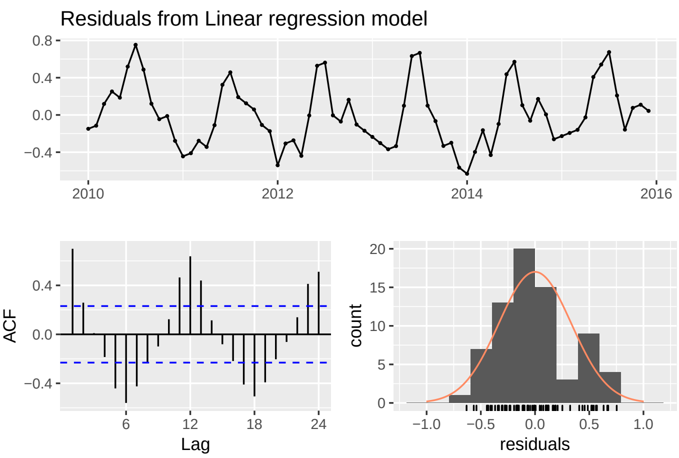
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2019	.	.	.	62.7	62.8	63.4	63.5	62.7	62.7	62.9	62.9	62.8
2020	62.8	63.0	63.0	.	.	.	.	.	.	.	.	.

6.1 Trend Only

As mentioned earlier, there is a strong downward trend in the training set, which isn't observed in the test set. This contributes to the reason why we reject any model with a trend component, since the trend does not seem to be indicative of future values. However if we only diagnose the fit the relevance in terms of fitting the in-sample data, the Trend and Seasonal Method should be used as the residuals are minimized, and are most closely centred around 0 in the bell-shaped distribution.

Table 3: Trend Only Regression Results

Coefficients	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)	64.782707	0.080159	808.17	¡2e-16 ***
trend	-0.033697	0.001908	-17.66	¡2e-16 ***
	$R^2 = 0.8166$	Adj.$R^2 = 0.814$	F-statistic = 311.7	

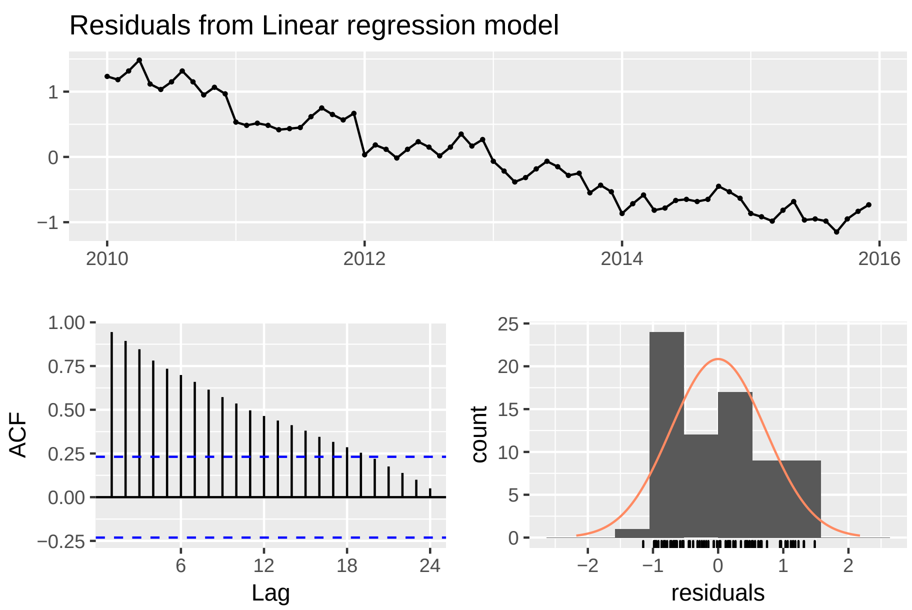


6.2 Seasonal Dummies

Below is a table reflecting the regression results using a seasonal forecast method over the training set. Both R squared values are too low, and the residualsare not centered about 0. According to these metrics and visual interpreations, this would not be a good model to use. We discuss later why this model gives us difficulty when only using the training set to model relationships within the data.

Table 4: Seasonal Dummies Regression Results

Coefficients	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)	63.36667	0.32223	196.649	¡2e-16 ***
season2	0.05000	0.45570	0.110	0.9130
season3	0.11667	0.45570	0.256	0.7988
season4	0.05000	0.45570	0.110	0.9130
season5	0.31667	0.45570	0.695	0.4898
season6	0.70000	0.45570	1.536	0.1298
season7	0.78333	0.45570	1.719	0.0908
season8	0.31667	0.45570	0.695	0.4898
season9	0.08333	0.45570	0.183	0.8555
season10	0.08333	0.45570	0.183	0.8555
season11	-0.03333	0.45570	-0.073	0.9419
season12	-0.23333	0.45570	-0.512	0.6105
	R^2 = 0.1355	Adj.R^2 = -0.02298	F-statistic = 0.855	

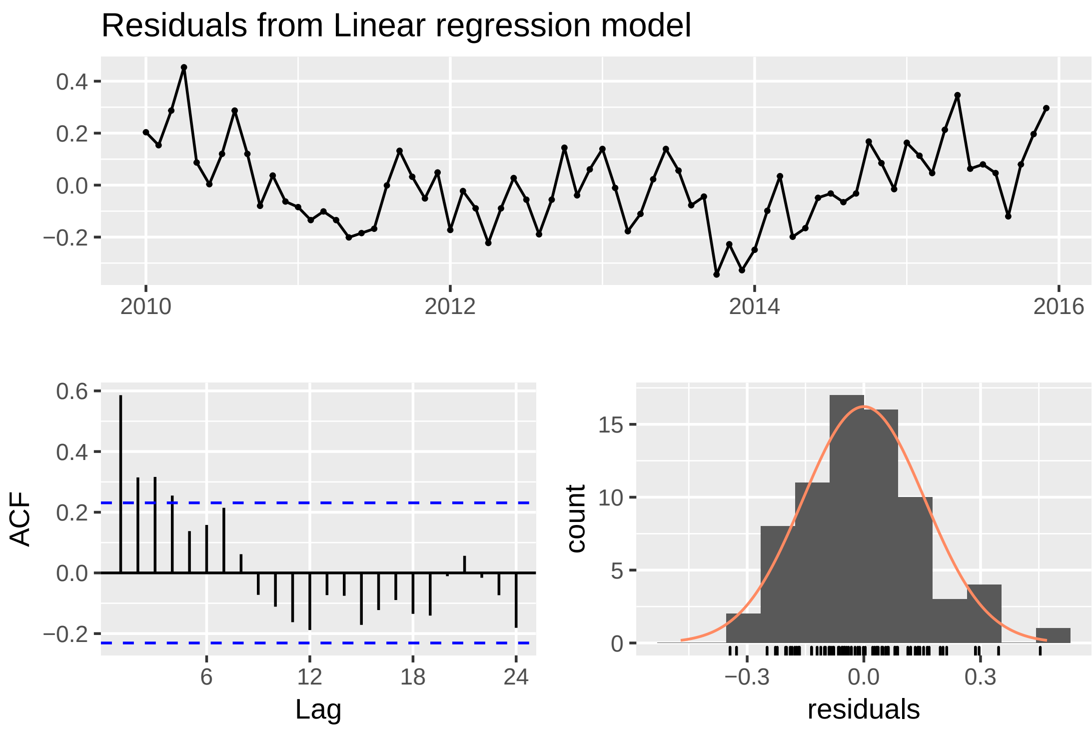


6.3 Seasonal with Trend

We find the highest R squared values as well as the most centred distribution of residuals. Thus, we would suggest using the seasonal with a trend model if only incorporating the in-sample data into our forecast.

Table 5: Seasonal Dummies with Trend Regression Results

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	64.4307540	0.0767262	839.749	< 2e-16 ***
trend	-0.0343254	0.0009904	-34.656	< 2e-16 ***
season2	0.0843254	0.0994451	0.848	0.399888
season3	0.1853175	0.0994599	1.863	0.067408 .
season4	0.1529762	0.0994846	1.538	0.129471
season5	0.4539683	0.0995191	4.562	2.62e-05 ***
season6	0.8716270	0.0995634	8.754	2.94e-12 ***
season7	0.9892857	0.0996176	9.931	3.33e-14 ***
season8	0.5569444	0.0996816	5.587	6.19e-07 ***
season9	0.3579365	0.0997554	3.588	0.000678 ***
season10	0.3922619	0.0998389	3.929	0.000226 ***
season11	0.3099206	0.0999322	3.101	0.002954 **
season12	0.1442460	0.1000352	1.442	0.154603
	R^2 = 0.9595	Adj.R^2 = 0.9513	F-statistic = 116.5	



7 Determining Best Model

7.1 The Best Model as per the CV method:

The coefficient of variation (CV) is used when comparing the dispersion of data points around the mean. In our case, this is the dispersion from our model values to our actual values. The coefficient of variation is the ratio of the standard deviation to the mean. Since it is used as a measure of fit, a smaller CV indicates that the model more closely represents the actual data.

In our case, the Seasonal and Trend model gives us the lowest coefficient of variation. It also reports the largest adjusted R^2 value. This tells us that the Seasonal and Trend model best represents the actual observations .

Table 6: Measurements of Fit

	Trend Only	Seasonal Only	Seasonal and Trend
CV	0.11623806	0.74760000	0.03650929
AIC	-152.84500	-21.19818	-239.61783
AICc	-152.49206	-14.92232	-232.24941
BIC	-146.015005	8.398477	-207.744503
AdjR^2	0.81401383	-0.02297799	0.95128941

Since these values measure the dispersion over the training set, it is clear that the in-sample model must have a trend component, otherwise there will be too much variation around the mean. This can be seen by the high CV value and low (negative?) adjusted R^2 values.

8 Test Set Forecasts

When comparing the forecasts of the training set to the actual values, we can use graphs to visually check how accurate each model is. In the cases where a trend is involved, we can see that the forecast differs radically from the actual data, with variation growing over time. Although the seasonal only model still deviates substantially from the actual data, it is the only model that does not have an increasing variation as time increases.

Table 7: Trend Only Forecasts

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2016	62.33	62.29	62.26	62.22	62.19	62.15	62.12	62.09	62.05	62.02	61.99	61.95
2017	61.92	61.88	61.85	61.82	61.78	61.75	61.72	61.68	61.65	61.62	61.58	61.55
2018	61.51	61.48	61.45	61.41	61.38	61.35	61.31	61.28	61.24	61.21	61.18	61.14
2019	61.11	61.08	61.04	.	.	.	.	.	.	.	.	.

See the following figure for a visual representation of the Trend Forecast over the Test Set:

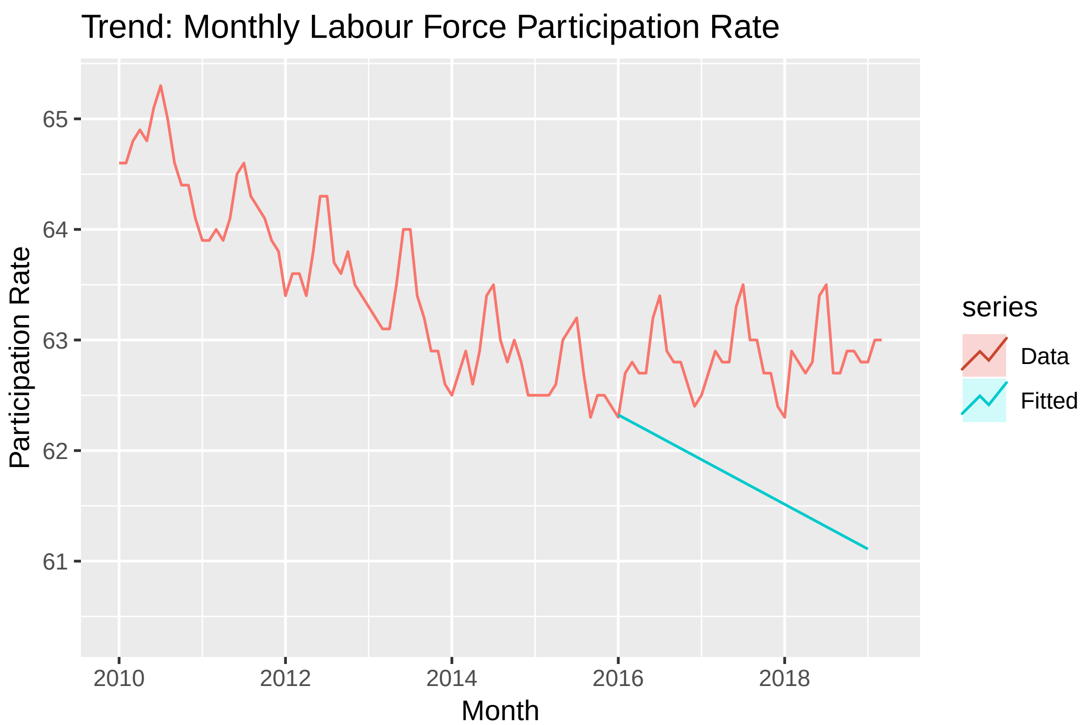


Table 8: Seasonal Only Forecasts

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2016	63.37	63.42	63.48	63.42	63.68	64.07	64.15	63.68	63.45	63.45	63.33	61.7
2017	63.37	63.42	63.48	63.42	63.68	64.07	64.15	63.68	63.45	63.45	63.33	63.13
2018	63.37	63.42	63.48	63.42	63.68	64.07	64.15	63.68	63.45	63.45	63.33	63.13
2019	63.37	63.42	63.48	.	.	.	.	.	.	.	.	.

As per the table of Seasonal Point Forecasts displayed above, see the following figure for a visual representation of the Seasonal Forecast over the test set:

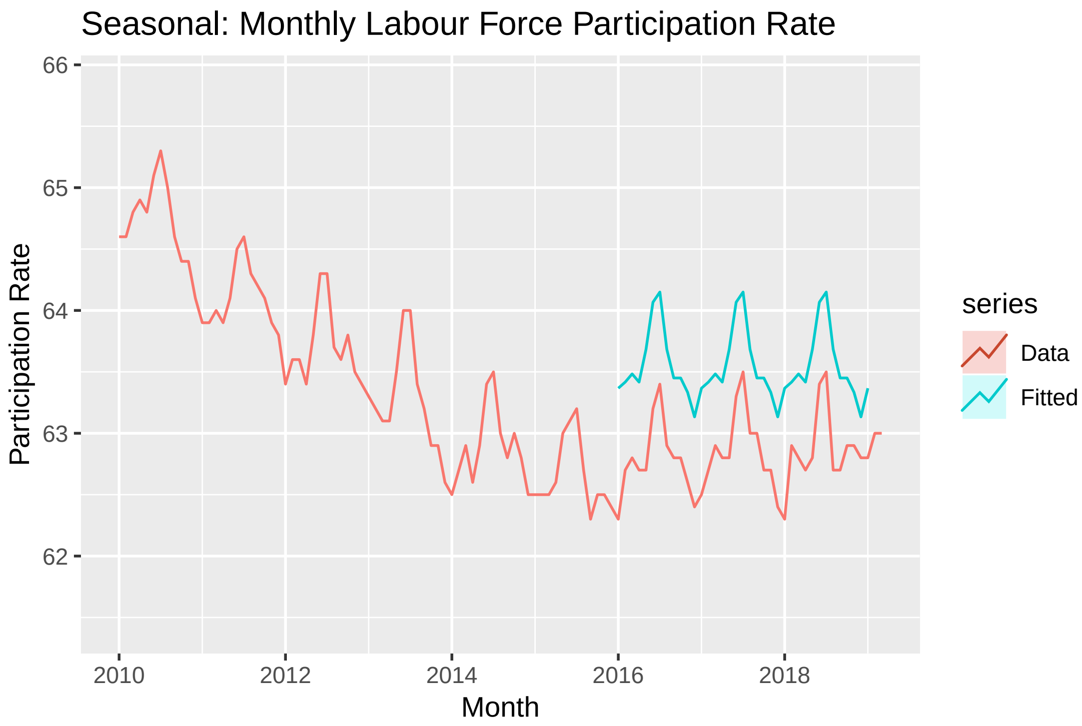
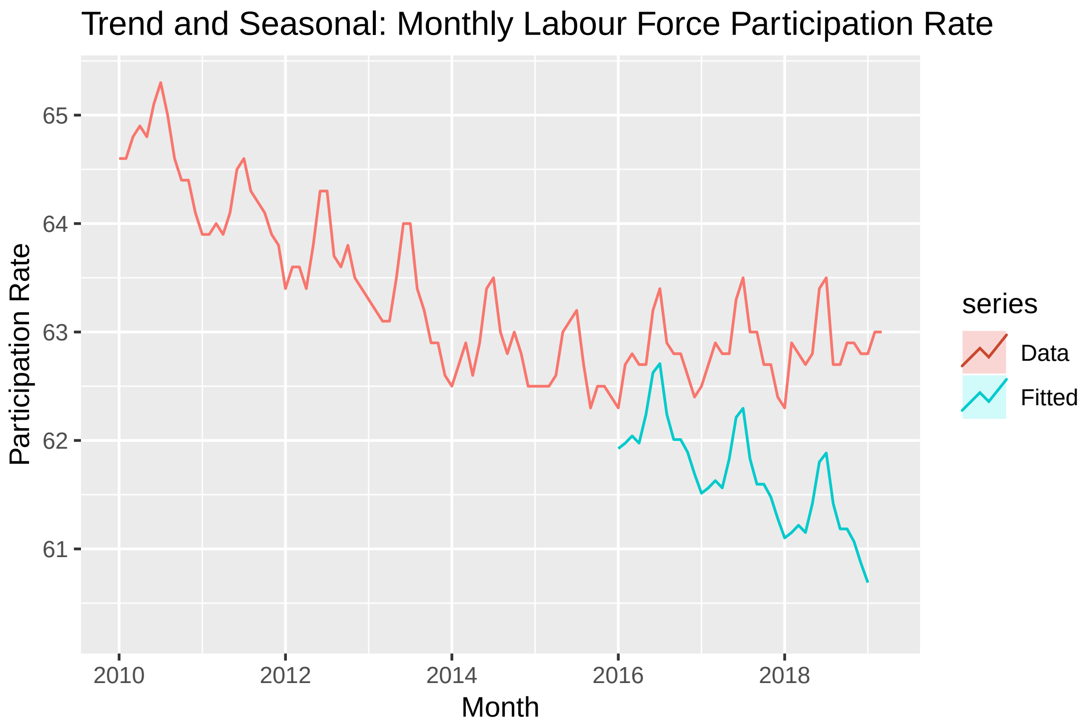


Table 9: Seasonal with Trend Forecasts

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2016	61.93	61.98	62.04	61.98	62.24	62.63	62.71	62.24	62.01	62.01	61.89	61.69
2017	61.51	61.56	61.63	61.56	61.83	62.21	62.30	61.83	61.60	61.60	61.48	61.28
2018	61.10	61.15	61.22	61.15	61.42	61.80	61.88	61.42	61.18	61.18	61.07	60.87
2019	60.69	60.74	60.81	.	.	.	.	.	.	.	.	.

As per the table of Seasonal Point Forecasts displayed above, see the following figure for a visual representation of the Trend + Seasonal Forecast over the test set:



9 Future Forecasts

In creating forecasts for the next 12 months, we have decided to use the Seasonal Only regression model. As shown in Table 10, this model produces the smallest RMSE and MAPE values, indicating that it is the best fit when compared to the full data set.

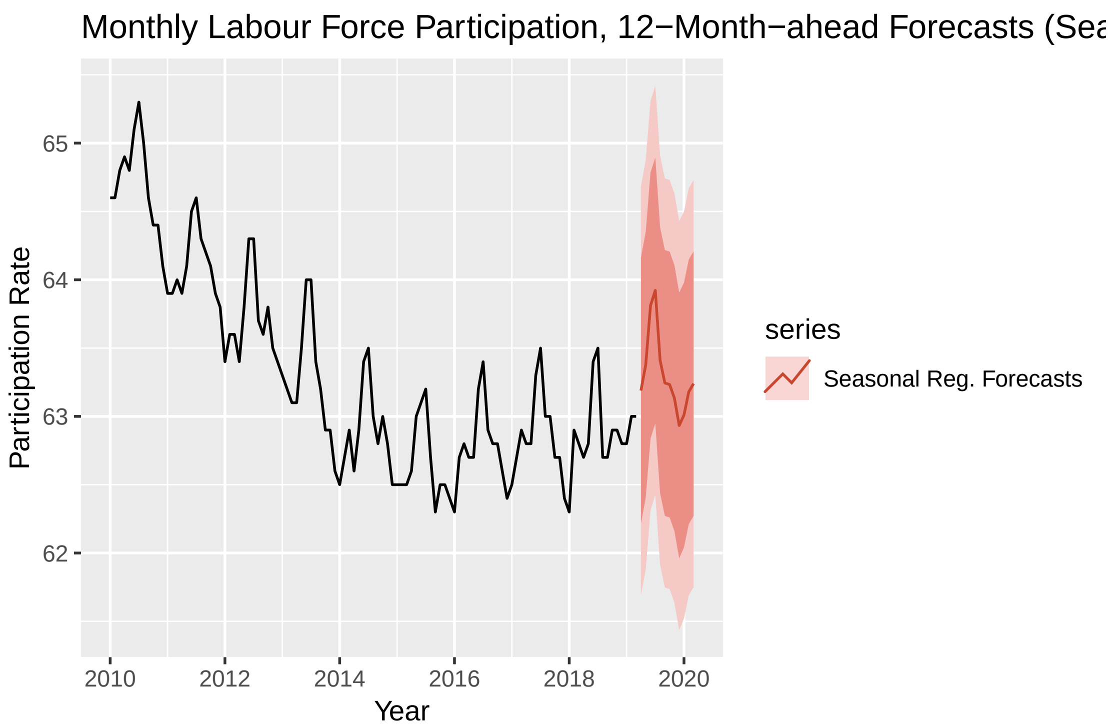
Table 10: Measures of Forecast Accuracy

	Trend Only	**Seasonal Only**	**Seasonal and Trend**
RMSE	1.2214409	0.7329647	1.2374900
MAPE	1.7757909	1.1388952	1.8463375

The Seasonal Only regression model produces a conservative 12 month forecast, which can be seen in the following table and graph. This model predicts that the civilian labor force participation rate will decrease from 62.13 to 62.08 in the next year, with minor seasonal fluctuations in between. This is a different model than our best in-sample model, and this is because the in-sample has an almost linear, downward sloping trend, while out-sample has an almost flat slope.

Table 11: Seasonal Only Model Forecasts

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2019	.	.	.	62.13	62.32	62.76	62.87	62.36	62.19	62.18	62.08	61.88
2020	61.85	62.019	62.08	.	.	.	.	.	.	.	.	.



10 Bibliography

- U.S. Bureau of Labor Statistics, Civilian Labor Force Participation Rate [LNU01300000], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LNU01300000, April 24, 2019.
- Lamarche, J. (2018, September). *Econometrics* Lectures and tutorials presented at Brock University in Econometrics 5P03, St. Catharines.
- Lamarche, J. (2019, February 12). *Time Series Graphics* Lecture presented at Brock University in Topics in Applied Economics 5P04, St. Catharines.
- Lamarche, J. (2019, March 5). *Forecasters Toolbox* Lecture presented at Brock University in Topics in Applied Economics 5P04, St. Catharines.
- Lamarche, J. (2019, March 11). *Time Series Regression* Lecture presented at Brock University in Topics in Applied Economics 5P04, St. Catharines.
- Lamarche, J. (2019, March 12). *Arima Models Short* Lecture presented at Brock University in Topics in Applied Economics 5P04, St. Catharines.

11 Figures

12 R Code

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


df<-read.csv("DATA.csv", header = TRUE)
cpr <- df$LNU01300000

# Creating time series object
tscpr <- ts(cpr,start=c(2010,1),frequency=12)
insample <- window(tscpr, start=c(2010, 1), end=c(2015, 12))
outsample <- window(tscpr, start=c(2016, 1), end=c(2019, 3))

#question1
autoplot(tscpr) + ggtitle("Civilian Labor Force Participation Rate") + xlab("Month") +
ylab("Participation Rate") ggseasonplot(tscpr, year.labels=TRUE, year.labels.left=TRUE) +
ylab("Labor Force Participation Rate") +
ggtitle("Seasonal plot: Civilian Labor
Force Participation Rate")

#question2
```

```
ggAcf(tscpr, lag=10) + ggtitle("Labour Force Participation Rate Correlogram") + xlab("Lag") +
ylab("ACF") labour_autocorr<-ggAcf(tscpr, plot= FALSE) labour_autocorr10<-labour_autocorr[1:10,]

#question3: forecasts
mean_for<-meanf(insample, h=39)
listedmeanfor<-mean_for$mean
naive<-naive(insample, h=39)
listednaive<-naive$mean
naives_for<-snaive(insample, h=39)
listednaives_for<-naives_for$mean

#plot of forecasts
insample_lfp<-window(tscpr, start=2010, end=c(2015,12))
autoplot(insample) + ggtitle("Monthly Labour Force Participation") + xlab("Year") +
ylab("Participation")

autoplot(insample) + autolayer(meanf(insample, h=39)$mean, series="Mean")
+ autolayer(naive(insample, h=39)$mean, series= "Naive")
+ autolayer(snaive(insample,h=39)$mean, series="Seasonal Naive")
+ xlab("Year") + ylab("Participation Rate") +guides(colour=guide_legend(title="Forecasts"))

#accuracy
acc_tscpr<-window(tscpr, start=2016)
accuracy(mean_for, acc_tscpr)
accuracy(naive, acc_tscpr)
accuracy(naives_for, acc_tscpr)

#Q4 using Seasonal Naive
seasnaivefullsampletwelvefore<-snaive(tscpr, h=12)
seasnaivefullsampletwelvefore
autoplot(tscpr) + ggtitle("Monthly Labour Force Participation (full-sample)")
+ xlab("Year") + ylab("Participation Rate")

#plotted
autoplot(tscpr) + autolayer(snaive(tscpr, h=12), series="Seasonal Forecasts (with CI)")+
ggtitle("LF Participation 12-Month Forecasts") +  xlab("Year") + ylab("Participation Rate")

#Q5
#Trend
trend_reg<-tslm(insample ~ trend)
summary(trend_reg)
checkresiduals(trend_reg)
#seasonal
seas_reg<-tslm(insample ~ season)
summary(seas_reg)
checkresiduals(seas_reg)
#trend and seas
trend_seas_reg<-tslm(insample ~ trend + season)
```

```
summary(trend_seas_reg)

checkresiduals((trend_seas_reg))

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

#Q6

checkmodel<-rbind(CV(trend_reg), CV(seas_reg), CV(trend_seas_reg))
checkmodel

#Q7 forecasting ocer test set... plot over total data

#trend reg
outsample<-window(tscpr, start=2016, end=c(2019,3))
trend_for<-forecast(trend_reg, newdata = outsample)
autoplot(tscpr, series="Data")+autolayer(trend_for, level = FALSE,PI = TRUE,series="Fitted") +
xlab("Month") +ylab("Participation Rate") +
ggtitle("Trend: Monthly Labour Force Participation Rate")

#seas reg
seas_for<-forecast(seas_reg, newdata = outsample)
autoplot(tscpr, series="Data")+autolayer(seas_for, level = FALSE,PI = TRUE,series="Fitted") +
xlab("Month") +ylab("Participation Rate") + ggtitle("Seasonal: Monthly Labour Force
Participation Rate")

#trend+seas reg
trend_seas_for<-forecast(trend_seas_reg, newdata = outsample)
autoplot(tscpr, series="Data")+autolayer(trend_seas_for, level = FALSE,PI = TRUE,
series="Fitted") + xlab("Month") +ylab("Participation Rate") + ggtitle("Trend and Seasonal:
Monthly Labour Force Participation Rate")

trendforecast<-data.frame(trend_for)
seasonforecast<-data.frame(seas_for)
trendandseasonforecast<-data.frame(trend_seas_for)

pointforecasts<-cbind(trendforecast$Point.Forecast, seasonforecast$Point.Forecast,
trendandseasonforecast$Point.Forecast)
pointforecasts

names(pointforecasts[1])<-"Trend forecast"
col.names(pointforecasts[2])<-"Season forecast"
names(pointforecasts[3])<-"Trend and Season"
pointforecasts

forecasts<-rbind(trend_for, seas_for, trend_seas_for)
forecasts
```

```
#Q8
forecastaccuracy<-rbind(accuracy(trend_for, outsample),
accuracy(seas_for, outsample), accuracy(trend_seas_for,outsample))
forecastaccuracy

#Q9 fav model - forecast next 12 months

total_tslm<-tslm(tscpr ~ season)
total_tslm_for<-forecast(total_tslm, h=12)
autoplot(tscpr) +ggtitle("Monthly Labour Force Participation,) + xlab("Year") +
ylab("Participation Rate")

#plotted
autoplot(tscpr) +autolayer(forecast(total_tslm, h=12), series="Seasonal Reg. Forecasts")+
ggtitle("Monthly Labour Force Participation, 12-Month-ahead Forecasts (Seasonal)")+ xlab("Year")
ylab("Participation Rate")
total_tslm_for
```