Assignment 7 Work

Matthew Dunne

August 5, 2018

Question 1

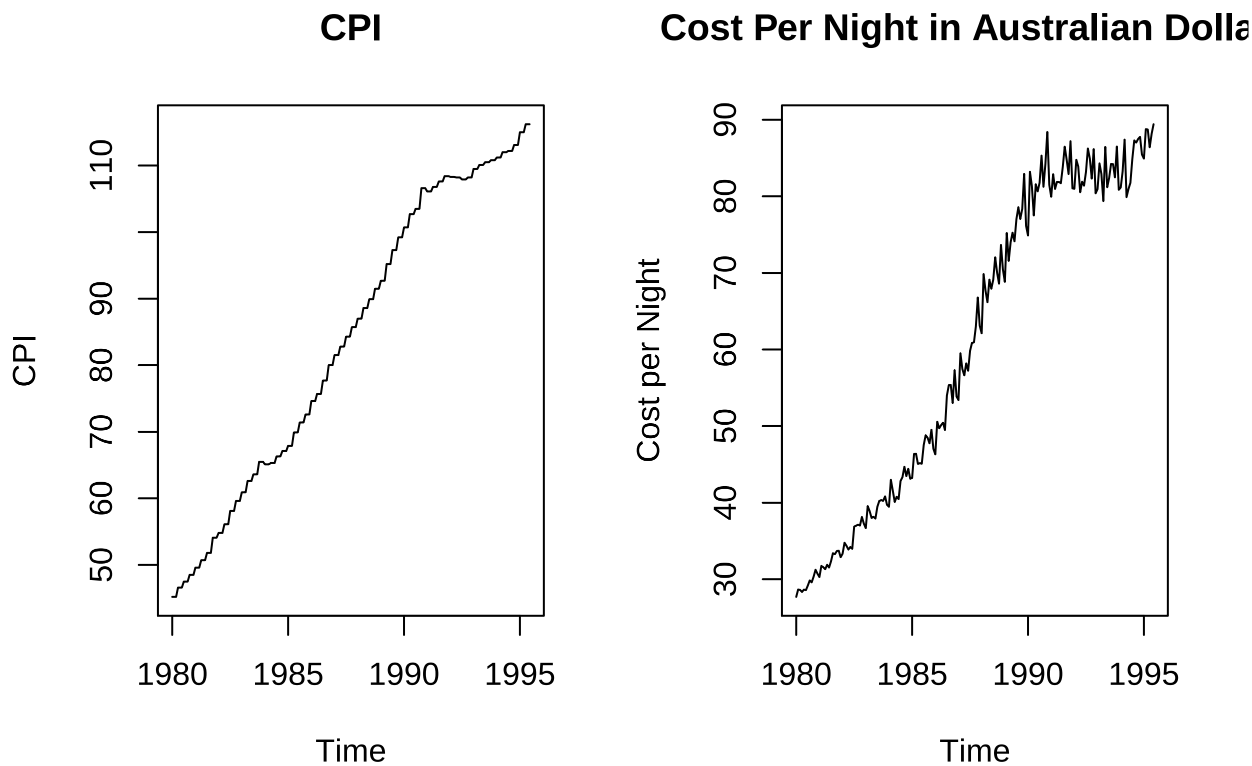
Load the data and calculate the average cost of a night's accommodation in Victoria each month (i.e. Cost variable).

```
library(fpp)
load("C:/Users/mjdun/Desktop/Time Series/Assignments/motel.rda")
data<-motel
#calculate average cost per night (Takings/RoomNights)*1000
cost<-(data[,2]/data[,1])*1000
data<-cbind(data, cost)
head(data[ ,1:4])
```

```
##          data.Roomnights data.Takings data.CPI     cost
## Jan 1980          276986         7673     45.2 27.70176
## Feb 1980          260633         7472     45.2 28.66866
## Mar 1980          291551         8339     45.2 28.60220
## Apr 1980          275383         7805     46.6 28.34235
## May 1980          275302         7889     46.6 28.65580
## Jun 1980          231693         6619     46.6 28.56798
```

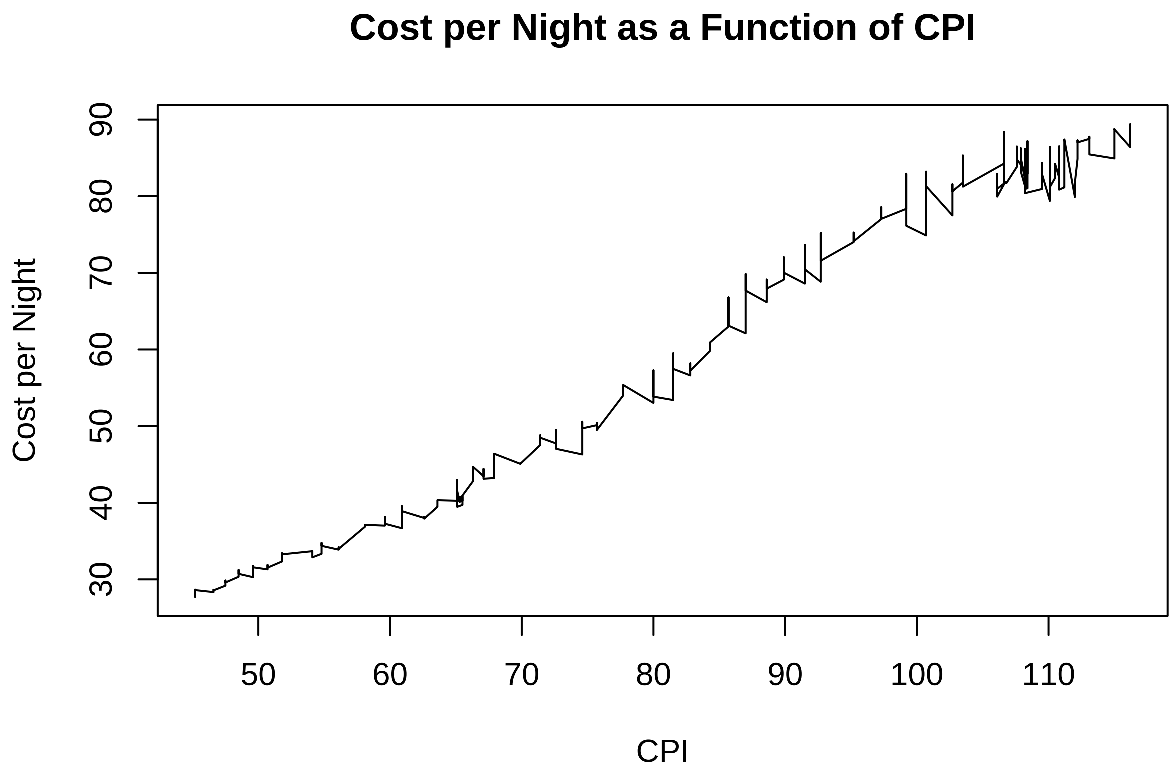
a) Plot the CPI and Cost time series

```
par(mfrow=c(1,2))
plot(data[,3], main="CPI", ylab="CPI")
plot(data[,4], main="Cost Per Night in Australian Dollars", ylab="Cost per Night")
```



b) Plot the Cost time series against the CPI time series and calculate the correlation between CPI and Cost.

```
plot(data[,4]~data[,3], main="Cost per Night as a Function of CPI", ylab='Cost per Night', xlab='CPI', 
```

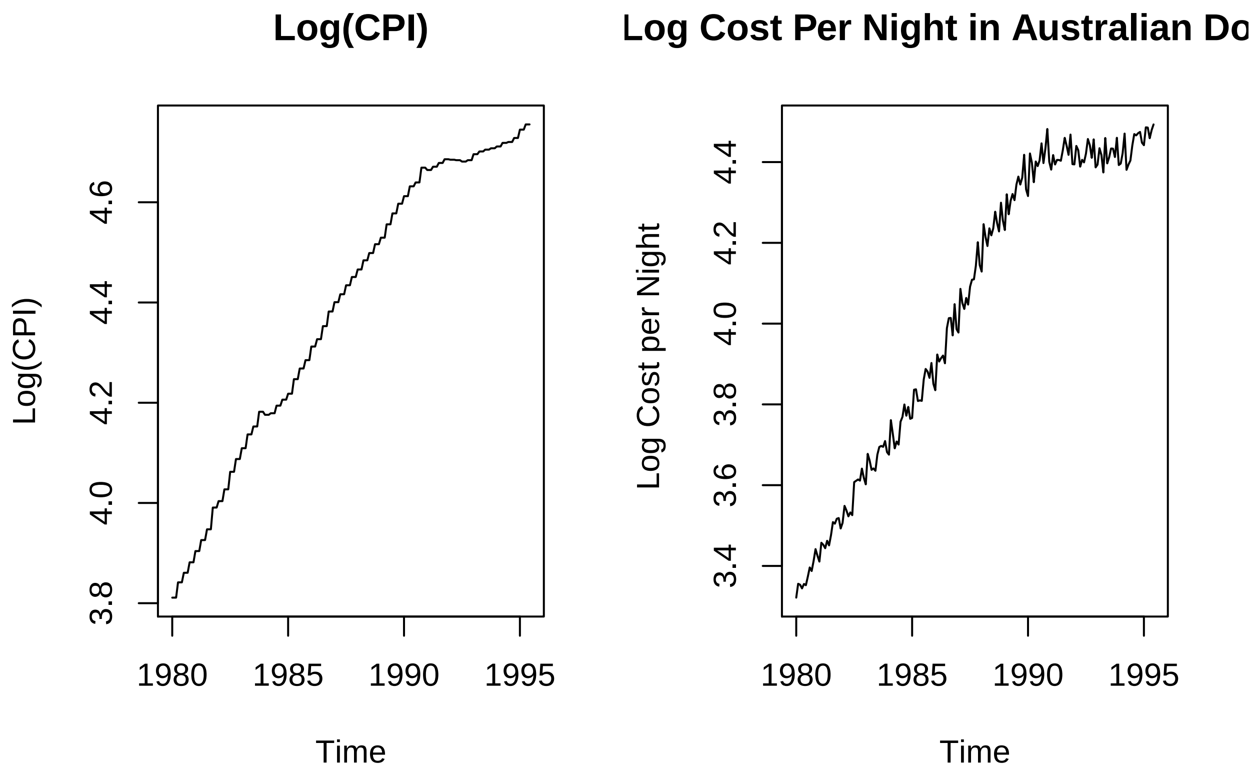


Discuss your results and explain why logarithms of both variables need to be taken before fitting any models.

The Cost per Night seems to go up in lock step with CPI. This means that Cost per Night is getting more expensive in nominal but not real terms. It is necessary to do a log transformation of the Cost variable because there is a clear change in variance as the level changes. Variance increases as a function of time. The same is not true of CPI, but in order to compare it with Cost it must be on the same scale. Log transformation of Cost means that we have to log transorm CPI if we are to accurately compare the two.

c) Plot the log(CPI) and log(Cost) time series and calculate the correlation between the logarithms of both CPI and Cost.

```
par(mfrow=c(1,2))
plot(log(data[,3]), main="Log(CPI)", ylab="Log(CPI)")
plot(log(data[,4]), main="Log Cost Per Night in Australian Dollars", ylab=" Log Cost per Night")
```




```
cor(data[,3],data[,4])
```

```
## [1] 0.9907186
```

There is a near perfect correlation between the log transformed CPI and the log transformed Cost per Night.

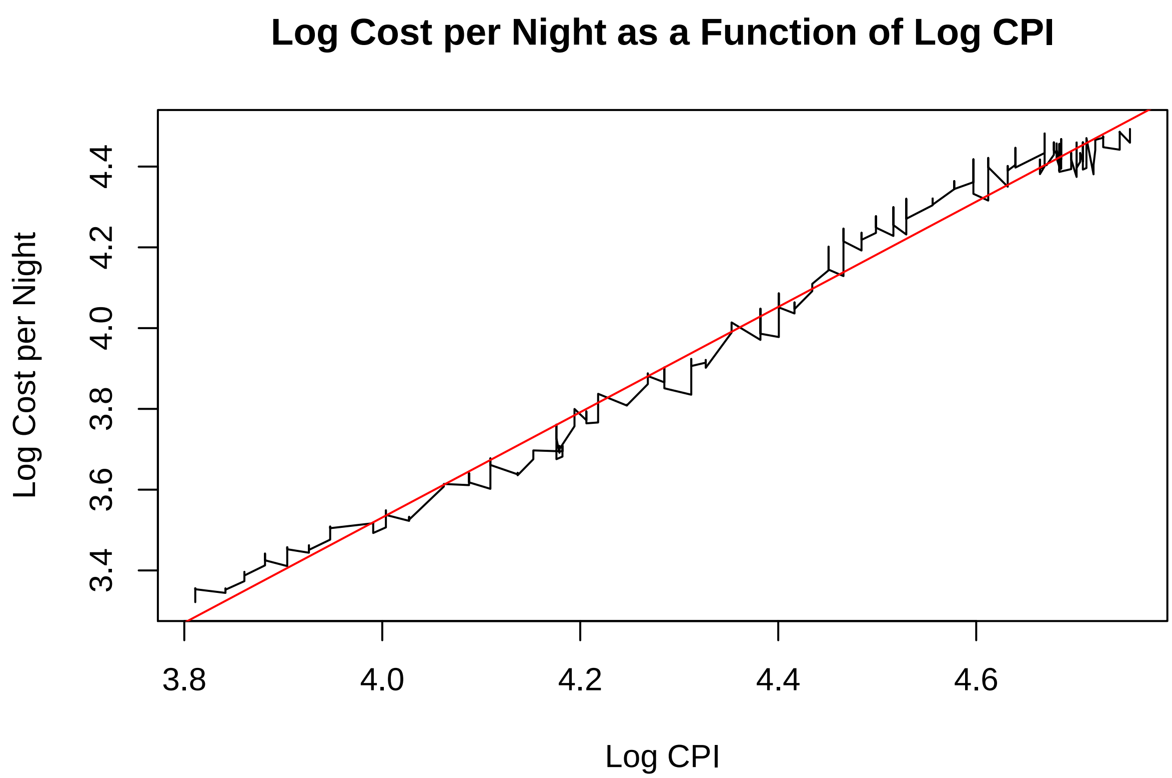
Question 2

a) Use the linear model with time series components function tslm() to fit a linear model to the Cost time series as a function of the CPI time series (i.e. CPI –> independent variable, Cost –> dependent variable). Remember to set the 'lambda' argument to 0 to reflect a logarithmic transformation.

```
logmodel<-tslm(data[,4]~data[,3], data=data, lambda=0)
```

b) Plot log(CPI) against log(Cost) and the fitted trend. Remember to apply the logarithmic transformation to the fitted trend.

```
plot(log(data[,4])~log(data[,3]), main="Log Cost per Night as a Function of Log CPI", ylab='Log Cost pe
abline(logmodel, untf = T, col=2)
```

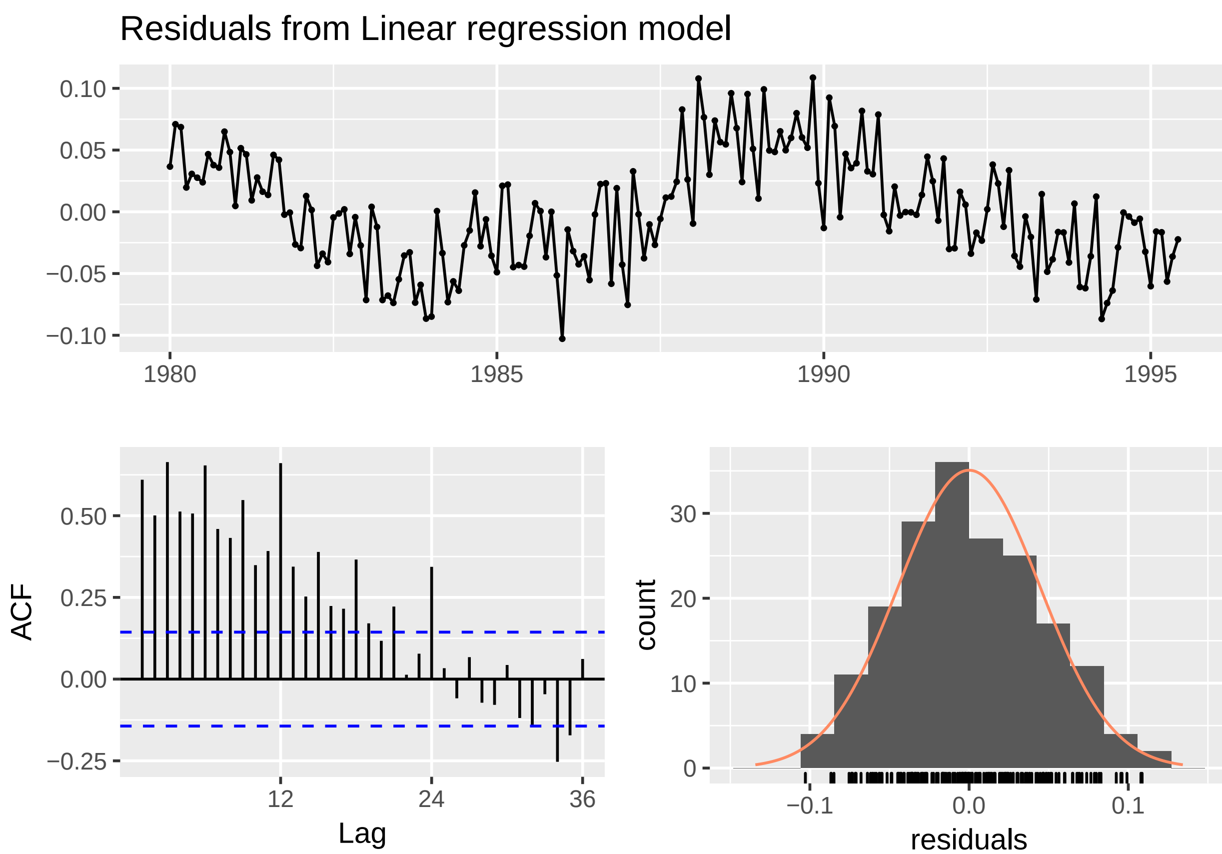


c) Use the summary() function to summarize the generated model, and the checkresiduals() function to evaluate the residuals.

```
summary(logmodel)
```

```
##
## Call:
## tslm(formula = data[, 4] ~ data[, 3], data = data, lambda = 0)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.102841 -0.033939 -0.002071  0.030363  0.108670
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.68246    0.05032  -33.44   <2e-16 ***
## data[, 3]    1.30339    0.01144  113.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04487 on 184 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 3.943e+07 on 1 and 184 DF,  p-value: < 2.2e-16
```

```
checkresiduals(logmodel)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 24
##
## data:  Residuals from Linear regression model
## LM test = 148.84, df = 24, p-value < 2.2e-16
```

The general model says that the log of Cost per Night increases by 1.3 for every increase of 1 in the log CPI. The R Squared is 1, which would suggest this model is more deterministics than stochastic.

Also we see from the checkresiduals function that are clearly not white noise. The decay in the residuals as seen in the ACF plot is too slow. And the p-value of the Breusch-Godfrey test also shows autocorrelation in the residuals

Discuss your results.

Overall we see that the Cost per Night largely moves with CPI. This makes sense given the underlying variables. We are measuring the the change in price of one thing against the price of things in general. The two should track closely.

Question 3

Use the auto.arima() function to fit an appropriate regression model with ARIMA errors to the CPI time series. Set the Order of seasonal-differencing argument, D, to 1 and the 'lambda' argument to 0 to reflect a logarithmic transformation.

```
arima_errors<-auto.arima(data[,4], D=1, lambda = 0, xreg = data[,3])
```

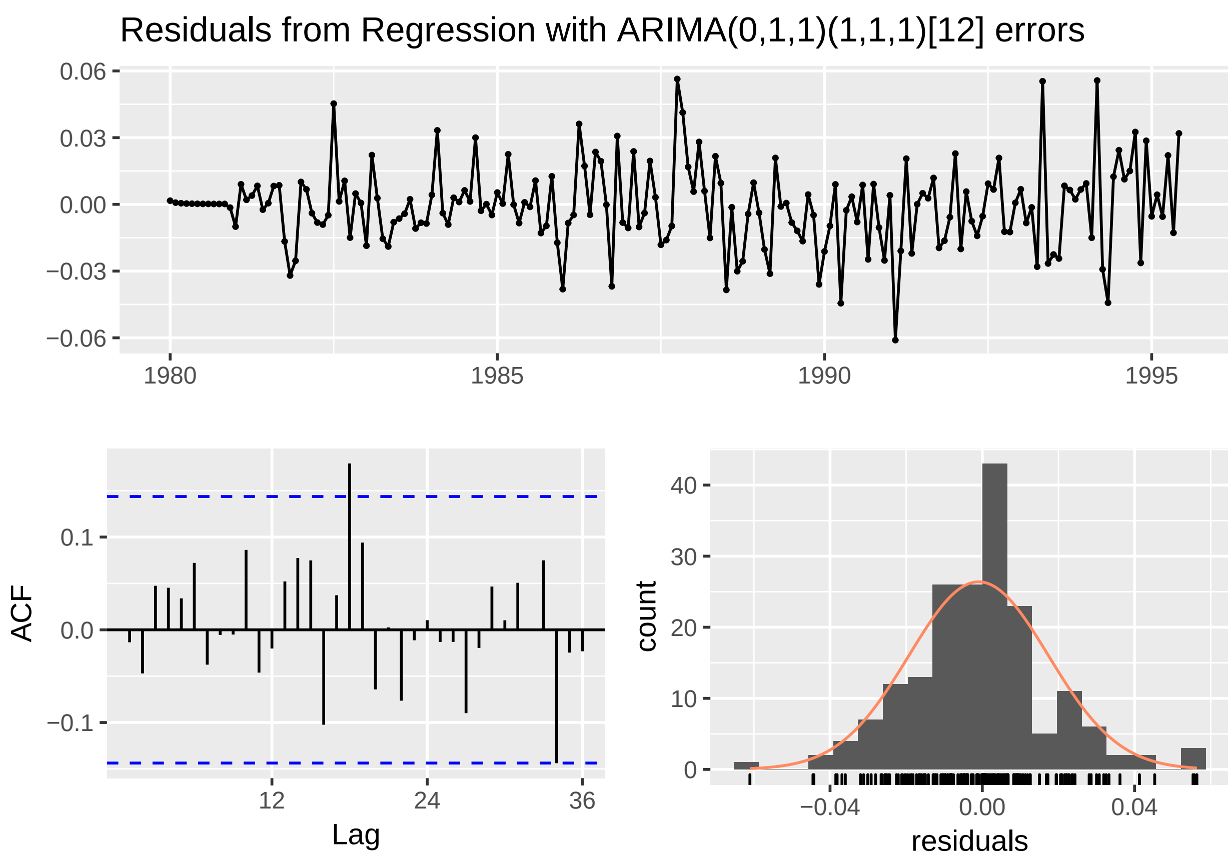
a) Use the summary() function to summarize the generated model.

```r
summary(arima_errors)
```

```
## Series: data[, 4]
## Regression with ARIMA(0,1,1)(1,1,1)[12] errors
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1      sar1     sma1    xreg
##       -0.5517  -0.3247  -0.3117  0.0099
## s.e.   0.0594   0.1521   0.1665  0.0025
##
## sigma^2 estimated as 0.0003693:  log likelihood=437.76
## AIC=-865.52   AICc=-865.16   BIC=-849.75
##
## Training set error measures:
##                       ME     RMSE       MAE        MPE    MAPE     MASE
## Training set -0.08006965 1.299879 0.8880394 -0.1126117 1.33579 0.209435
##                     ACF1
## Training set -0.07078119
```

b) Use the checkresiduals() function to evaluate the residuals.

```r
checkresiduals(arima_errors)
```




```
##
##  Ljung-Box test
##
```

```
## data:  Residuals from Regression with ARIMA(0,1,1)(1,1,1)[12] errors
## Q* = 20.853, df = 20, p-value = 0.4058
##
## Model df: 4.   Total lags used: 24
```

Discuss your results.

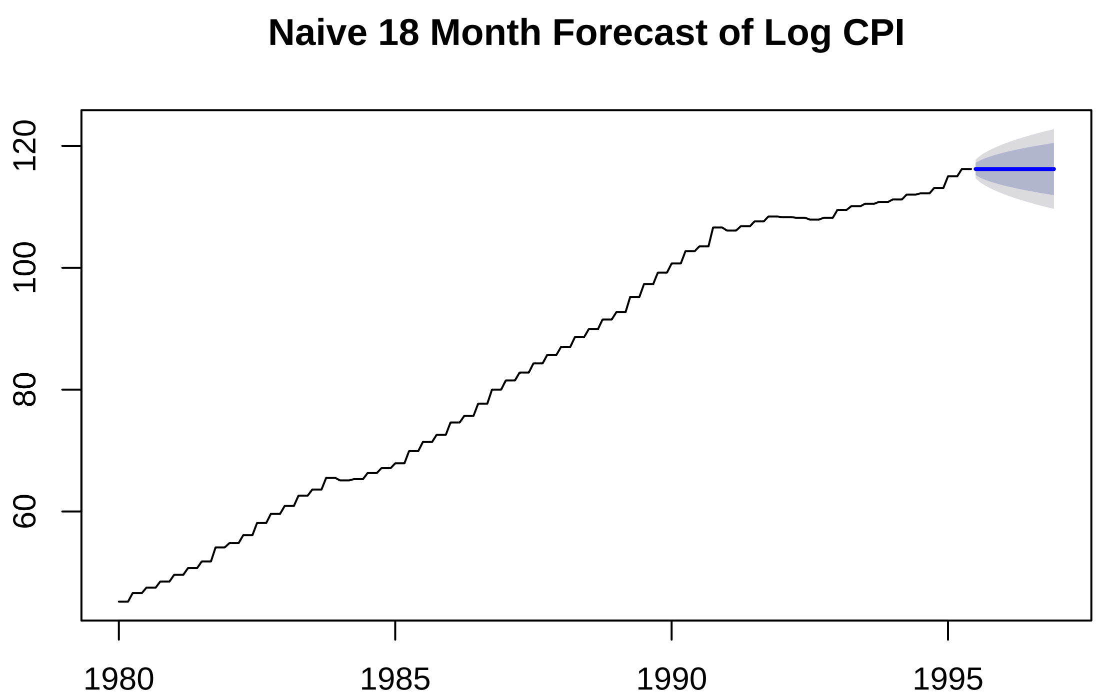
We get a ARIMA(0,1,0)(1,1,2)[12] model (seasonal). The residuals look extremely low until we remember that we are using log values rather than real values. The ACF plot and the p-value from the Ljung-Box Test say that the residuals are White Noise. The spike at lag 18 is probably just coincidental.

Question 4

a) Calculate and plot a naïve forecast of CPI for the next 18 months.

Because the model in Question 3 uses a log transformation of the data, I assume no log transformation is necessary for generating the naive forecast of CPI.

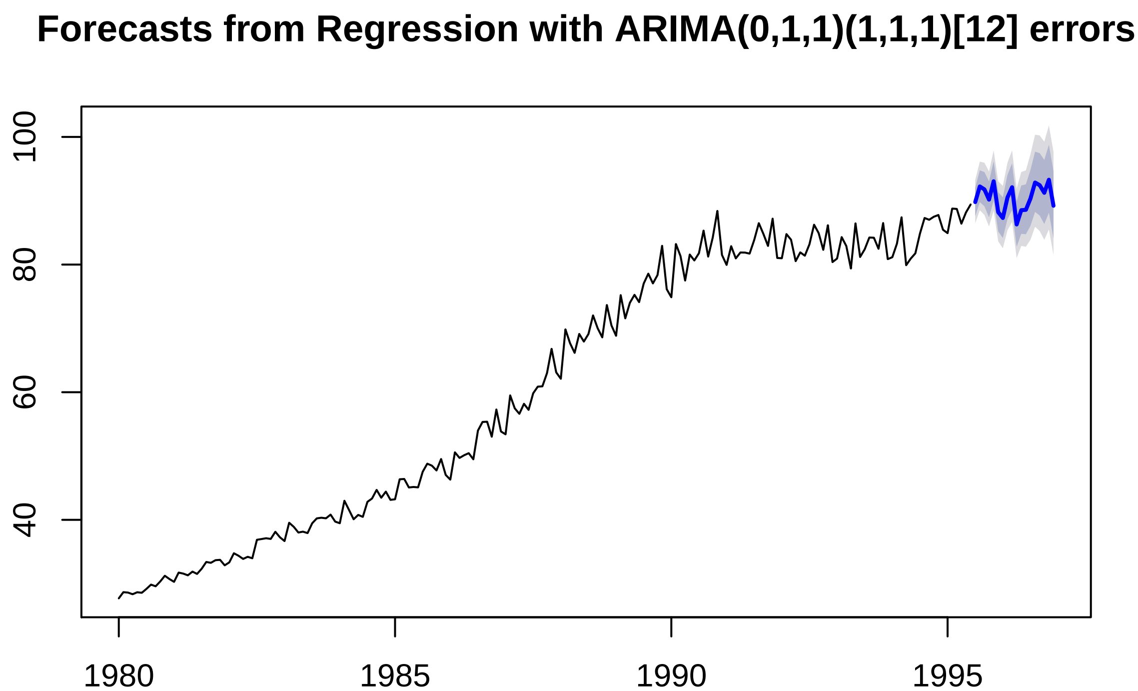
```
cpi_forecast<-naive(data[,3], h=18)
plot(cpi_forecast, main = "Naive 18 Month Forecast of Log CPI")
```




```
#get just the naive forecast values, not the entirety of CPI values, to plug into next forecast
cpi_forecast<-naive(data[,3], h=18)$mean
```

b) Forecast and plot the average price per room (i.e., Cost) for the next 18 months using the fitted model from Question 3 and the naïve forecast of CPI.

```r
#use the arima_errors model with the additional xreg input of only the 18 naive forecast values
arima_error_forecast<-forecast(arima_errors, xreg=cpi_forecast, h=18)
plot(arima_error_forecast)
```



Discuss your results.

Because our outside variable is based on a naive forecast of CPI, the level of the forecast of Cost per Night does not appear to change much if at all. The forecast does appear to mimic the seasonal character of Cost per Night. It is not a bad forecast if we assume CPI will stay level.

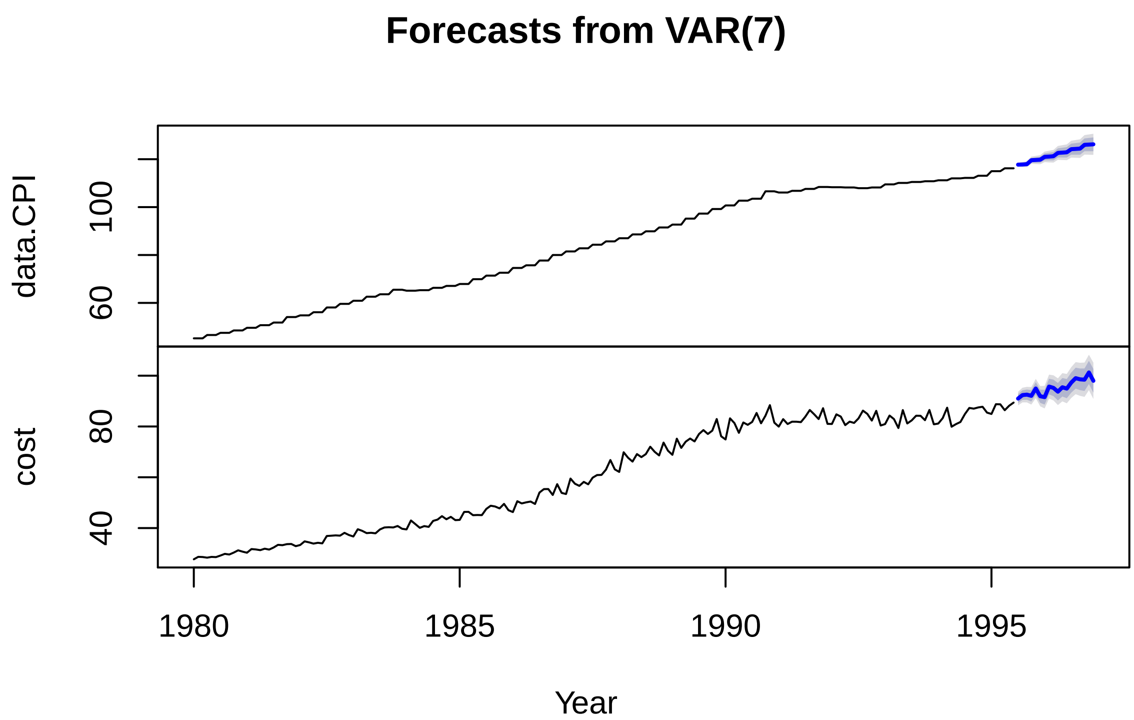
Question 5

a) Use the VAR() function to fit a VAR(7) model to the Cost and CPI time series. Set the 'type' and 'season' arguments to 'both' and 12, respectively.

```r
library(vars)
var_model<-VAR(data[,3:4], p=7, type = "both", season = 12)
```

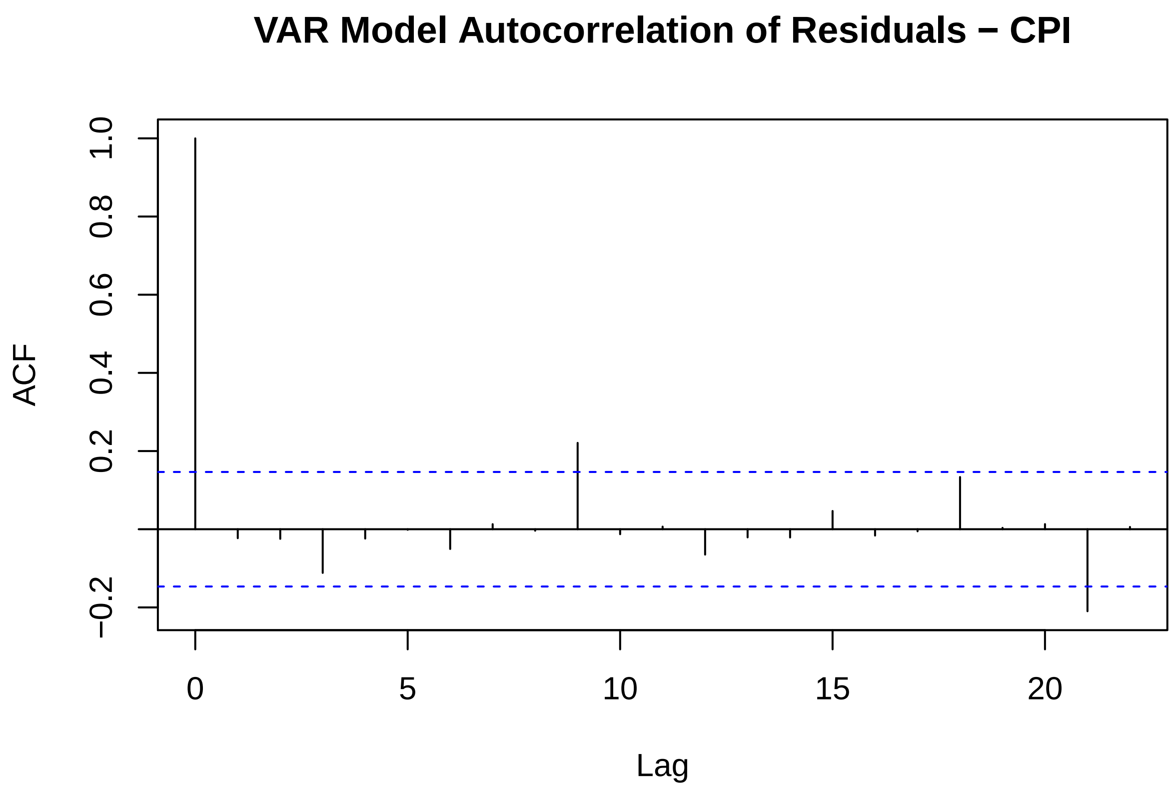
b) Forecast and plot the average price per room (i.e., Cost) and CPI for the next 18 months using your fitted model.

```r
var_forecast<-forecast(var_model, h=18)
plot(var_forecast, xlab="Year")
```

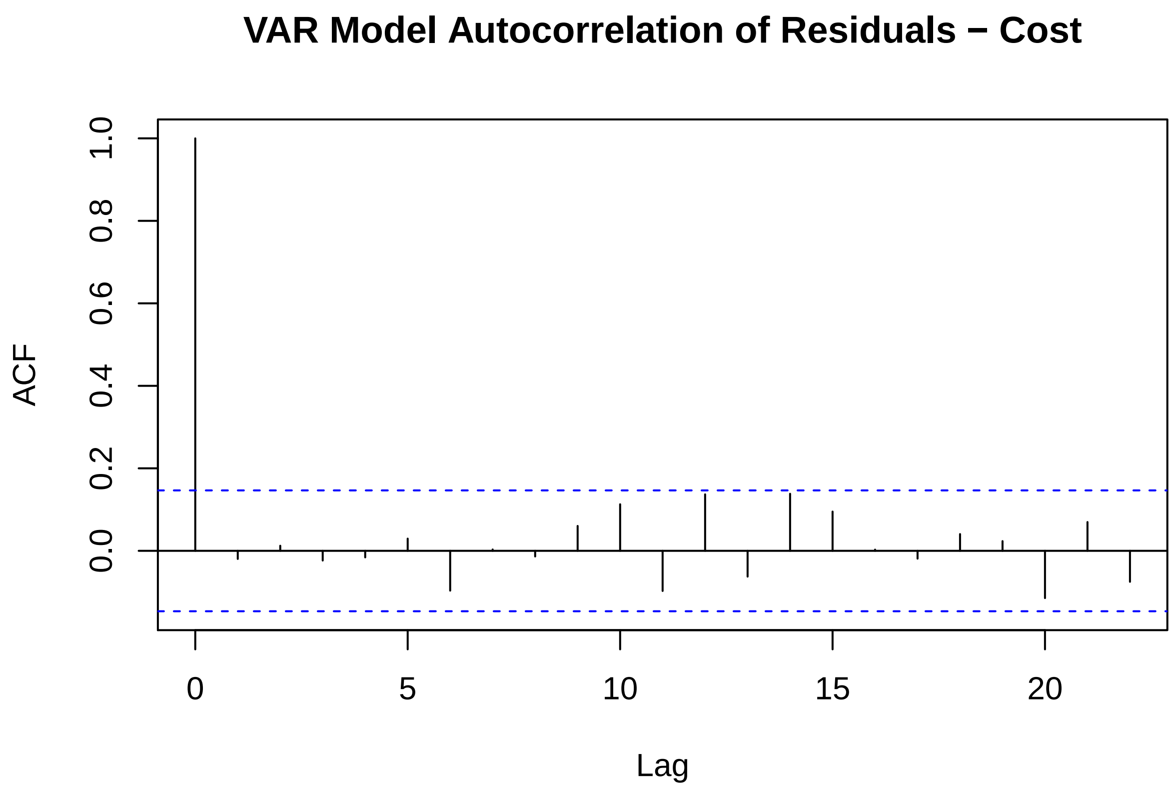


c) Plot the acf of residuals

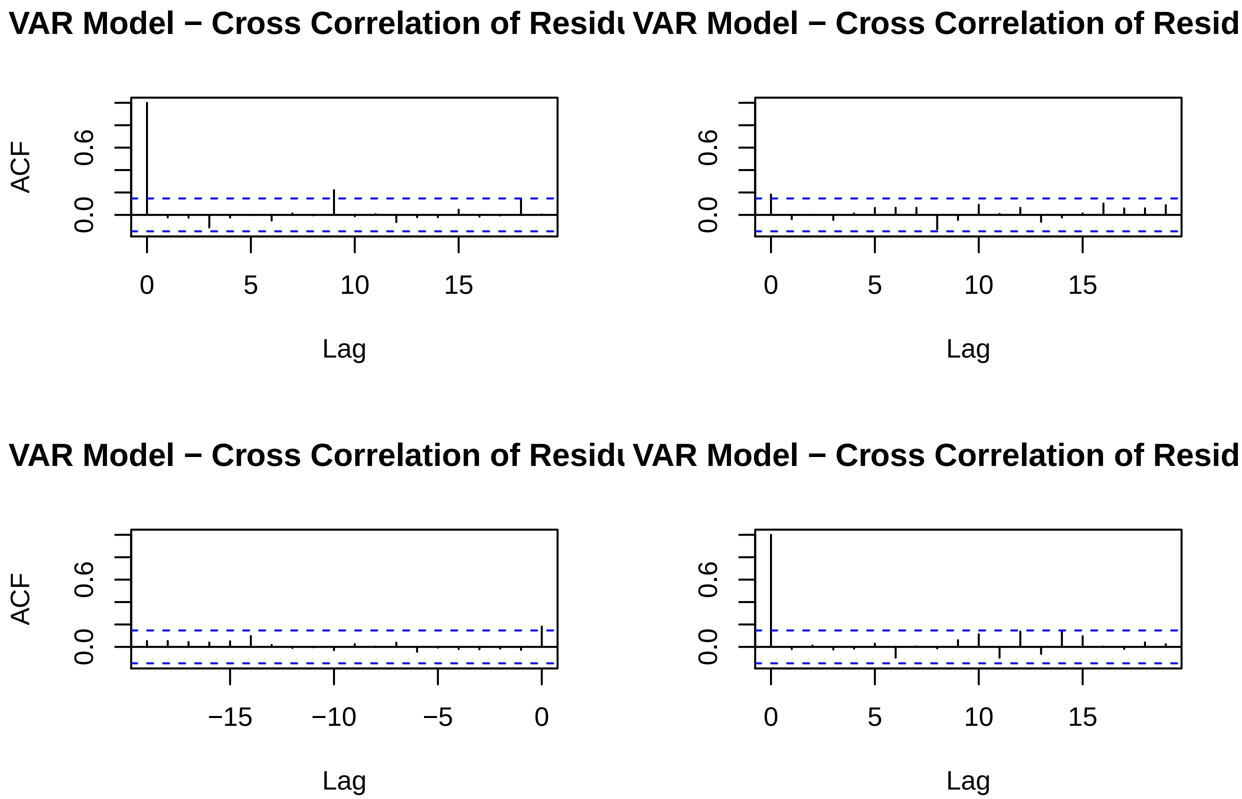
```
acf(residuals(var_model)[,1], main="VAR Model Autocorrelation of Residuals - CPI")
```




```
acf(residuals(var_model)[,2], main="VAR Model Autocorrelation of Residuals - Cost")
```




```
acf(residuals(var_model), main="VAR Model - Cross Correlation of Residuals")
```



Discuss your results.

The forecast of the Cost per Night and CPI show a tandem upward trend. The seasonality of Cost per Night is noticeable, as is its heteroskedasticity. For both variables, the confidence interval gets wider the further ahead in time we forecast.

As for the correlations in the residuals, we would hope to see what amounts to White Noise, both within variables (autocorrelation) and between variables (cross-correlation). The ACF plot of the residuals for CPI is a little ambiguous, but the spikes at 9 and 21 are more than likely coincidental. There are no such spikes in the ACF of the residuals for Cost per Night. So we can say that, within variables, the autocorrelation of residuals is White Noise.

For correlation of residuals between variables we need to look at the cross-correlation matrix of the residuals. This shows that the model does a pretty good job of capturing the complete association between the variables in time. We see the cross-correlation plots of each set of residuals in the upper right and lower left graphs. For a good model these should look like White Noise, and they do so here.