Scenario 3-4

Nipun

22/04/2021

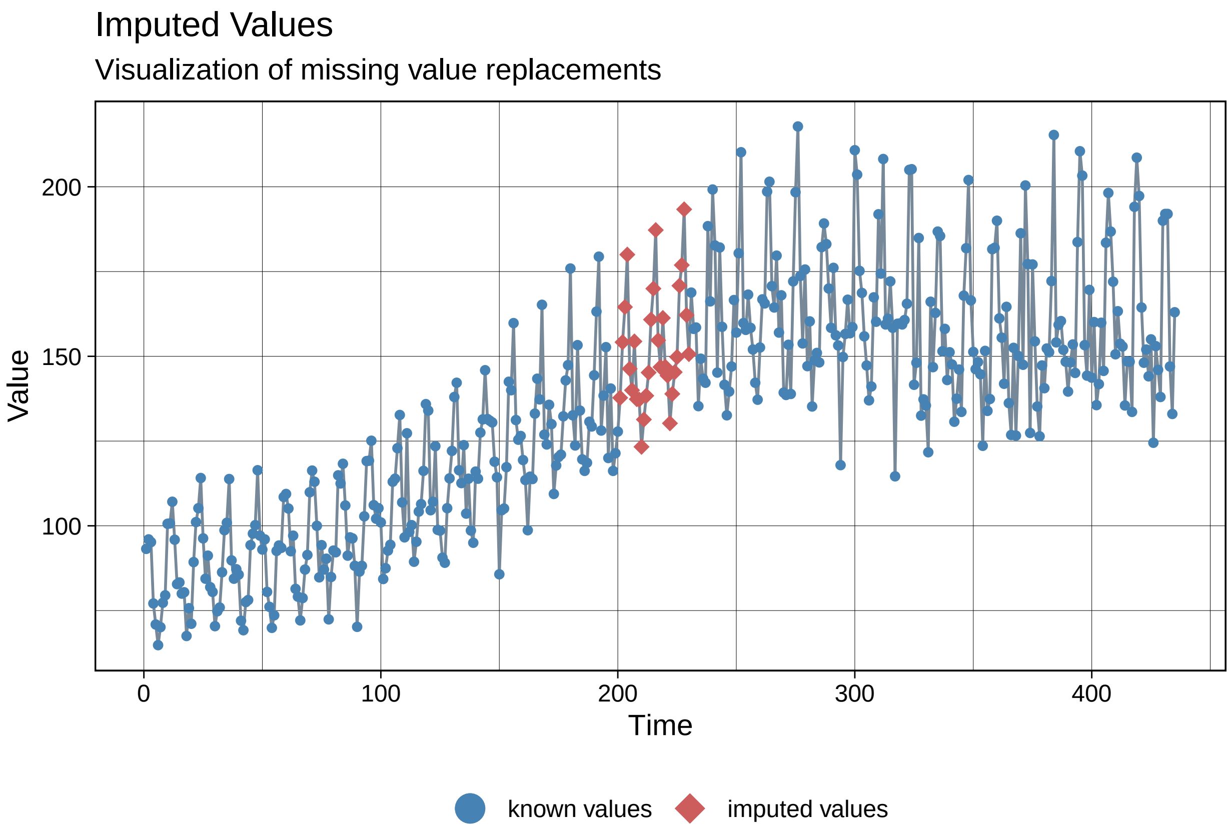
```
library(imputeTS)
library(vars)
library(fpp2)
```

```
prod_table <- read.table("prod_target.txt", header = TRUE, stringsAsFactors = FALSE, sep=",")
prod_1_table <- read.table("prod_1.txt", header = TRUE, stringsAsFactors = FALSE, sep=",")
prod_2_table <- read.table("prod_2.txt", header = TRUE, stringsAsFactors = FALSE, sep=",")
eng_1_table <- read.table("eng_1.txt", header = TRUE, stringsAsFactors = FALSE, sep=",")
eng_2_table <- read.table("eng_2.txt", header = TRUE, stringsAsFactors = FALSE, sep=",")
temp_table <- read.table("temp.txt", header = TRUE, stringsAsFactors = FALSE, sep=",")
beer <- ts(prod_table$V2, start=c(1956,1), frequency=12)
car <- ts(prod_1_table$V2, start=c(1961,7), frequency=12)
steel <- ts(prod_2_table$V2, start=c(1956,1), frequency=12)
gas <- ts(eng_1_table$V2, start=c(1956,1), frequency=12)
elec <- ts(eng_2_table$V2, start=c(1956,1), frequency=12)
temp <- ts(temp_table$X20.4, start=c(1943,11), frequency=12)
```

Scenario 3

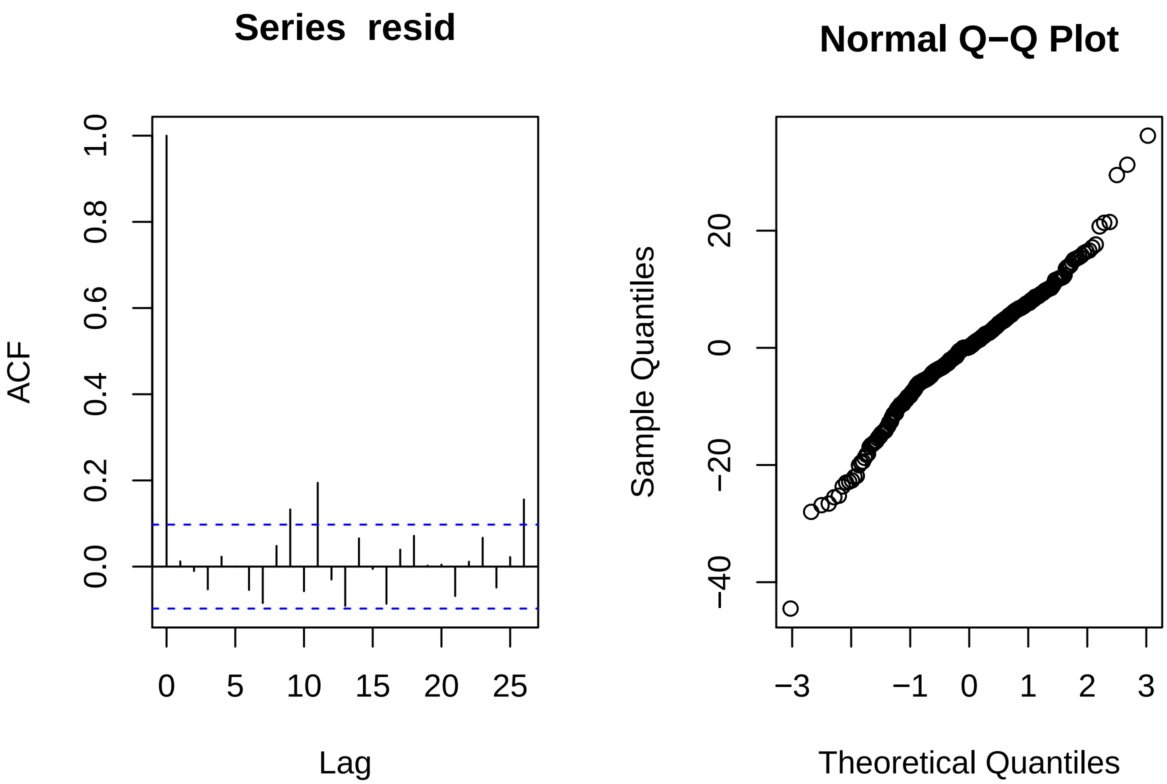
We will impute the 30 missing values in the middle of beer time series by using Kalman smoothing with model input from auto.arima

```
na_indices <- which(is.na(beer))
beerAR <- auto.arima(beer)
imp.cmort.kal = na_kalman(beer, model = beerAR$model)
ggplot_na_imputations(beer, imp.cmort.kal)
```



We will check if this model is fitting the beer time series well by looking at the ACF plot of residuals and QQ plot to make sure that these residuals are white

```
resid = na_remove(beerAR$residuals)
par(mfrow=c(1,2))
acf(resid)
qqnorm(resid)
```



From the ACF plot of residuals, we observe that there are large values at lag 8,11 and 24. But otherwise, the residuals seem uncorrelated to each other. Additionally, the residuals look gaussian after looking at their QQ-plot.

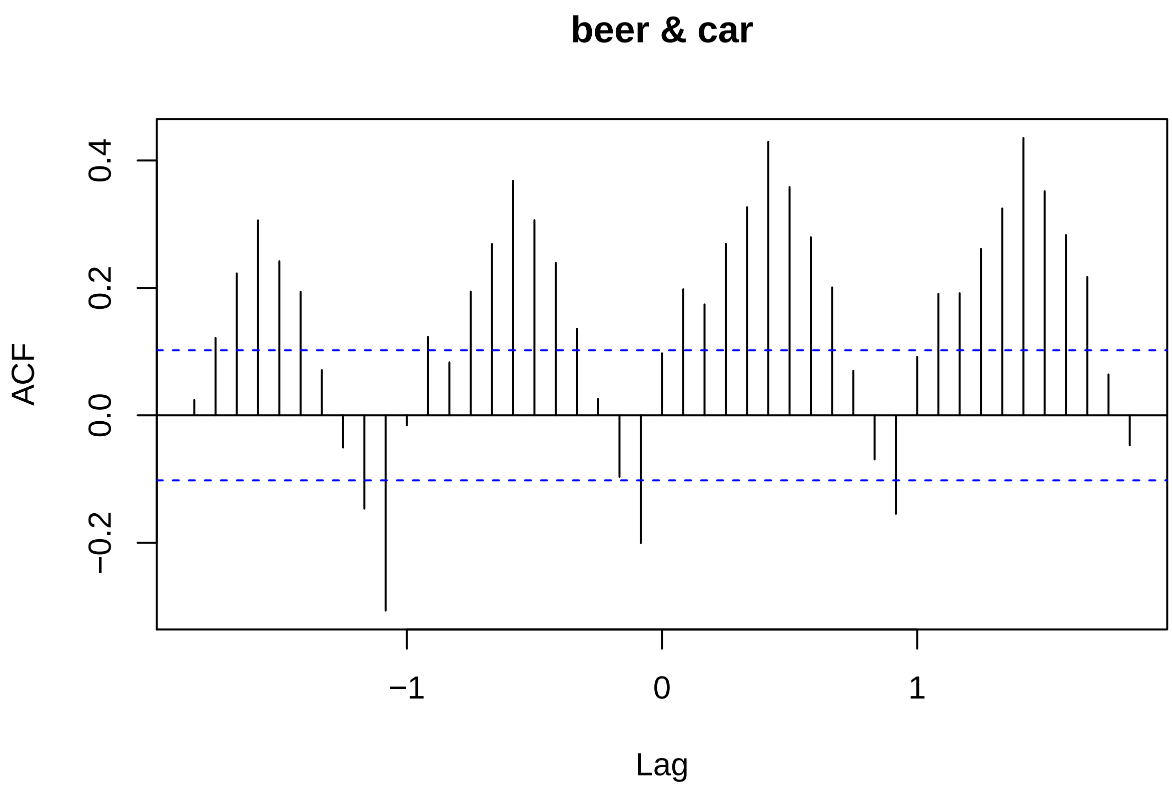
```
imputation3 = imp.cmort.kal[na_indices]
write(imputation3, file = paste("Scenario3_","Lamba","20751692",".txt", sep = ""), ncolumns = 1)
```

Scenario 4

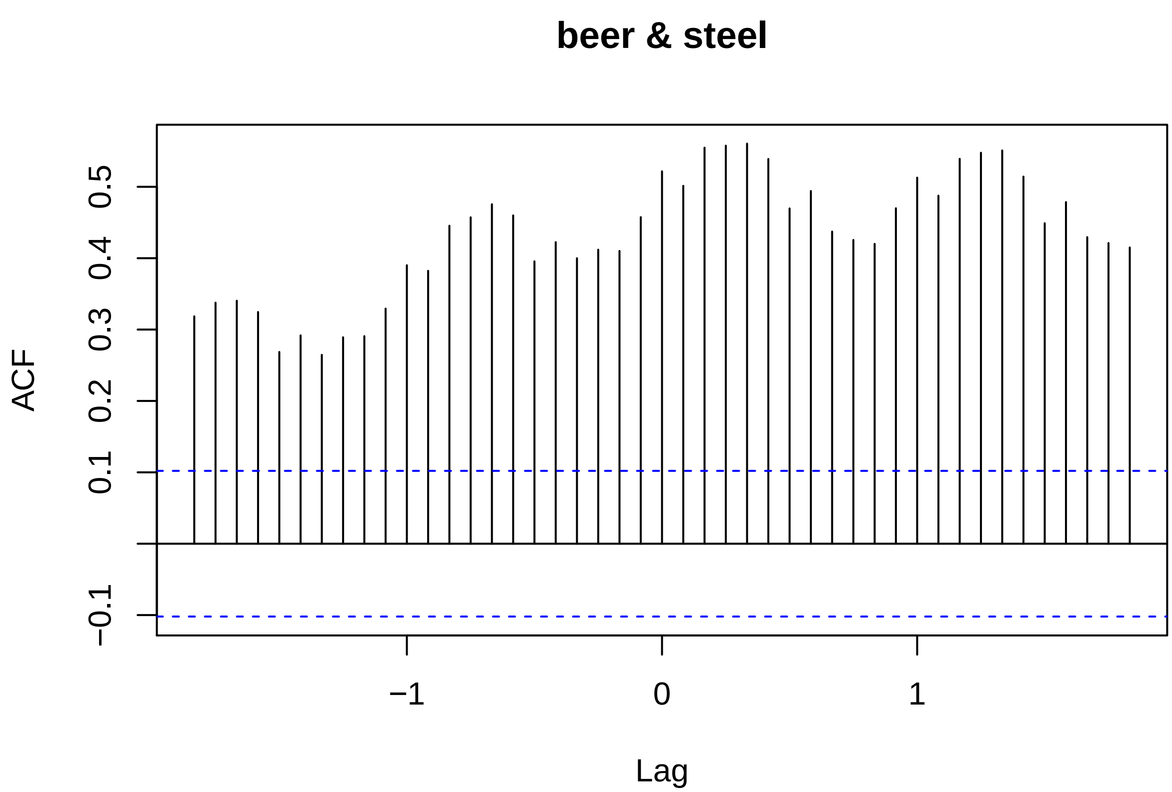
We will forecast the beer time series 24 steps ahead by using VAR model and regression with ARIMA errors. Then, we will compare their results using cross validation.

First, we will examine the cross-correlation plots to analyse the relationships of car, steel, gas, elec and temp time series on beer time series

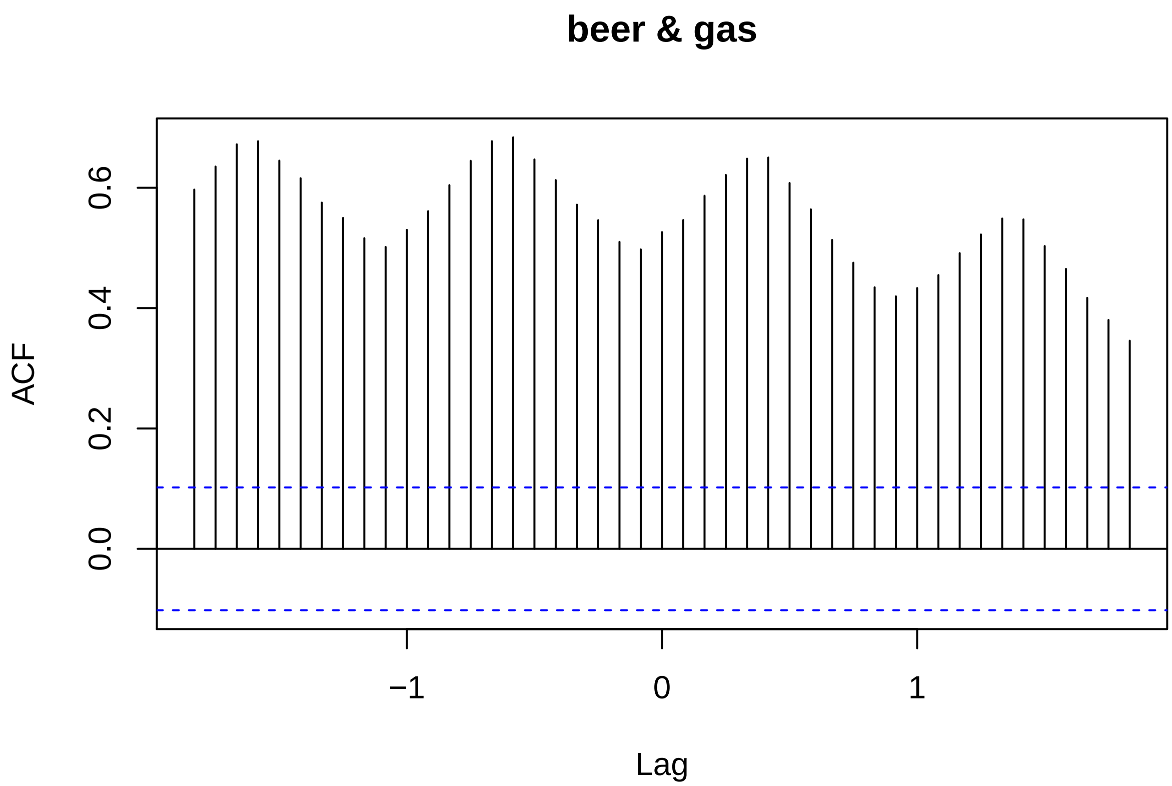
```
beer <- ts(imp.cmort.kal[c(67:435)],start=c(1961,7), frequency=12)
elec <- ts(elec[c(67:435)],start=c(1961,7), frequency=12)
gas <- ts(gas[c(67:435)],start=c(1961,7), frequency=12)
steel <- ts(steel[c(67:435)],start=c(1961,7), frequency=12)
temp <- ts(temp[c(213:581)],start=c(1961,7), frequency=12)

ccf(beer, car)
```



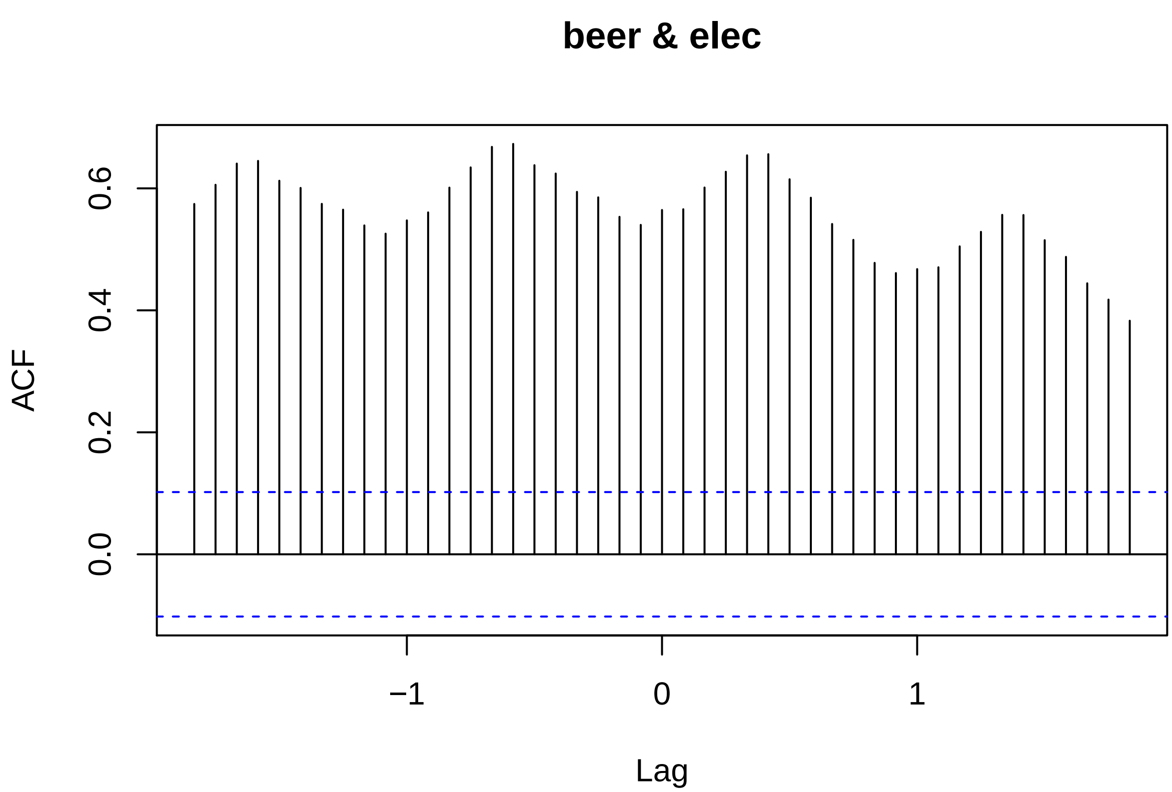

```
ccf(beer, steel)
```



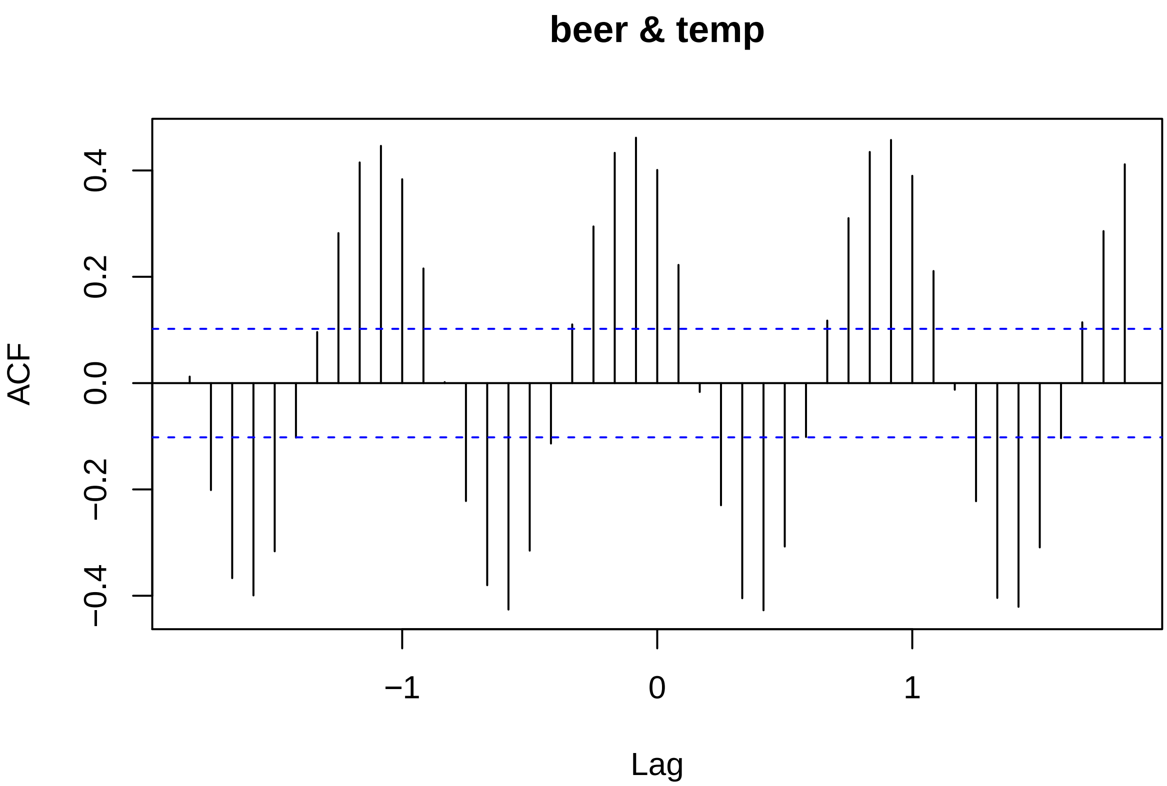
```
ccf(beer, gas)
```




```
ccf(beer, elec)
```



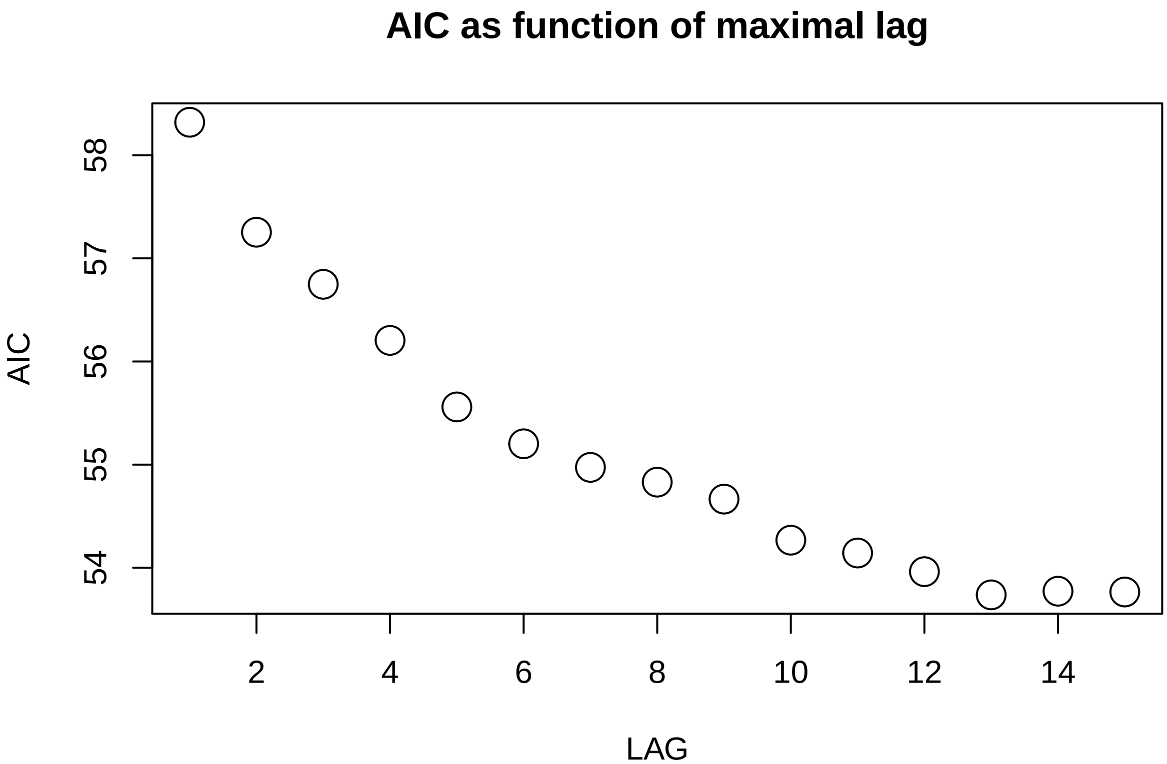
```
ccf(beer, temp)
```



From the above, cross correlations plot, we conclude that there is significant correlation between the t+h time of beer and t time of car, steel, gas, elec and temp. Hence, it makes sense to include all these time series in forecasting beer time series

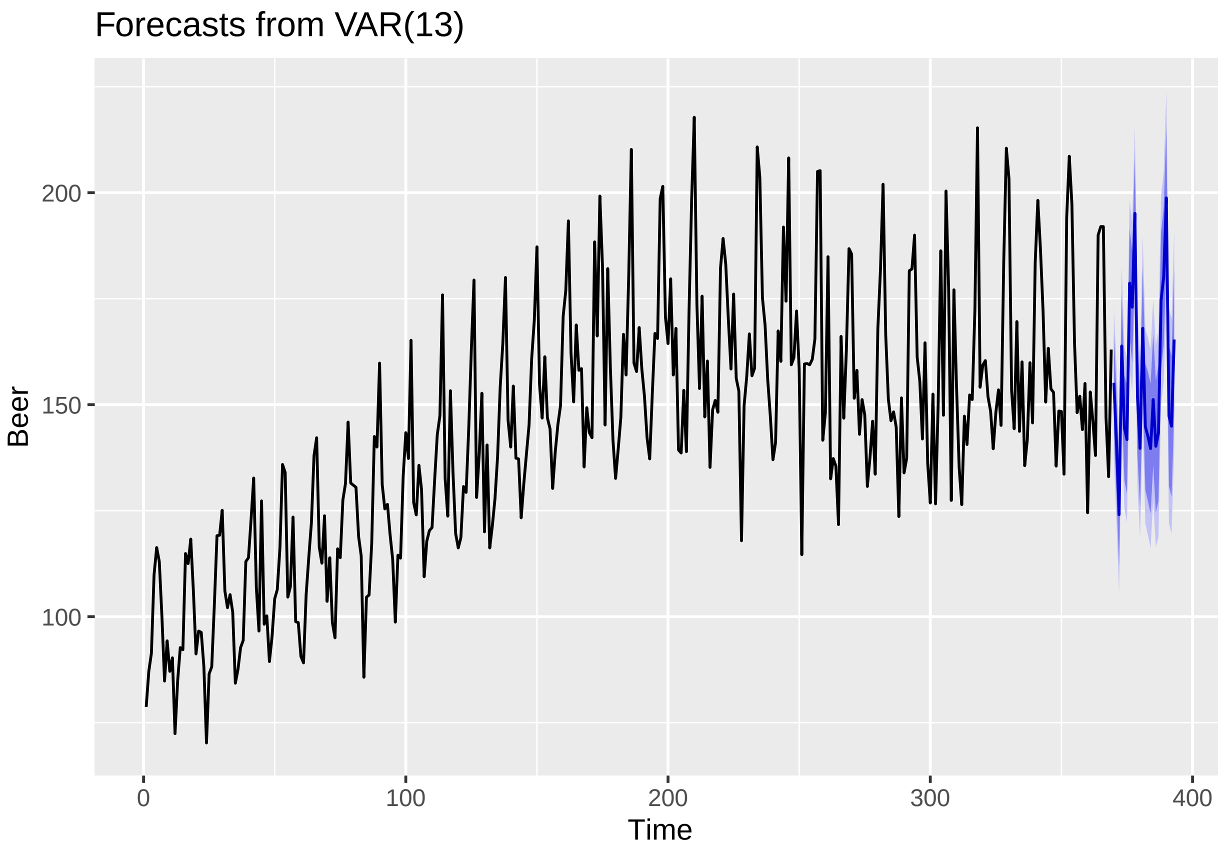
First, we use Vector Auto Regression to forecast the beer time series 24 steps ahead

```
dat.mat=cbind(as.numeric(car),as.numeric(steel),as.numeric(gas),as.numeric(elec), as.numeric(temp))
dat.mat.full =cbind(as.numeric(beer),as.numeric(car),as.numeric(steel),as.numeric(gas),as.numeric(elec)
colnames(dat.mat.full)=c("Beer","Car","Steel", "Gas", "Elec", "Temp")


x=VARselect(dat.mat.full, lag.max = 15)$criteria[1,] #computes AICs for VAR models up to lag 20
plot(x, main="AIC as function of maximal lag", xlab= "LAG", ylab="AIC",cex=2)
```



We are calculating the AIC of VAR fits to this mutlivariate time series from lag=1 and lag=15 and see the lag at which the AIC value is the lowest. We observe that for lag=13, the AIC is the lowest. Hence, we will use lag=13 as the lag in VAR model

```
var_mort <- VAR(as.ts(dat.mat.full), p = 13, type = "const")
var_mort_prd <- forecast(var_mort, h=24)
autoplot(var_mort_prd$forecast$Beer)
```




```
serial.test(var_mort,lags.pt = 13)
```

```
##
##  Portmanteau Test (asymptotic)
##
## data:  Residuals of VAR object var_mort
## Chi-squared = 207.11, df = 0, p-value < 2.2e-16
```

From the Pormanteau test, we observe that p-value is very small which shows significant evidence supporting that residuals are serially correlated

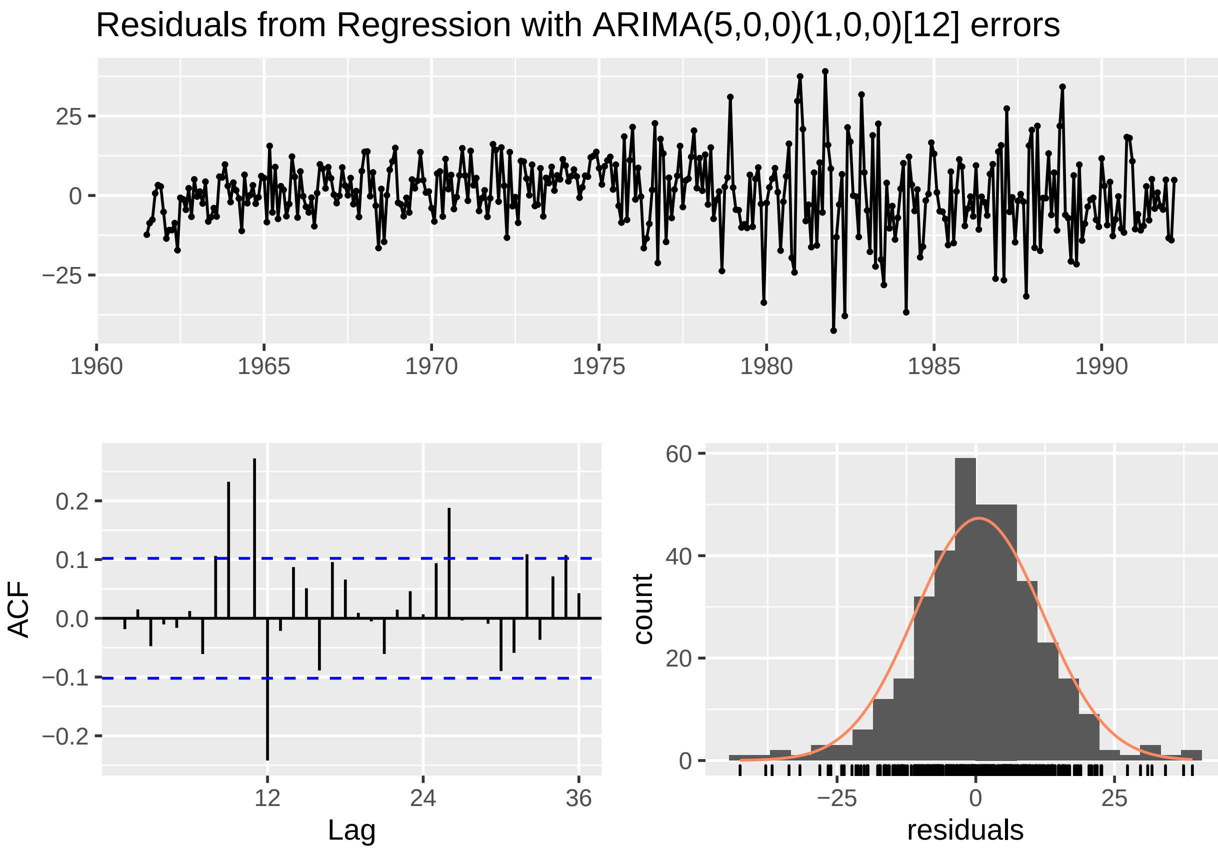
Now we fit beer time series by performing regression with ARIMA errors. First, we will forecast covariate series car, steel, gas, elec and temp 24 steps ahead. Then, we will use these forecasts to forecast beer time series 24 steps ahead

```
car.mod = auto.arima(car)
steel.mod = auto.arima(steel)
gas.mod = auto.arima(gas)
elec.mod = auto.arima(elec)
temp.mod = auto.arima(temp)

car.for=forecast(car.mod, h=24)
steel.for=forecast(steel.mod, h=24)
gas.for=forecast(gas.mod, h=24)
elec.for=forecast(elec.mod, h=24)
temp.for=forecast(temp.mod, h=24)
```

```r
car.for=ts(car.for$mean,frequency = 12)
steel.for=ts(steel.for$mean,frequency = 12)
gas.for=ts(gas.for$mean,frequency = 12)
elec.for=ts(elec.for$mean,frequency = 12)
temp.for=ts(temp.for$mean,frequency = 12)

dat.mat.for=cbind(car.for,steel.for,gas.for,elec.for,temp.for)

ar.regf=auto.arima(beer, xreg=dat.mat)

ar.regf
```

```
## Series: beer
## Regression with ARIMA(5,0,0)(1,0,0)[12] errors
##
## Coefficients:
##           ar1      ar2     ar3      ar4     ar5    sar1  intercept  xreg1
##       -0.0254  -0.0367  0.1398  -0.0564  0.1475  0.8575    63.2196  3e-04
## s.e.   0.0573   0.0545  0.0519   0.0522  0.0526  0.0323    13.9183  1e-04
##        xreg2  xreg3   xreg4   xreg5
##       0.0118  2e-04  0.0037  1.3517
## s.e.  0.0102  3e-04  0.0016  0.4590
##
## sigma^2 estimated as 137.3:  log likelihood=-1433.7
## AIC=2893.4   AICc=2894.43   BIC=2944.24
```

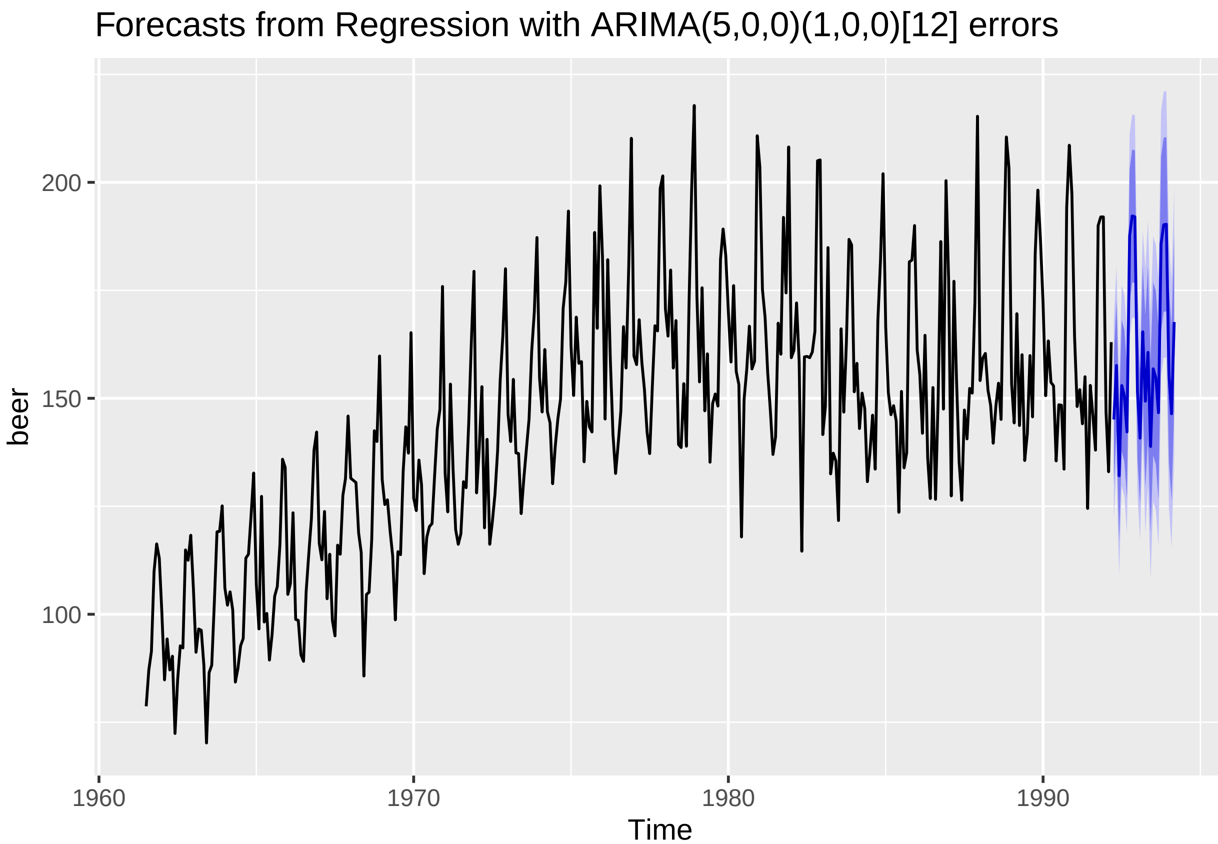
```r
checkresiduals(ar.regf)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(5,0,0)(1,0,0)[12] errors
## Q* = 93.139, df = 12, p-value = 1.21e-14
##
## Model df: 12.   Total lags used: 24
```

By looking at the histogram of residuals plot, we observe that they look gaussian. The ACF plot has few large values but other than that, the residuals seem to be somewhat uncorrelated.

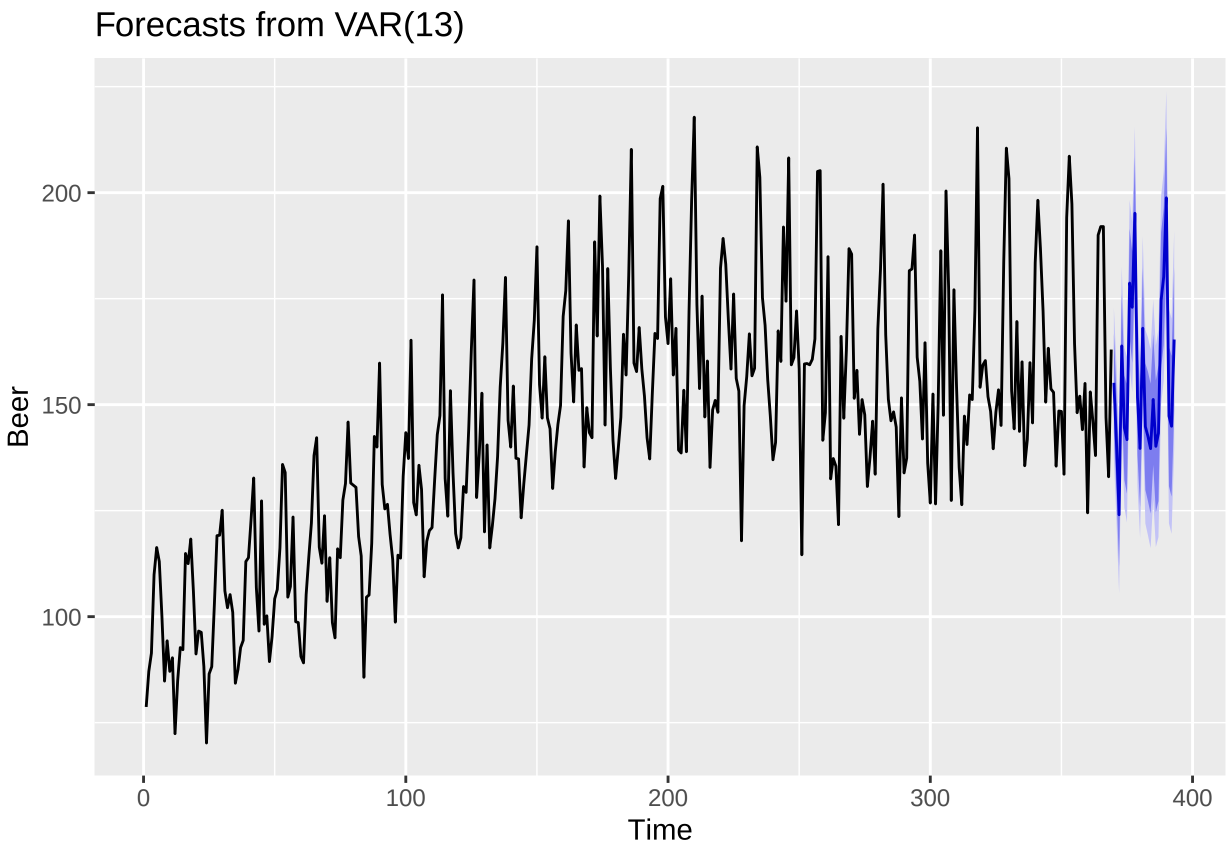
```
x=forecast(ar.regf,xreg=dat.mat.for,h=24)
autoplot(x)
```



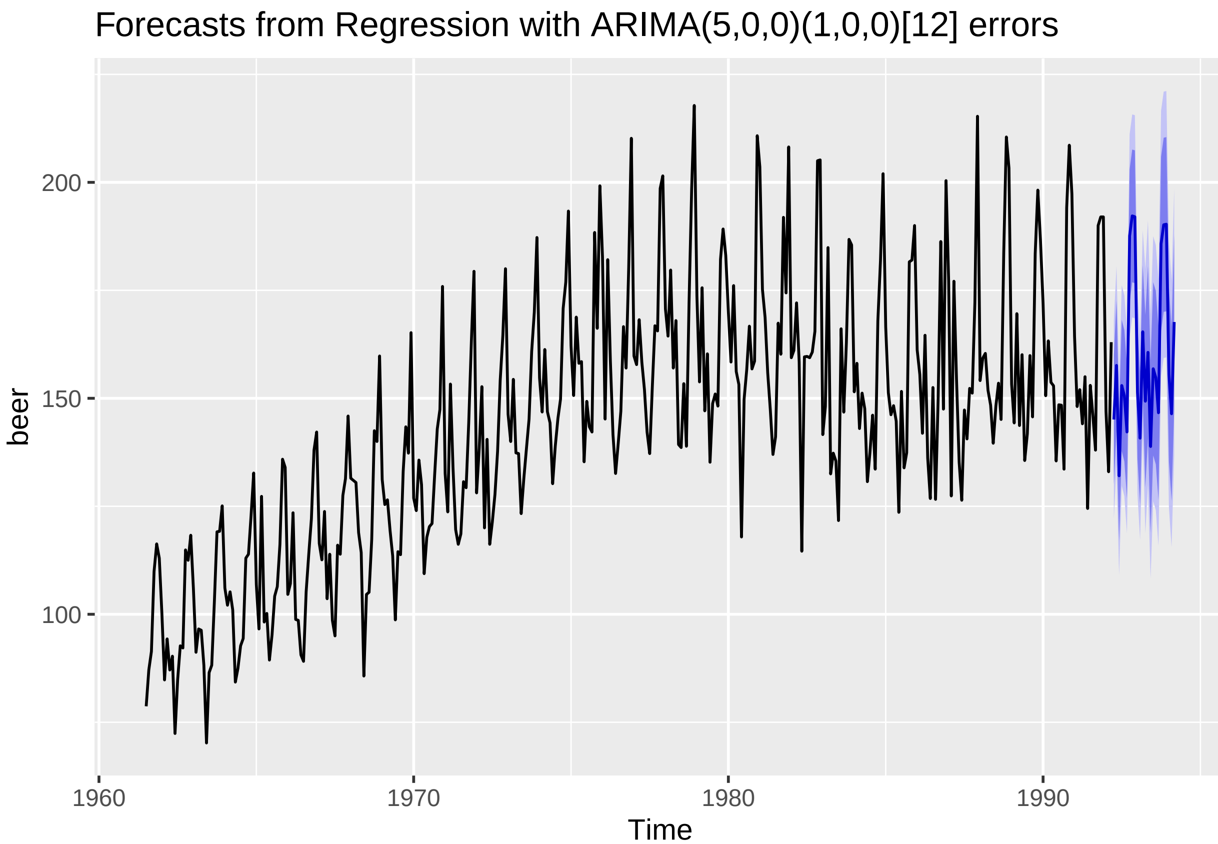
We have the 24 steps ahead forecast for beer time series based on the regression with ARIMA errors model. The forecast looks reasonably plausible given the past observations. The 95% prediction level is wide enough to capture the variations in the forecasts.

Now we will compare the 24 steps ahead forecasts of VAR model and regression with ARIMA errors model side by side

```
autoplot(var_mort_prd$forecast$Beer)
```




```
autoplot(x)
```



After comparing the 24 steps ahead forecasts of both VAR and regression with ARIMA errors model, we observe that regression with ARIMA errors forecasts fit the past observations better than VAR model. Hence, we will be choosing regression with ARIMA errors model as our model for beer time series.

```
forecast4 = x$mean
write(forecast4, file = paste("Scenario4_","Lamba","20751692",".txt", sep = ""), ncolumns = 1 )
```