Time Series Analysis Homework07 Explaination

Hung-Ta, Wang (R13546017)

December 08, 2024

5. Consider the well-known time series data "co2" (monthly carbon dioxide readings through 11 years in Alert, Canada).

5.a. Fit a deterministic regression model in terms of months and time. Are the regression coefficients significant? What is the adjusted R-squared? (Note that the month variable should be treated as categorical and transformed into 11 dummy variables.)

```r
library(forecast)
library(TSA)
library(tseries)
library(tidyverse)

data <- read.csv("C:/Git_Code/Some-practice/TSA HW07.co2.csv")
data$month <- factor(data$month, levels = unique(data$month))
model <- lm(co2_level ~ time_trend + month, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = co2_level ~ time_trend + month, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.73874 -0.59689 -0.06947  0.54086  2.15539
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3290.5412    44.1790 -74.482  < 2e-16 ***
## time_trend      1.8321     0.0221  82.899  < 2e-16 ***
## monthFeb        0.6682     0.3424   1.952 0.053319 .
## monthMar        0.9637     0.3424   2.815 0.005715 **
## monthApr        1.2311     0.3424   3.595 0.000473 ***
## monthMay        1.5275     0.3424   4.460 1.87e-05 ***
## monthJun       -0.6761     0.3425  -1.974 0.050696 .
## monthJul       -7.2851     0.3426 -21.267  < 2e-16 ***
## monthAug      -13.4414     0.3426 -39.232  < 2e-16 ***
## monthSep      -12.8205     0.3427 -37.411  < 2e-16 ***
## monthOct       -8.2604     0.3428 -24.099  < 2e-16 ***
## monthNov       -3.9277     0.3429 -11.455  < 2e-16 ***
## monthDec       -1.3367     0.3430  -3.897 0.000161 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8029 on 119 degrees of freedom
## Multiple R-squared:  0.9902, Adjusted R-squared:  0.9892
## F-statistic: 997.7 on 12 and 119 DF,  p-value: < 2.2e-16
```

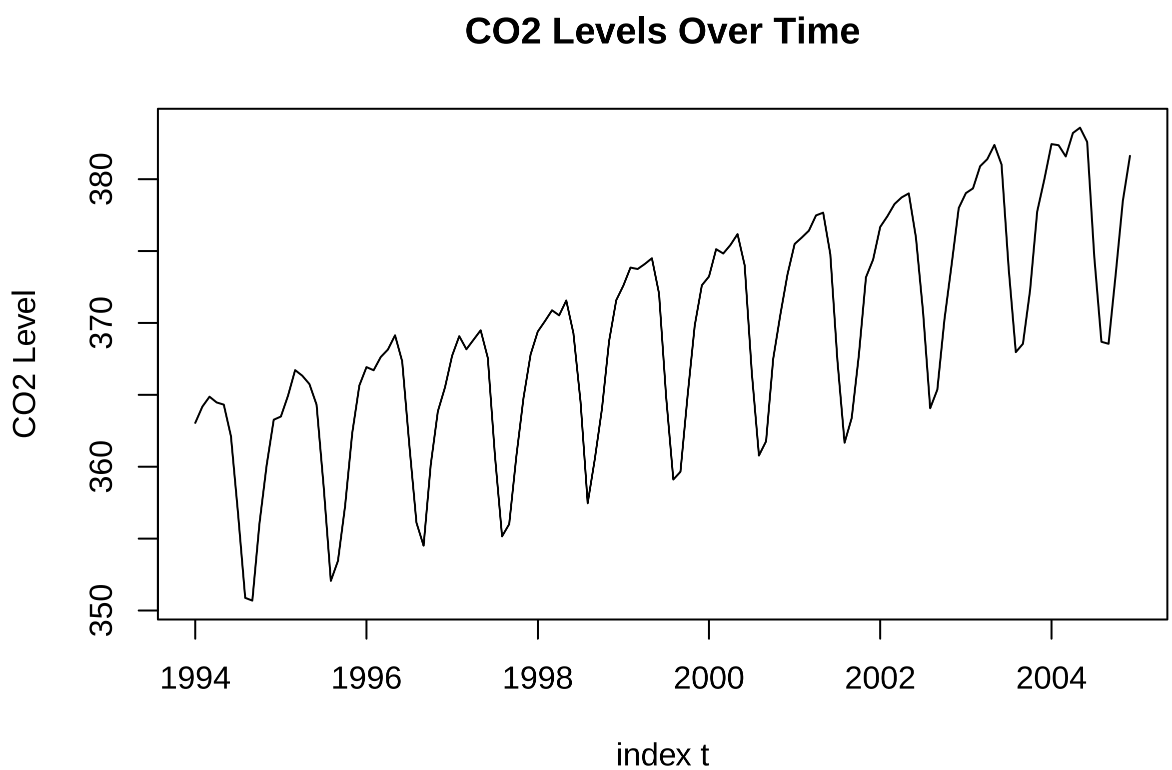
```
cat("Adjusted R-squared:", summary(model)$adj.r.squared)
```

```
## Adjusted R-squared: 0.9891657
```

Explanation: All regression coefficients are significant except monthFeb and monthJun. The adjusted R-squared is 0.9891657.

5.b. Identify, estimate the SARIMA model for the co2 level.

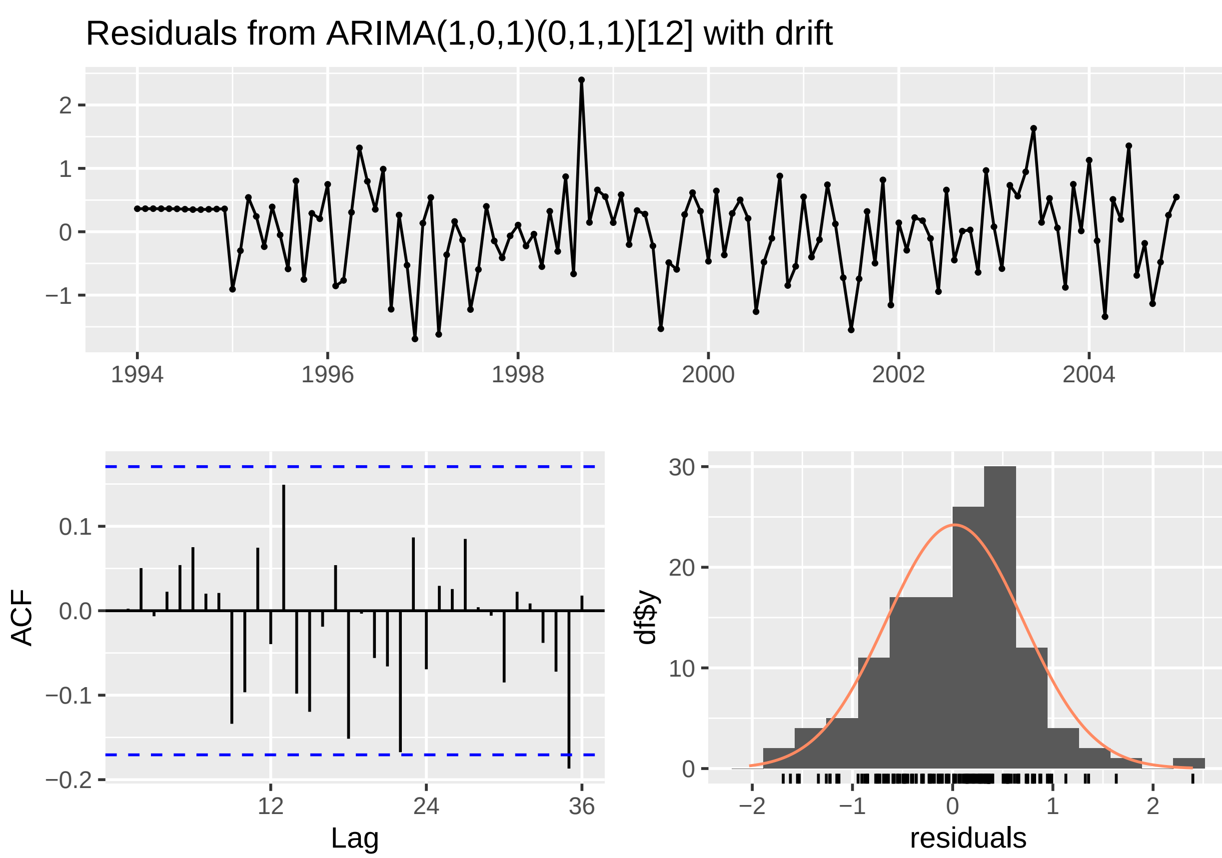
```
co2_ts <- ts(data$co2_level, start = c(1994, 1), frequency = 12)
plot(co2_ts, main = "CO2 Levels Over Time", ylab = "CO2 Level", xlab = "index t")
```




```
sarima_model <- auto.arima(co2_ts, seasonal = TRUE, stepwise = FALSE, approximation = FALSE)
summary(sarima_model)
```

```
## Series: co2_ts
## ARIMA(1,0,1)(0,1,1)[12] with drift
##
## Coefficients:
##          ar1      ma1     sma1   drift
##       0.8349  -0.4630  -0.8487  0.1520
## s.e.  0.0819   0.1246   0.1274  0.0052
##
## sigma^2 = 0.5288:  log likelihood = -136.09
## AIC=282.18   AICc=282.7   BIC=296.11
##
## Training set error measures:
##                      ME      RMSE      MAE         MPE      MAPE     MASE
## Training set 0.02075345 0.6817172 0.542085 0.005102291 0.1469965 0.288126
##                     ACF1
## Training set 0.002372469
```

```r
checkresiduals(sarima_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,1)(0,1,1)[12] with drift
## Q* = 25.4, df = 21, p-value = 0.2302
##
## Model df: 3.   Total lags used: 24
```

Explanation: We can see that SARIMA(1,0,1)(0,1,1)[12] can fit the data and Ljung Box test also shows that the residual is stationary process.

5.c. Compare the two models above, what do you observe?

```
calculate_mape <- function(actual, predicted) {
  return(mean(abs((actual - predicted) / actual)) * 100)
}

deterministic_predictions <- as.numeric(predict(model, newdata = data))
sarima_predictions <- as.numeric(fitted(sarima_model))

mape_deterministic <- calculate_mape(data$co2_level, deterministic_predictions)
mape_sarima <- calculate_mape(data$co2_level, sarima_predictions)

cat("MAPE for Deterministic Regression:", mape_deterministic, "%\n")
```

```
## MAPE for Deterministic Regression: 0.1693098 %
```

```
cat("MAPE for SARIMA Model:", mape_sarima, "%\n")
```

```
## MAPE for SARIMA Model: 0.1469965 %
```

Explanation: We can see that the MAPE for Deterministic Regression is 0.1693098 % and MAPE for SARIMA Model is 0.1469965 %. Thus, we can conclude that SARIMA is slightly better than Deterministic Regression. Both models has a great performance in the dataset.