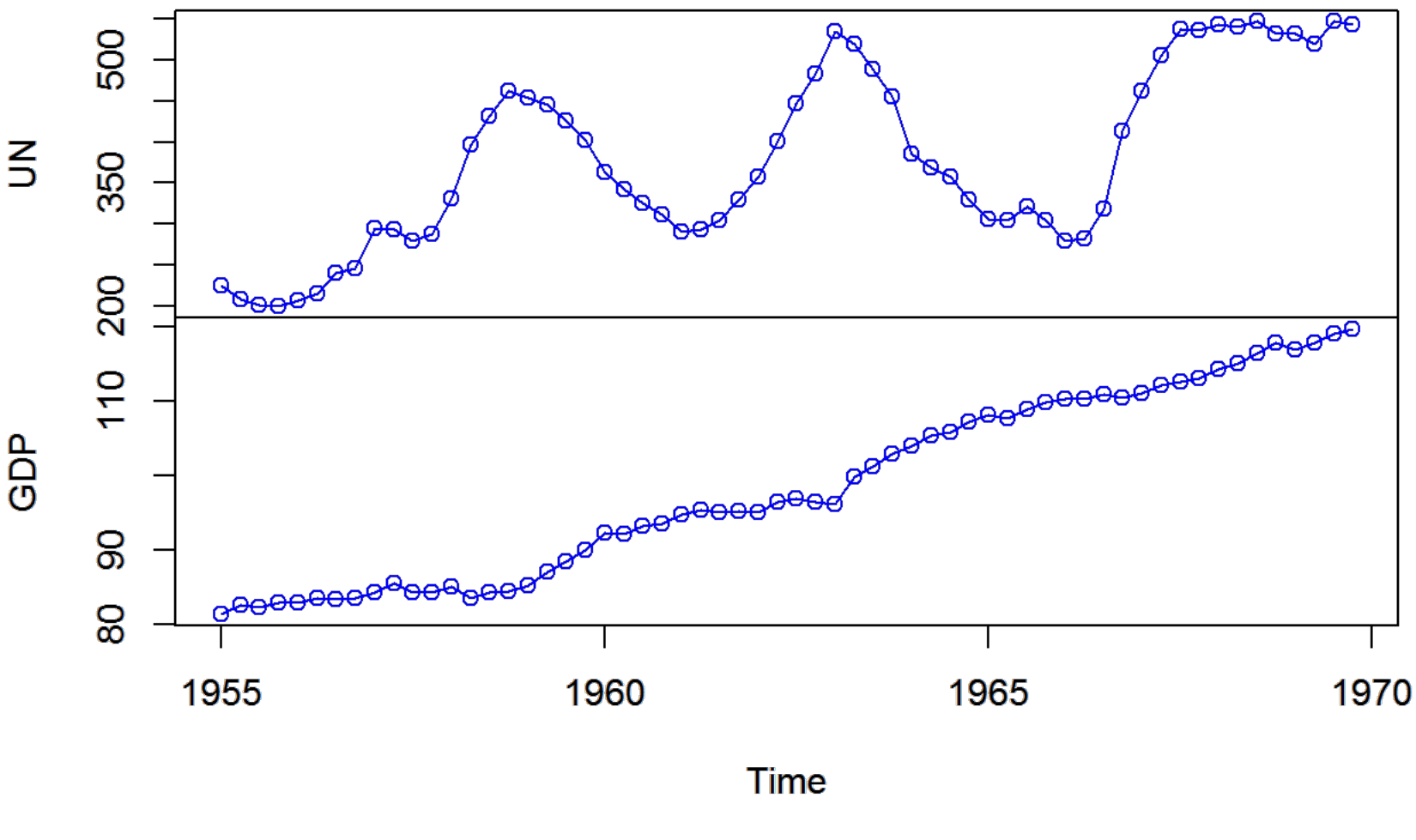
Project 2 Solution

```
data <- ts(read.csv("unem_gdp.csv"), start=1955, frequency=4)
un <- data[,1]
gdp <- data[,2]
print(window(data, end=c(1956,2)))
```

```
##           UN   GDP
## 1955 Q1 225 81.37
## 1955 Q2 208 82.60
## 1955 Q3 201 82.30
## 1955 Q4 199 83.00
## 1956 Q1 207 82.87
## 1956 Q2 215 83.60
```

Plot of data

```
plot(data, type="o", col="blue", main="Time Series of UN and GDP")
```



Both the UN and GDP series have an upward trend and are therefore not stationary. We would not expect the ARMA models without differencing to be adequate.

Building ARIMA models

```
library(forecast)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## Loading required package: timeDate
## This is forecast 5.6
```

```
pre.un <- window(un, end=c(1968,4))
un.arima <- auto.arima(pre.un); un.arima
```

```
## Series: pre.un
## ARIMA(1,1,0) with drift
##
## Coefficients:
##          ar1   drift
##       0.6574  4.1744
## s.e.  0.1005  8.7225
##
## sigma^2 estimated as 523.1:  log likelihood=-245.42
## AIC=496.84   AICc=497.31   BIC=502.86
```

```
pre.gdp <- window(gdp, end=c(1968,4))
gdp.arima <- auto.arima(pre.gdp); gdp.arima
```

```
## Series: pre.gdp
## ARIMA(0,1,0) with drift
##
## Coefficients:
##        drift
##       0.6624
## s.e.  0.1162
##
## sigma^2 estimated as 0.7429:  log likelihood=-68.1
## AIC=140.21   AICc=140.44   BIC=144.22
```

```
c(ndiffs(pre.un), ndiffs(pre.gdp))
```

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

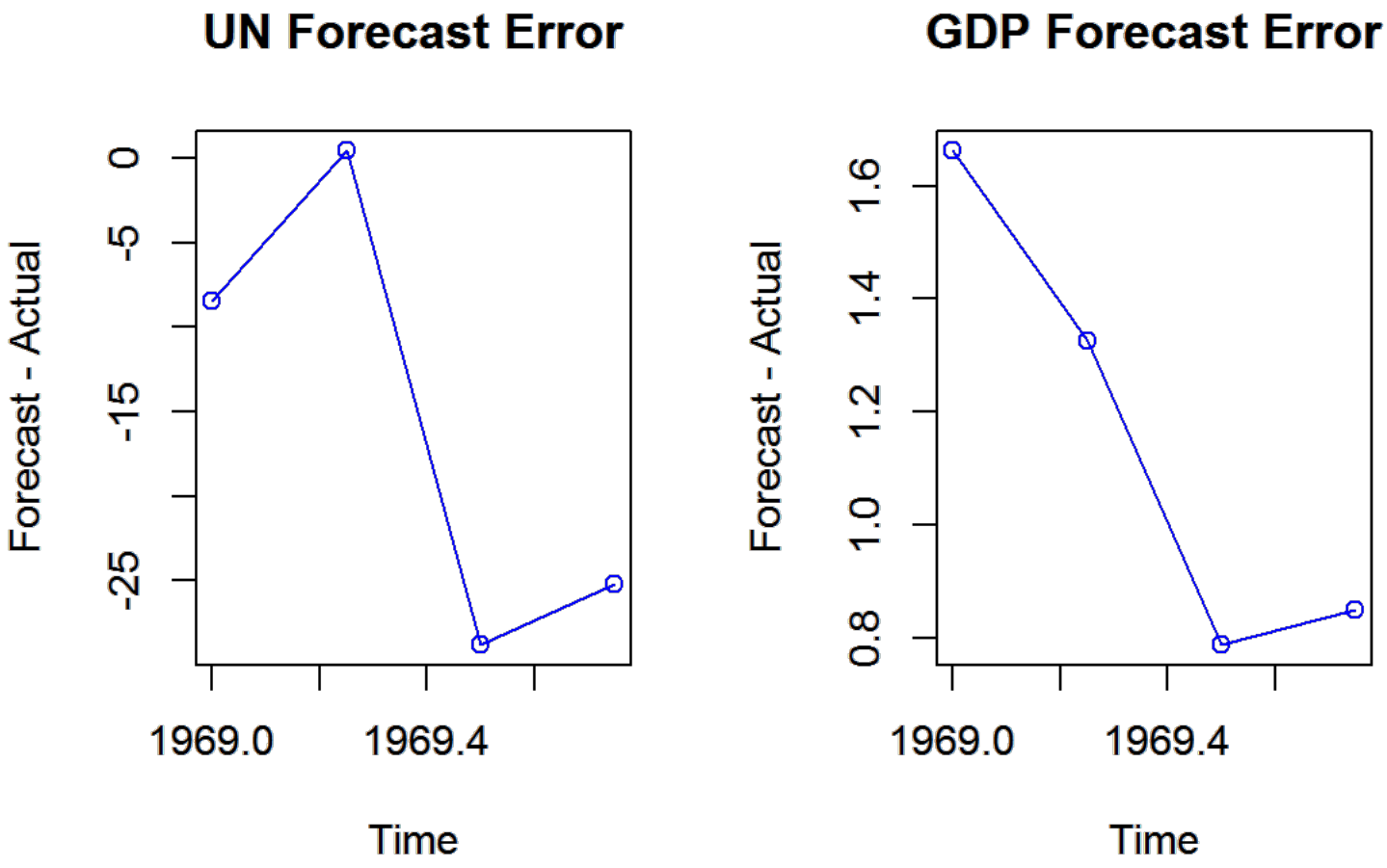
For both the UN and GDP series, both the auto selection algorithm and the KPSS test suggest ARIMA models with $d = 1$.

Forecasting with ARIMA models

```
un.forecast <- forecast(un.arima, h=4)
un.error <- un.forecast$mean - window(un, start=c(1969,1))

gdp.forecast <- forecast(gdp.arima, h=4)
gdp.error <- gdp.forecast$mean - window(gdp, start=c(1969,1))

par(mfrow=c(1,2))
plot(un.error, type="o", col="blue", main="UN Forecast Error", ylab="Forecast - Actual")
plot(gdp.error, type="o", col="blue", main="GDP Forecast Error", ylab="Forecast - Actual")
```




```
c(sum(un.error^2), sum(gdp.error^2))
```

```
## [1] 1537.62873    5.85944
```

The sums of squared errors are 1537.6 and 5.86 for the UN and GDP series respectively.

Building linear regression models

```
gdp.on.un <- lm(pre.gdp ~ pre.un)
summary(gdp.on.un)
```

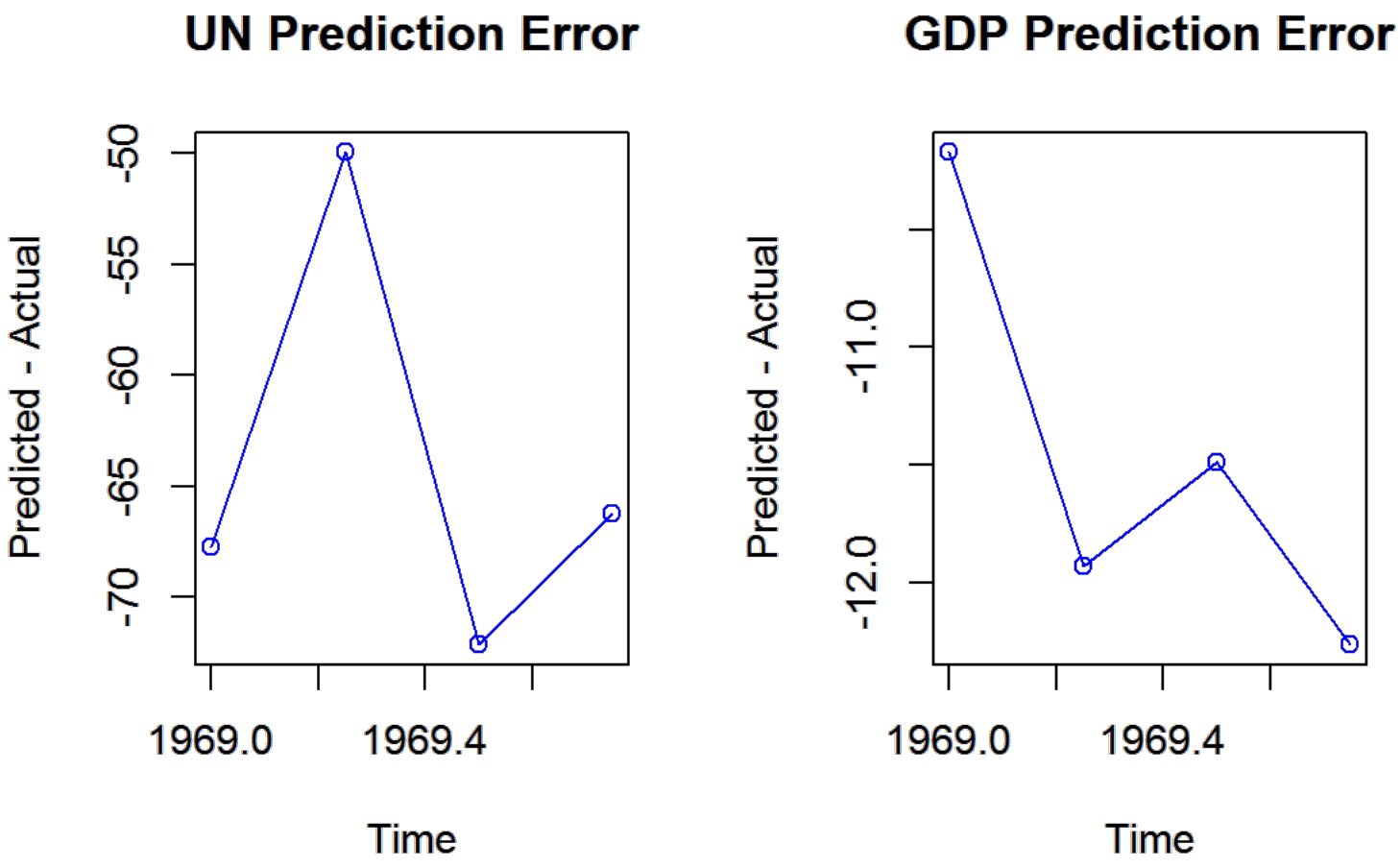
```
##
## Call:
## lm(formula = pre.gdp ~ pre.un)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.024  -7.146  -1.932   8.024  18.411
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 75.42255    4.87117  15.483  < 2e-16 ***
## pre.un       0.05866    0.01271   4.616 2.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.84 on 54 degrees of freedom
## Multiple R-squared:  0.283,  Adjusted R-squared:  0.2697
## F-statistic: 21.31 on 1 and 54 DF,  p-value: 2.453e-05
```

```
un.on.gdp <- lm(pre.un ~ pre.gdp)
summary(un.on.gdp)
```

```
##
## Call:
## lm(formula = pre.un ~ pre.gdp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -153.40  -67.99  -17.84   84.76  170.32
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -99.147    102.132  -0.971    0.336
## pre.gdp        4.824      1.045   4.616 2.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.23 on 54 degrees of freedom
## Multiple R-squared:  0.283,  Adjusted R-squared:  0.2697
## F-statistic: 21.31 on 1 and 54 DF,  p-value: 2.453e-05
```

```
gdp.error <- predict(gdp.on.un, data.frame(pre.un=window(un, start=c(1969,1)))) - window(gd
p, start=c(1969,1))
un.error <- predict(un.on.gdp, data.frame(pre.gdp=window(gdp, start=c(1969,1)))) - window(u
n, start=c(1969,1))

par(mfrow=c(1,2))
plot(un.error, type="o", col="blue", main="UN Prediction Error", ylab="Predicted - Actual")
plot(gdp.error, type="o", col="blue", main="GDP Prediction Error", ylab="Predicted - Actual
")
```




```
c(sum(un.error^2), sum(gdp.error^2))
```

```
## [1] 16681.0860   528.2449
```

In our case, UN should be the independent variable and GDP should be the dependent variable because unemployment affects the GDP but not the other way around.