Introduction

Overview

This report is concerned with performing exploratory analysis and creating the best possible model for a dataset describing returns on a share market trader’s investment portfolio. The data provided contains 179 days out of a possible 252 trading days in a year and only has consecutive trading days (i.e. weekends are not included).

Objectives

- Obtain summaries of the data using the relevant functions and plotting the time series data, ACF, PACF, and histograms.
- Fit the created models and obtain summaries of the models using residuals and QQ plots.
- Using the optimal model, provide predictions for the next 5 days.

Method

Determining Frequency

In the first step, the dataset is loaded into the environment and summary obtained. The summaries show a mean return of 5704.30 AUD, a max of 21461.10 AUD and a min of -4916.70 AUD.

```
setwd("C:/Work/MC242/Sem3/Time Series/TS Assignment 1")

returns = read.csv("assignment1Data2024.csv", col.names = c("Day", "Returns"))
summary(returns)
```

```
##       Day           Returns
##  Min.   :  1.0   Min.   :-49.167
##  1st Qu.: 45.5   1st Qu.: -2.685
##  Median : 90.0   Median : 51.105
##  Mean   : 90.0   Mean   : 57.043
##  3rd Qu.:134.5   3rd Qu.:117.781
##  Max.   :179.0   Max.   :214.611
```

The returns column is converted to a time-series object with a frequency of 1 since it is daily data. The series is then plotted to observe any trends or seasonality. The plot shows a decreasing trend till about day 100, after which the trend steadily increases. Throughout the series, there seem to be seasonal components which increase in variance as the days progress. This series can be said to be non-stationary for the reasons discussed (Canvas, 2024).

The ACF plot return significant peaks over the confidence intervals with a seasonal wave pattern to the peaks. The repeating wave pattern is observed to have a wavelength of 13. Therefore, it will be assumed for the rest of this report that the series has a seasonal component with a period of 13 days. The series is re-converted into a time-series object with a frequency of 13.

```
returns.ts = ts(returns[,c(2)], frequency = 1)
plot(returns.ts, ylab='Returns in AUD100', xlab='Day', type='o', main = "Time series plot of daily returns in AUD100.")
```

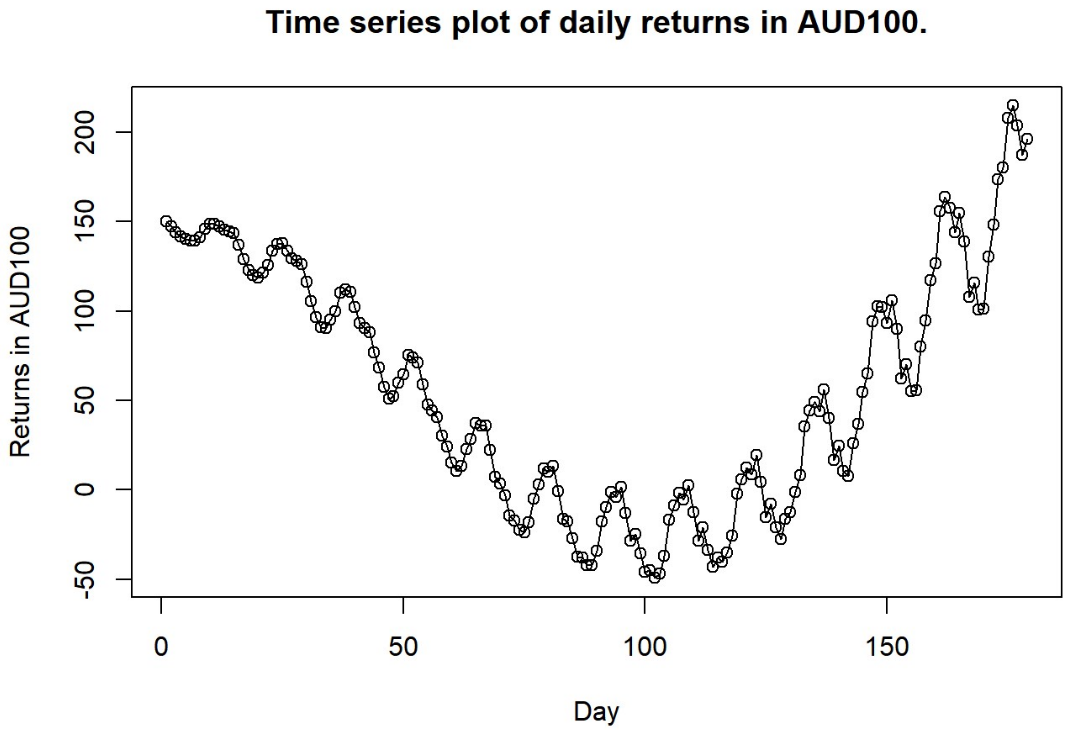

Figure 1: Plot of the time series returns data in AUD100

```
# changing frequency based on ACF plot
returns.ts= ts(returns[,c(2)], frequency = 13)
```

```
acf(returns.ts, lag.max = 60)
```

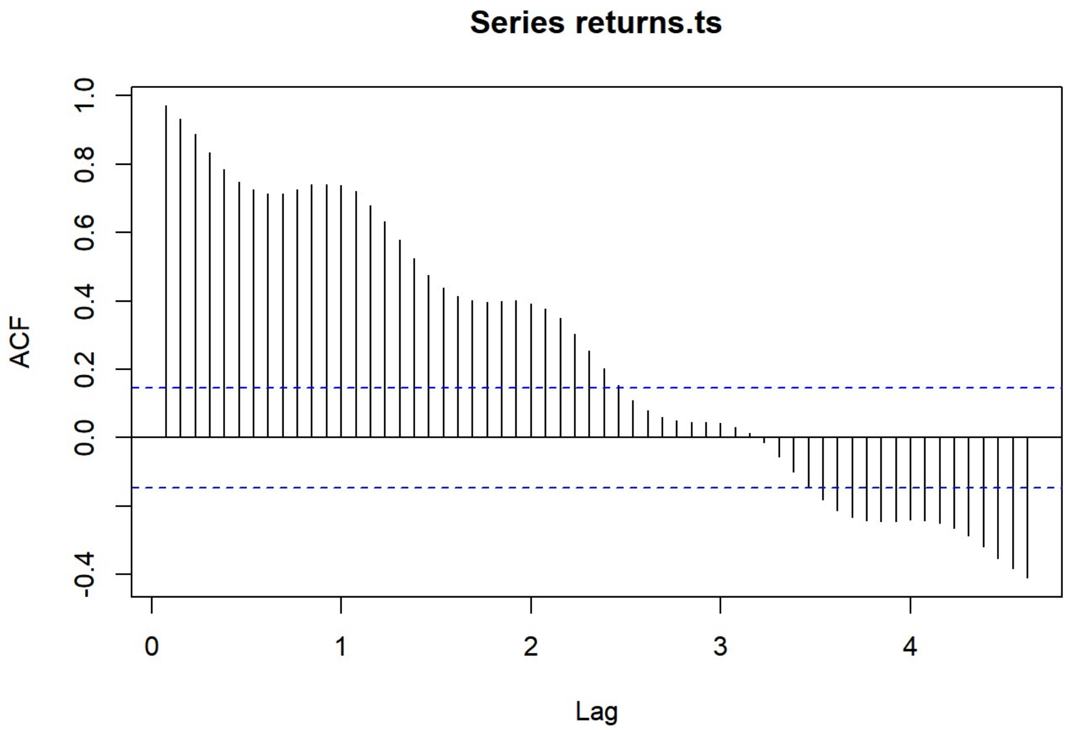

Figure 2: ACF plot of the time series returns data

Linear Model

A linear model is fitted to the data and the results are plotted. The resulting model (in red) is a slightly decreasing line that follows the linear trend of the data, decreasing from around AUD 7000 to slightly below AUD 5000. The summary of the model shows that the model is not significant, with a p-value of 0.08812 (>0.05) and an R-squared value of 0.01079 shows that only about 1% of the data is explained by the model. Overall, this is not a suitable model to represent the series being investigated.

```
# Fit linear model
linear.returns.ts = lm(returns.ts~time(returns.ts))
summary(linear.returns.ts)
```

```
##
## Call:
## lm(formula = returns.ts ~ time(returns.ts))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -104.186  -58.689   -5.424   57.532  172.072
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        74.245     11.245   6.602 4.55e-10 ***
## time(returns.ts)   -2.192      1.278  -1.715   0.0881 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.99 on 177 degrees of freedom
## Multiple R-squared:  0.01634,    Adjusted R-squared:  0.01079
## F-statistic: 2.941 on 1 and 177 DF,  p-value: 0.08812
```

```
plot(returns.ts, xaxt = "n", type='o', xlab = 'Days', ylab='Returns in AUD100', main = "Fitted linear line to returns data in AUD100")
axis(1, at=seq(1, 14, by=3), labels=seq(1, 179, by=36))
abline(linear.returns.ts, lty=2,col="red")
```

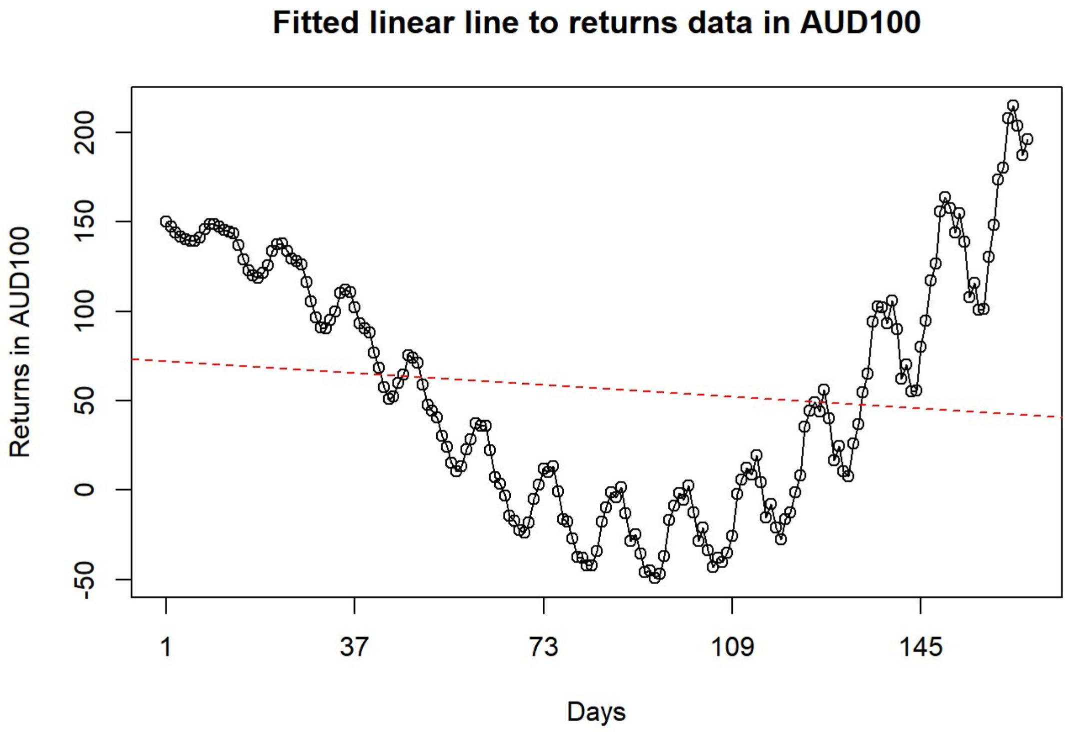

Figure 3: Plot of the time series returns data in AUD100 with fitted linear line

The *rstudent()* function is used to obtain the residuals for the linear model and a function called *plot_residuals()* is created to simplify the process of plotting the standardized residuals and their histogram, QQ plot, ACF plot, and PACF plot. These are summarized below:

- **Standardized residuals:** Seasonal components and trend remain in the residuals
- **Histogram:** Histogram of residuals does not show normal distribution, suggesting residuals are significant
- **QQ plot:** QQ plot shows two tail ends departing the reference line, indicating there is no normally distributed stochastic component in the model
- **ACF plot:** Significant peaks at all lags with a wave pattern visible indicating seasonality in residuals.
- **PACF plot:** Several significant peaks beyond the confidence intervals are recorded at multiple lags, indicating that the series is not comparable to a white noise process.

A Sharpiro-Wilk test produces a p-value that is less than 0.05, confirming that the residuals are not normal due to the null hypothesis being rejected.

```
res.linear.returns.ts = rstudent(linear.returns.ts)

plot_residuals = function(res, time.returns) {
  par(mfrow=c(3,2))
  par(mar=c(4,4,3,1.5))
  plot(y = res, xaxt = "n", x = as.vector(time.returns), xlab = 'Days', ylab='Standardized Residuals',type
='l', main = "Standardised residuals from model.")
  axis(1, at=seq(1, 14, by=3), labels=seq(1, 179, by=36))
  hist(res,xlab='Standardized Residuals', main = "Histogram of standardised residuals.")
  qqnorm(y=res, main = "QQ plot of standardised residuals.")
  qqline(y=res, col = 2, lwd = 1, lty = 2)
  acf(res, main = "ACF of standardized residuals.")
  pacf(res, main = "PACF of standardized residuals.")
  par(mfrow=c(1,1))
}

plot_residuals(res.linear.returns.ts, time(returns.ts))
```

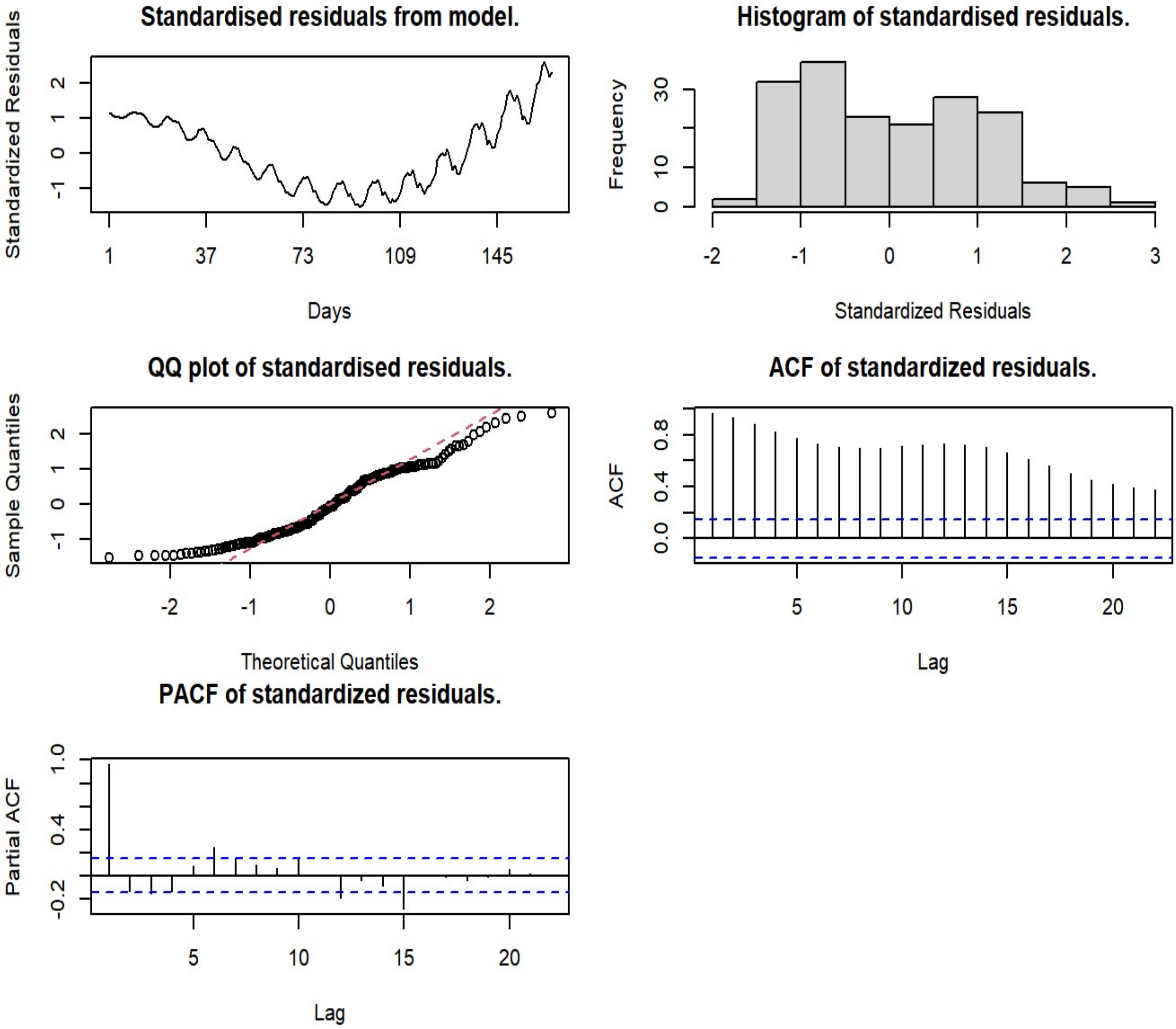

Figure 4: Plot of the residuals along with their histogram, QQ plot, ACF plot, and PACF plot

```
shapiro.test(res.linear.returns.ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res.linear.returns.ts
## W = 0.9525, p-value = 1.022e-05
```

Quadratic Model

A quadratic model may better explain the trend of the series since it is changing throughout the time period. A quadratic model is fitted to the data and the results are plotted. The resulting model (in red) is a curve that follows the trend of the data. The summary of the model shows that the model is significant, with a low p-value of 2.2e-16 (<0.05) and an R-squared value of 0.8523 shows that about 85% of the data is explained by the model. Overall, this is a very suitable model to represent the series being investigated.

```
# Fit quadratic model
t = time(returns.ts)
t2 = t^2
quad.returns.ts = lm(returns.ts ~ t + t2)
summary(quad.returns.ts)
```

```
##
## Call:
## lm(formula = returns.ts ~ t + t2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -59.258 -19.321   0.571  19.992  53.165
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  276.295      7.703   35.87   <2e-16 ***
## t            -71.476      2.236  -31.96   <2e-16 ***
## t2             4.415      0.139   31.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.27 on 176 degrees of freedom
## Multiple R-squared:  0.8539, Adjusted R-squared:  0.8523
## F-statistic: 514.4 on 2 and 176 DF,  p-value: < 2.2e-16
```

```
plot(ts(fitted(quad.returns.ts)), ylim = c(min(c(fitted(quad.returns.ts),
    as.vector(returns.ts))), max(c(fitted(quad.returns.ts),as.vector(returns.ts)))),
    xlab = 'Days', ylab='Returns data in AUD100', main = "Fitted quadratic curve to returns data in AUD100",
type="l",lty=2,col="red")
# axis(1, at=seq(1, 14, by=3), labels=seq(1, 179, by=36))
lines(as.vector(returns.ts),type="o")
```

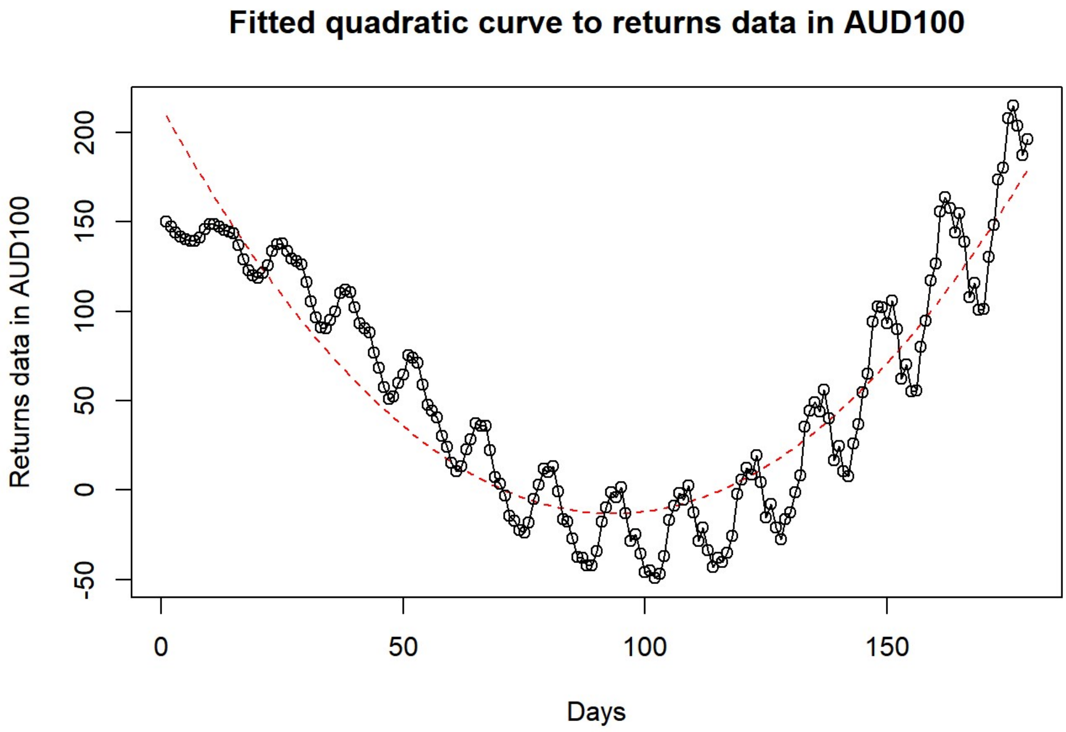

Figure 5: Plot of the time series returns data in AUD100 with fitted quadratic curve

The residuals and their relevant plots are summarized below:

- **Standardized residuals:** Seasonal components remain in the residuals, but less trend compared to the linear model
- **Histogram:** Histogram of residuals shows a normal distribution, suggesting residuals are not significant
- **QQ plot:** QQ plot shows points that do not depart from the reference line, indicating there is a normally distributed stochastic component in the model
- **ACF plot:** Fewer significant peaks compared to the linear model, with a wave pattern visible indicating seasonality in residuals.
- **PACF plot:** Several significant peaks beyond the confidence intervals are recorded at multiple lags, indicating that the series is not comparable to a white noise process.

A Sharpiro-Wilk test produces a p-value that is slightly less than 0.05, suggesting that the residuals are not normal due to the null hypothesis being rejected. However, the fact that it is quite close to 0.05 casts doubt onto that assumption.

```
res.quad.returns.ts = rstudent(quad.returns.ts)

plot_residuals(res.quad.returns.ts, time(returns.ts))
```

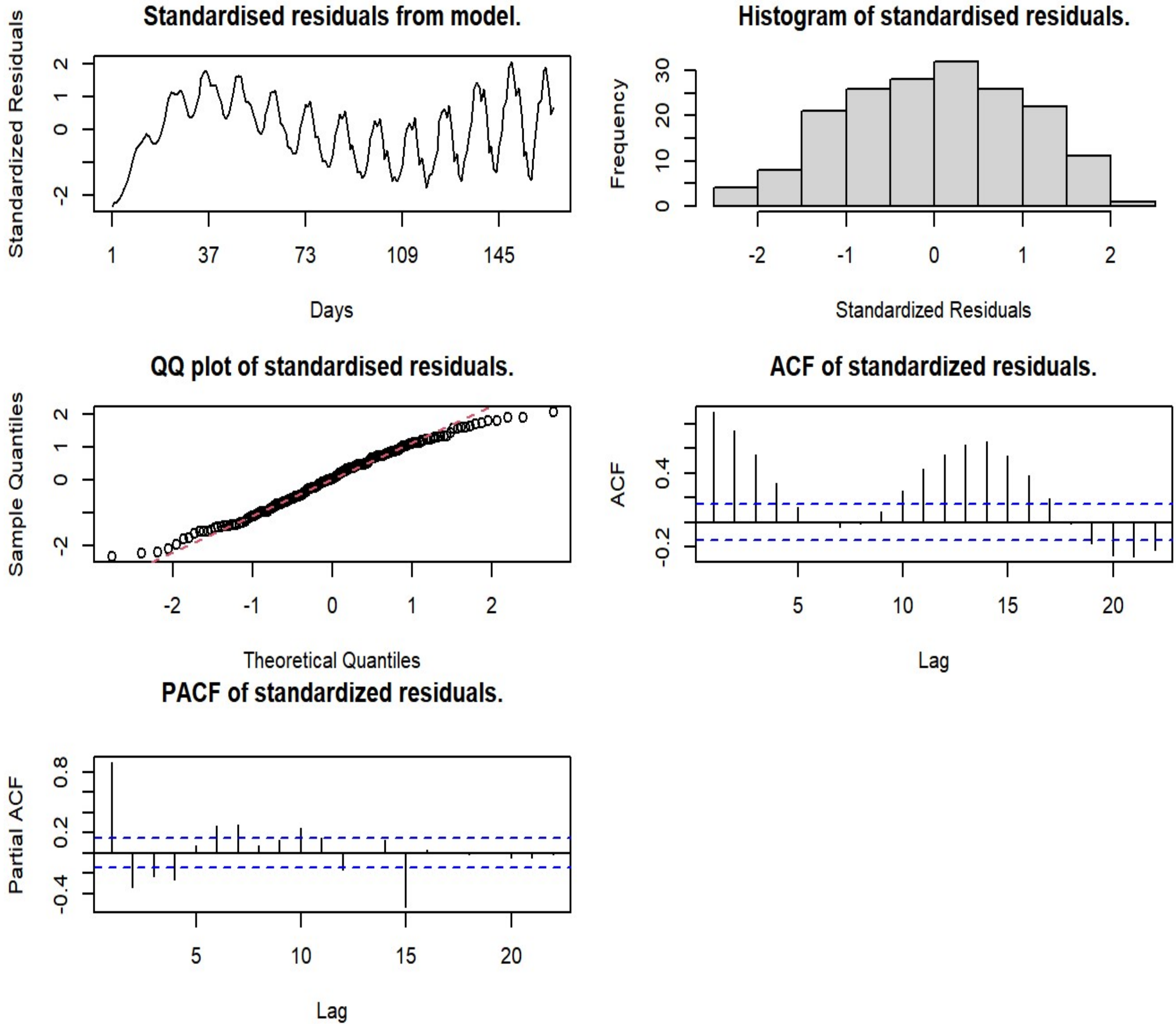

Figure 6: Plot of the residuals along with their histogram, QQ plot, ACF plot, and PACF plot

```
shapiro.test(res.quad.returns.ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res.quad.returns.ts
## W = 0.98397, p-value = 0.03799
```

Harmonic Model

A cosine or harmonic model could be used to model some of the seasonal components of the data since the amplitude, frequency and phase of a cosine wave could approximated the seasonal tendencies of the series (Canvas, 2024). As such, a harmonic model is fitted to the data and the results are plotted. The resulting model (in red) is a constant sinusoidal pattern that attempts to capture the seasonal components of the data but does not capture the trend. The summary of the model shows that the model is not significant, with a p-value of 0.1618 (>0.05) and an R-squared value of 0.009354 shows that only about 0.9% of the data is explained by the model. The cosine component is highly insignificant for this model, while the sine component is only slightly insignificant, with a p-value of 0.0601. Overall, this is not a suitable model to represent the series being investigated.

```
# Fit harmonic model
har. = harmonic(returns.ts, 1)
data = data.frame(returns.ts,har.)
har.returns.ts = lm(returns.ts ~ cos.2.pi.t. + sin.2.pi.t. , data = data)
summary(har.returns.ts)
```

```
##
## Call:
## lm(formula = returns.ts ~ cos.2.pi.t. + sin.2.pi.t., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -96.240 -63.559  -6.592  59.015 152.142
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   56.847      5.087  11.176   <2e-16 ***
## cos.2.pi.t.   -2.427      7.178  -0.338   0.7356
## sin.2.pi.t.   13.641      7.209   1.892   0.0601 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68.04 on 176 degrees of freedom
## Multiple R-squared:  0.02049,    Adjusted R-squared:  0.009354
## F-statistic:  1.84 on 2 and 176 DF,  p-value: 0.1618
```

```
fitted.har.returns.ts = fitted(har.returns.ts)

plot(ts(fitted(har.returns.ts)), ylim = c(min(c(fitted(har.returns.ts),
     as.vector(returns.ts))), max(c(fitted(har.returns.ts),as.vector(returns.ts)))),
     xlab = 'Days', ylab='Returns data in AUD100' , main = "Fitted cosine curve to returns data in AUD100", t
ype="l",lty=2,col="red")

lines(as.vector(returns.ts),type="o")
```

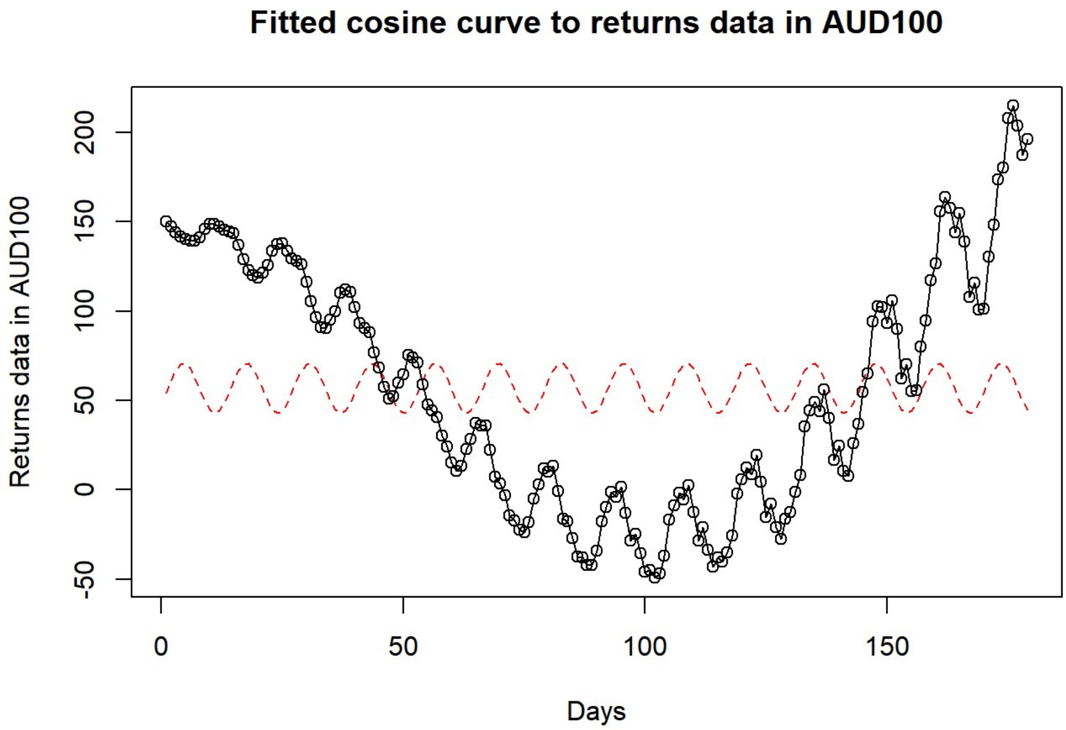

Figure 7: Plot of the time series returns data in AUD100 with fitted cosine curve

The residuals and their relevant plots are summarized below:

- **Standardized residuals:** Seasonal components and trend remain in the residuals
- **Histogram:** Histogram of residuals does not show normal distribution, suggesting residuals are significant
- **QQ plot:** QQ plot shows two tail ends departing the reference line, indicating there is no normally distributed stochastic component in the model
- **ACF plot:** Significant peaks at all lags with a wave pattern visible indicating seasonality in residuals.
- **PACF plot:** Several significant peaks beyond the confidence intervals are recorded at multiple lags, indicating that the series is not comparable to a white noise process.

A Sharpiro-Wilk test produces a p-value that is less than 0.05, confirming that the residuals are not normal due to the null hypothesis being rejected.

```
res.har.returns.ts = rstudent(har.returns.ts)

plot_residuals(res.har.returns.ts, time(returns.ts))
```

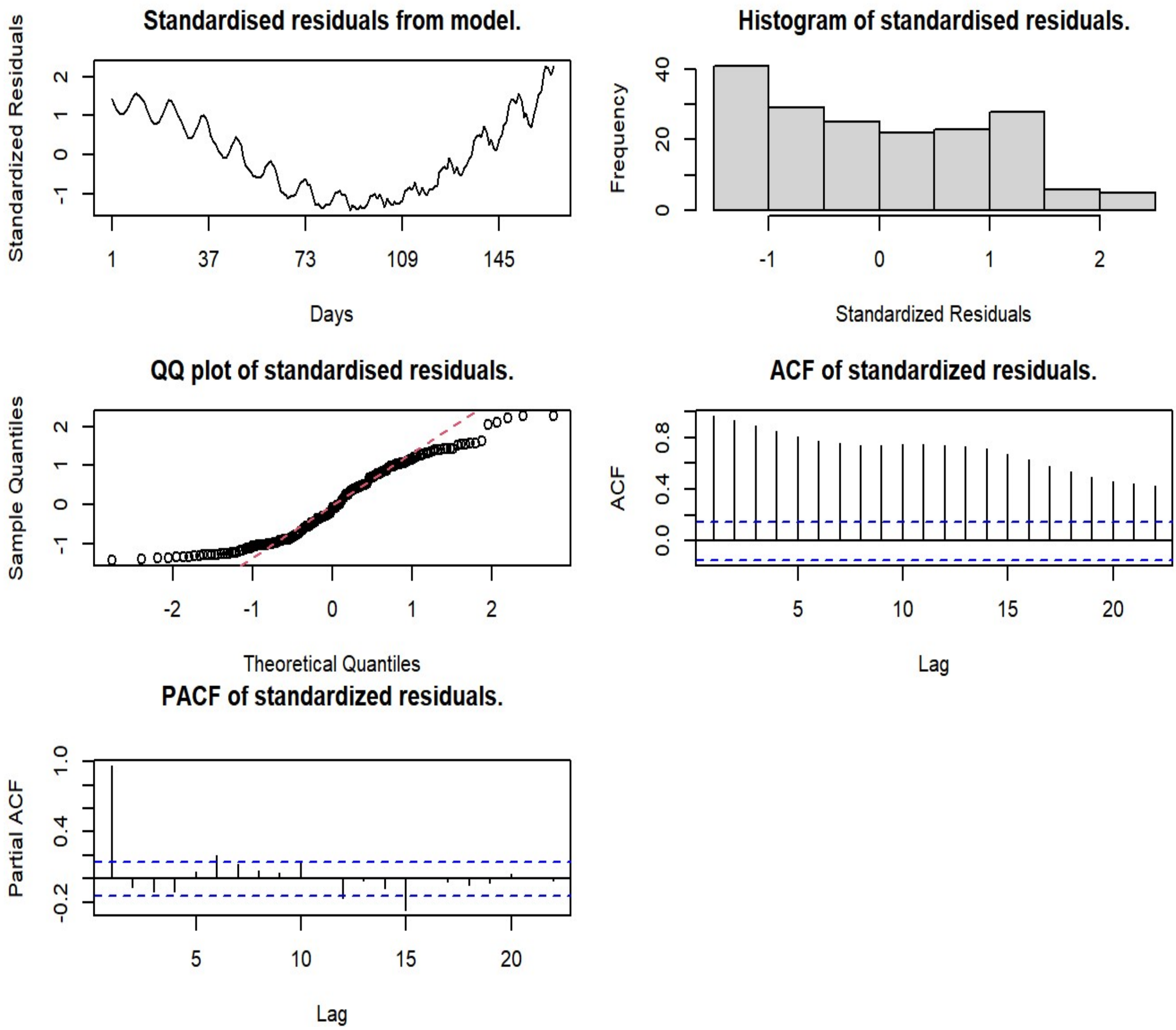

Figure 8: Plot of the residuals along with their histogram, QQ plot, ACF plot, and PACF plot

```
shapiro.test(res.har.returns.ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res.har.returns.ts
## W = 0.93475, p-value = 3.077e-07
```

Seasonal Model

A seasonal model is fitted to the data and the results are plotted. The resulting model (in red) is a constant repeating pattern that attempts to capture the seasonal components of the data but does not capture the trend. The shape of the pattern seems to capture the specific seasonality of this series better than the harmonic model. The summary of the model shows that the model is significant, with a p-value of 1.747e-14 (<0.05) and an R-squared value of 0.3798 shows that about 38% of the data is explained by the model. Overall, this is not a suitable model to represent the series being investigated despite the significance of the model since it does not describe a sufficient proportion of the data.

```
# Fit seasonal model
seas.=season(returns.ts)
seas.returns.ts=lm(returns.ts~seas.-1) # -1 removes the intercept term
summary(seas.returns.ts)
```

```
##
## Call:
## lm(formula = returns.ts ~ seas. - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -97.730 -64.795  -7.509  58.863 153.035
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## seas.Season-1     57.85      18.71   3.092 0.002332 **
## seas.Season-2     63.74      18.71   3.407 0.000824 ***
## seas.Season-3     67.61      18.71   3.614 0.000399 ***
## seas.Season-4     69.48      18.71   3.714 0.000278 ***
## seas.Season-5     68.52      18.71   3.663 0.000335 ***
## seas.Season-6     66.39      18.71   3.549 0.000504 ***
## seas.Season-7     61.58      18.71   3.291 0.001218 **
## seas.Season-8     56.22      18.71   3.005 0.003068 **
## seas.Season-9     52.16      18.71   2.788 0.005924 **
## seas.Season-10    50.12      18.71   2.679 0.008123 **
## seas.Season-11    38.22      19.41   1.969 0.050660 .
## seas.Season-12    41.32      19.41   2.128 0.034778 *
## seas.Season-13    45.03      19.41   2.320 0.021577 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 70 on 166 degrees of freedom
## Multiple R-squared:  0.4249, Adjusted R-squared:  0.3798
## F-statistic: 9.433 on 13 and 166 DF,  p-value: 1.747e-14
```

```
plot(ts(fitted(seas.returns.ts)), ylim = c(min(c(fitted(seas.returns.ts),
     as.vector(returns.ts))), max(c(fitted(seas.returns.ts),as.vector(returns.ts)))),
     xlab = 'Days', ylab='Returns data in AUD100', main = "Fitted seasonal curve to returns data", type="l",l
ty=2,col="red")
lines(as.vector(returns.ts),type="o")
```

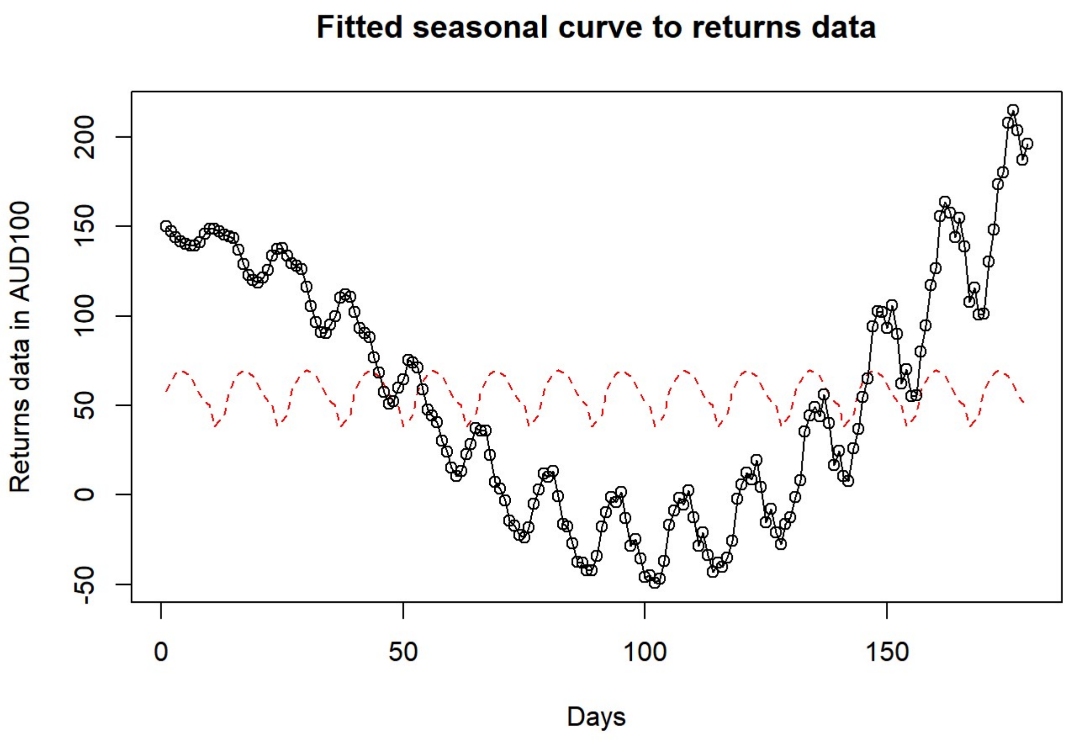

Figure 9: Plot of the time series returns data in AUD100 with fitted seasonal curve

The residuals and their relevant plots are summarized below:

- **Standardized residuals:** Seasonal components and trend remain in the residuals
- **Histogram:** Histogram of residuals does not show normal distribution, suggesting residuals are significant
- **QQ plot:** QQ plot shows two tail ends departing the reference line, indicating there is no normally distributed stochastic component in the model
- **ACF plot:** Significant peaks at all lags with a wave pattern visible indicating seasonality in residuals.
- **PACF plot:** Several significant peaks beyond the confidence intervals are recorded at multiple lags, indicating that the series is not comparable to a white noise process.

A Sharpiro-Wilk test produces a p-value that is less than 0.05, confirming that the residuals are not normal due to the null hypothesis being rejected.

```
res.seas.returns.ts = rstudent(seas.returns.ts)
plot_residuals(res.seas.returns.ts, time(returns.ts))
```

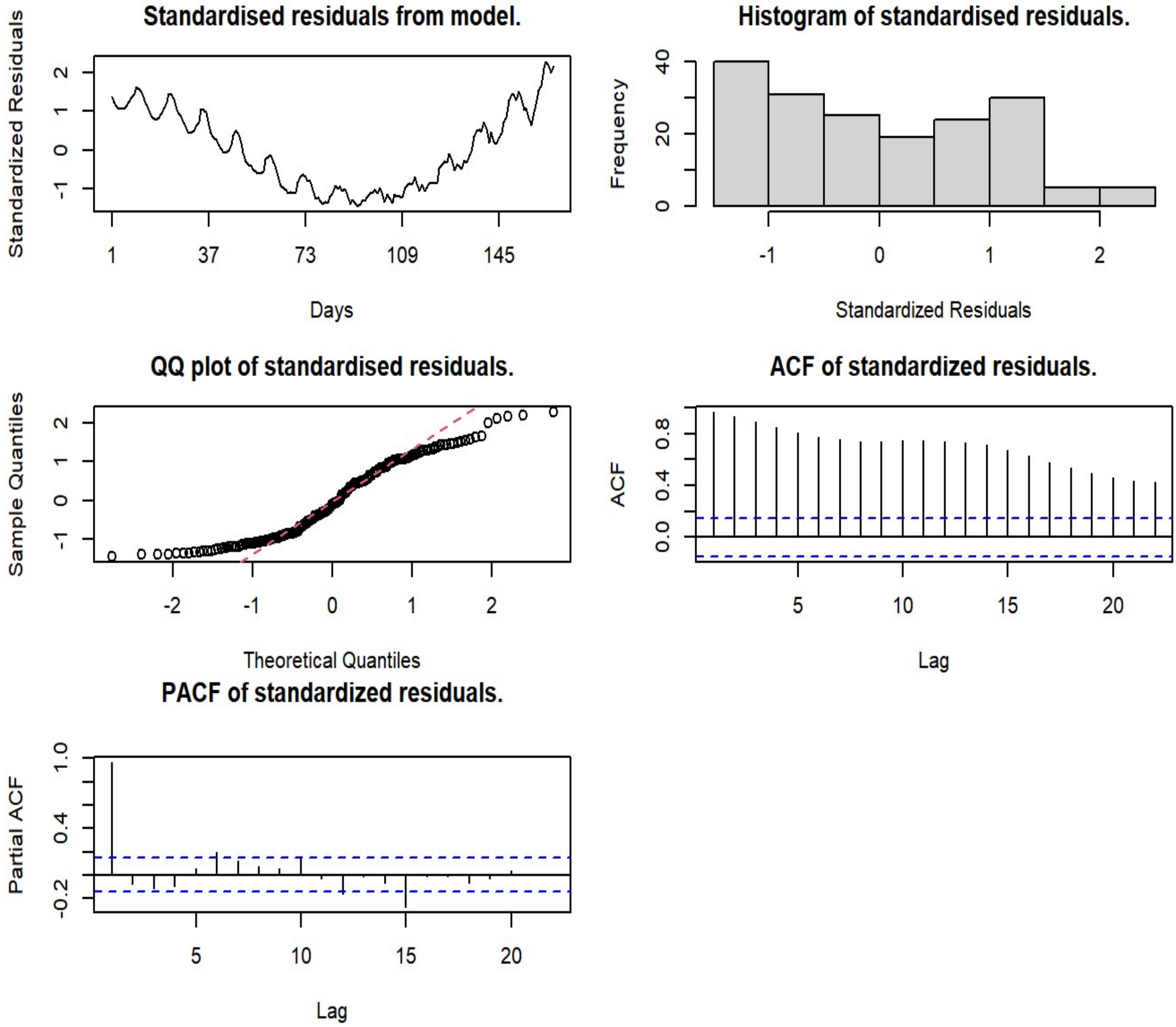

Figure 10: Plot of the residuals along with their histogram, QQ plot, ACF plot, and PACF plot

```
shapiro.test(res.seas.returns.ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res.seas.returns.ts
## W = 0.93578, p-value = 3.716e-07
```

Harmonic-Quadratic Model

A model combining harmonic and quadratic components could provide improvements to the quadratic model due to its ability to capture seasonal components as well. A model combining both harmonic and seasonal components is fitted to the data and the results are plotted. The resulting model (in red) is a constant repeating pattern that attempts to capture the seasonal components as well as the trend. The summary of the model shows that the model is significant, with a p-value of 2.2e-16 (<0.05) and an R-squared value of 0.8616 shows that about 86% of the data is explained by the model. The cosine component of the model is insignificant, however. For these reasons, this is a suitable model to represent the series being investigated.

```
# Fit harmonic-quadratic model
har. = harmonic(returns.ts, 1)
data = data.frame(returns.ts,har.)
t = time(returns.ts)
t2 = t^2
harquad.returns.ts=lm(returns.ts~ cos.2.pi.t. + sin.2.pi.t. + t + t2, data = data) # -1 removes the intercept
term
summary(harquad.returns.ts)
```

```
##
## Call:
## lm(formula = returns.ts ~ cos.2.pi.t. + sin.2.pi.t. + t + t2,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.928 -18.761  -2.631  18.422  54.048
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 274.9555     7.4641  36.837  < 2e-16 ***
## cos.2.pi.t.  -0.5770     2.6853  -0.215 0.830106
## sin.2.pi.t.  10.0413     2.6970   3.723 0.000266 ***
## t           -71.1307     2.1665 -32.832  < 2e-16 ***
## t2            4.3957     0.1347  32.645  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.43 on 174 degrees of freedom
## Multiple R-squared:  0.8647, Adjusted R-squared:  0.8616
## F-statistic: 278.1 on 4 and 174 DF,  p-value: < 2.2e-16
```

```
plot(ts(fitted(harquad.returns.ts)), ylim = c(min(c(fitted(harquad.returns.ts),
     as.vector(returns.ts))), max(c(fitted(harquad.returns.ts),as.vector(returns.ts)))),
     xlab = 'Days', ylab='Returns data in AUD100', main = "Fitted harmonic-quadratic curve to returns data",
type="l",lty=2,col="red")
lines(as.vector(returns.ts),type="o")
```

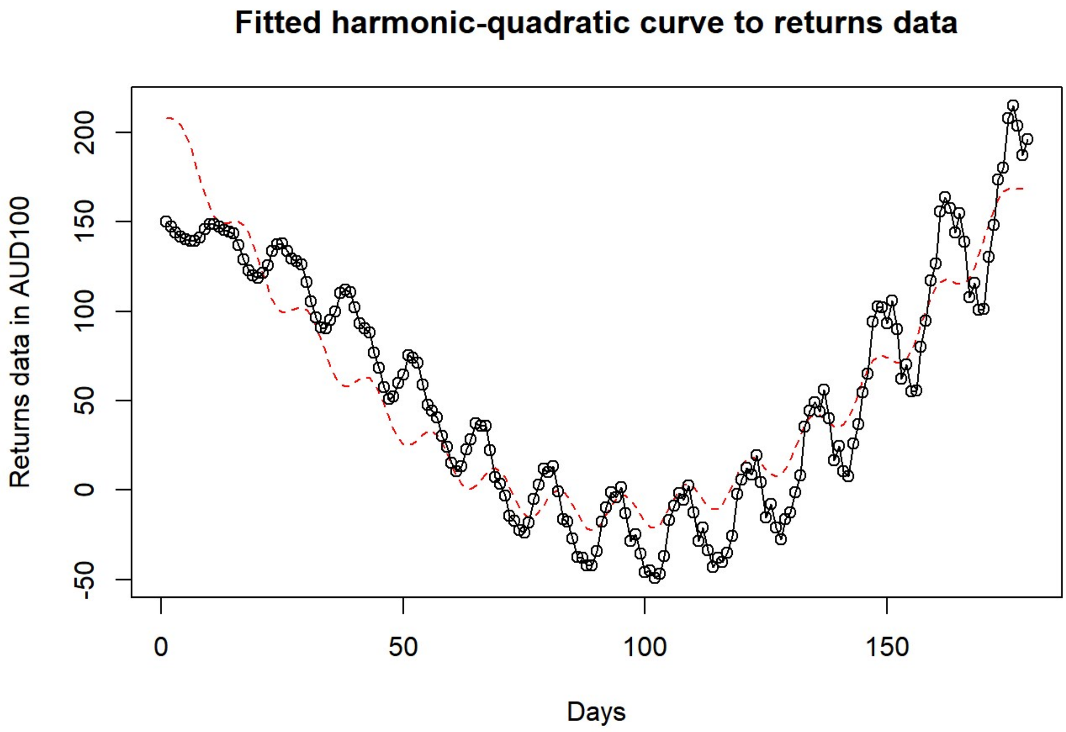

Figure 11: Plot of the time series returns data in AUD100 with fitted harmonic-quadratic curve

The residuals and their relevant plots are summarized below:

- **Standardized residuals:** Seasonal components and a slight trend remain in the residuals
- **Histogram:** Histogram of residuals does not show normal distribution, suggesting residuals are significant
- **QQ plot:** QQ plot shows most of the points not departing the reference line, except for small disturbances in the two tails, indicating there may be a normally distributed stochastic component in the model
- **ACF plot:** Significant peaks at most lags with a wave pattern visible indicating seasonality in residuals.
- **PACF plot:** Several significant peaks beyond the confidence intervals are recorded at multiple lags, indicating that the series is not comparable to a white noise process.

A Sharpiro-Wilk test produces a p-value that is larger than 0.05, confirming that the residuals are likely normal due to the null hypothesis not being rejected. However, the fact that it is only slightly larger (0.06225) casts doubt on that assumption.

```
res.harquad.returns.ts = rstudent(harquad.returns.ts)
plot_residuals(res.harquad.returns.ts, time(returns.ts))
```

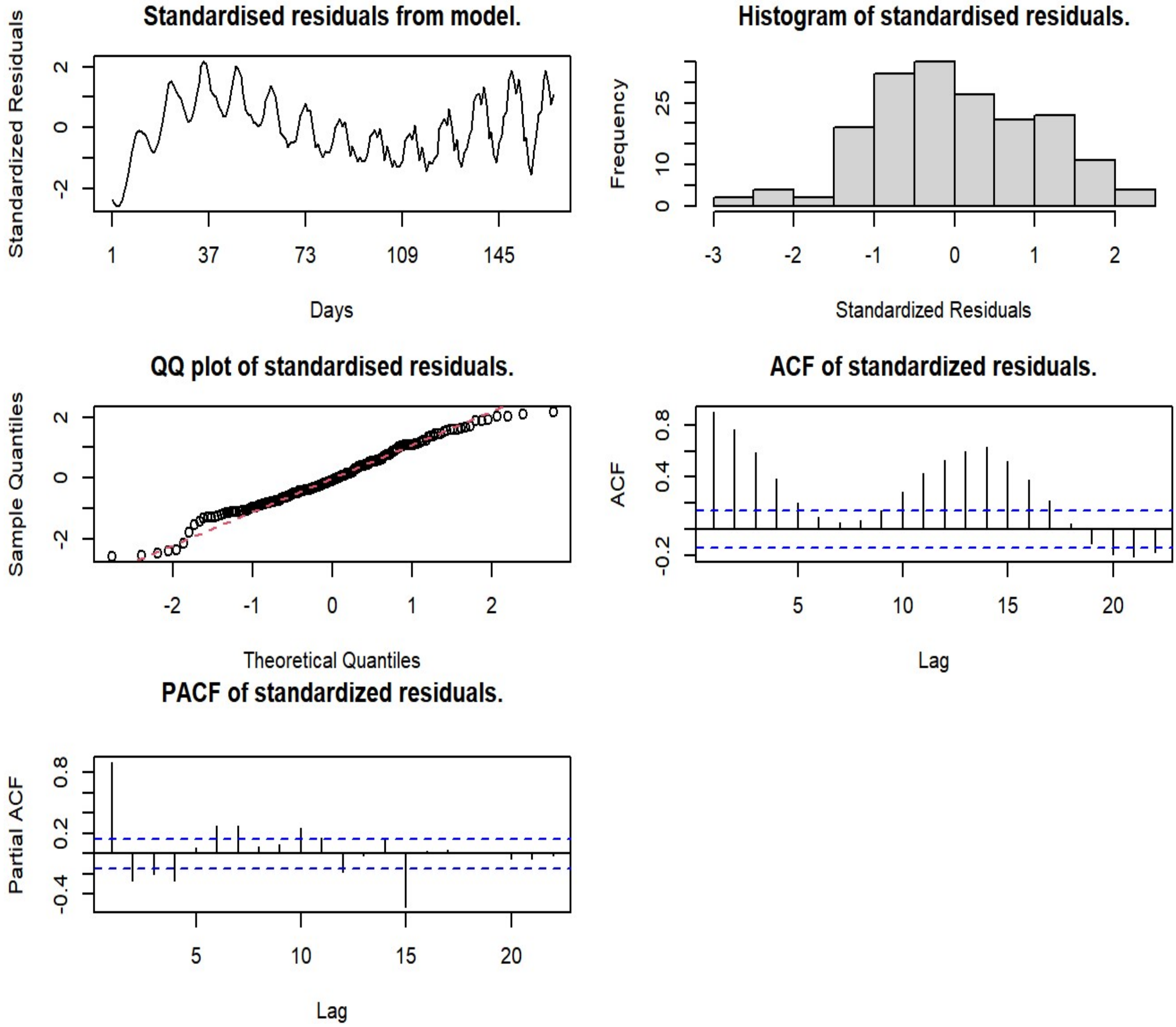

Figure 12: Plot of the residuals along with their histogram, QQ plot, ACF plot, and PACF plot

```
shapiro.test(res.harquad.returns.ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res.harquad.returns.ts
## W = 0.98556, p-value = 0.06225
```

Seasonal-Quadratic Model

A model combining seasonal and quadratic components could provide improvements to the quadratic model due to its ability to capture seasonal components as well. A model combining both seasonal and quadratic components is fitted to the data and the results are plotted. The resulting model (in red) is a constant repeating pattern that attempts to capture the seasonal components as well as the trend. The summary of the model shows that the model is significant, with a p-value of 2.2e-16 (<0.05) and an R-squared value of 0.9132 shows that about 91% of the data is explained by the model. All the components of the model are significant as well. This was the best performing model to represent the series that has been implemented thus far.

```
# Fit seasonal-quadratic model
seas.=season(returns.ts)
t = time(returns.ts)
t2 = t^2
seasquad.returns.ts=lm(returns.ts~ seas. + t + t2 -1) # -1 removes the intercept term
summary(seasquad.returns.ts)
```

```
##
## Call:
## lm(formula = returns.ts ~ seas. + t + t2 - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.196 -18.423  -2.084  18.663  53.120
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## seas.Season-1  272.7195    10.0982   27.01   <2e-16 ***
## seas.Season-2  278.9824    10.1163   27.58   <2e-16 ***
## seas.Season-3  283.1832    10.1334   27.95   <2e-16 ***
## seas.Season-4  285.3199    10.1496   28.11   <2e-16 ***
## seas.Season-5  284.5758    10.1649   28.00   <2e-16 ***
## seas.Season-6  282.6070    10.1793   27.76   <2e-16 ***
## seas.Season-7  277.9117    10.1928   27.27   <2e-16 ***
## seas.Season-8  272.6147    10.2055   26.71   <2e-16 ***
## seas.Season-9  268.5631    10.2172   26.29   <2e-16 ***
## seas.Season-10 266.4862    10.2281   26.05   <2e-16 ***
## seas.Season-11 264.0656    10.5339   25.07   <2e-16 ***
## seas.Season-12 267.3595    10.5483   25.35   <2e-16 ***
## seas.Season-13 271.2121    10.5617   25.68   <2e-16 ***
## t              -71.1527     2.2332  -31.86   <2e-16 ***
## t2               4.3969     0.1388   31.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.18 on 164 degrees of freedom
## Multiple R-squared:  0.9205, Adjusted R-squared:  0.9132
## F-statistic: 126.6 on 15 and 164 DF,  p-value: < 2.2e-16
```

```
plot(ts(fitted(seasquad.returns.ts)), ylim = c(min(c(fitted(seasquad.returns.ts),
     as.vector(returns.ts))), max(c(fitted(seasquad.returns.ts),as.vector(returns.ts)))),
     xlab = 'Days', ylab='Returns data in AUD100', main = "Fitted seasonal-quadratic curve to returns data",
type="l",lty=2,col="red")
lines(as.vector(returns.ts),type="o")
```

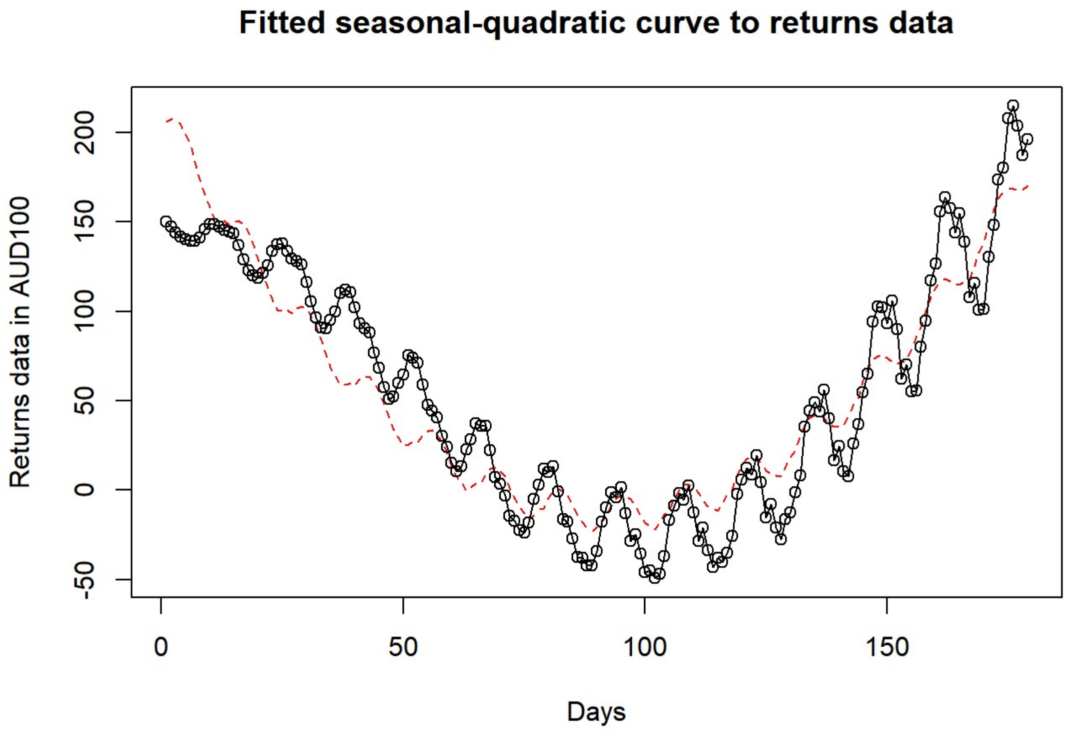

Figure 13: Plot of the time series returns data in AUD100 with fitted seasonal-quadratic curve

The residuals and their relevant plots are summarized below:

- **Standardized residuals:** Seasonal components and a slight trend remain in the residuals
- **Histogram:** Histogram of residuals shows an almost normal distribution except for the left tail, suggesting residuals may be insignificant
- **QQ plot:** QQ plot shows most of the points not departing the reference line, except for very small disturbances in the two tails, indicating there may be a normally distributed stochastic component in the model
- **ACF plot:** Significant peaks at most lags with a wave pattern visible indicating seasonality in residuals.
- **PACF plot:** Several significant peaks beyond the confidence intervals are recorded at multiple lags, indicating that the series is not comparable to a white noise process.

A Sharpiro-Wilk test produces a p-value that is larger than 0.05, confirming that the residuals are likely normal due to the null hypothesis not being rejected. However, the fact that it is only slightly larger (0.06655) casts doubt on that assumption.

```
res.seasquad.returns.ts = rstudent(seasquad.returns.ts)
plot_residuals(res.seasquad.returns.ts, time(returns.ts))
```

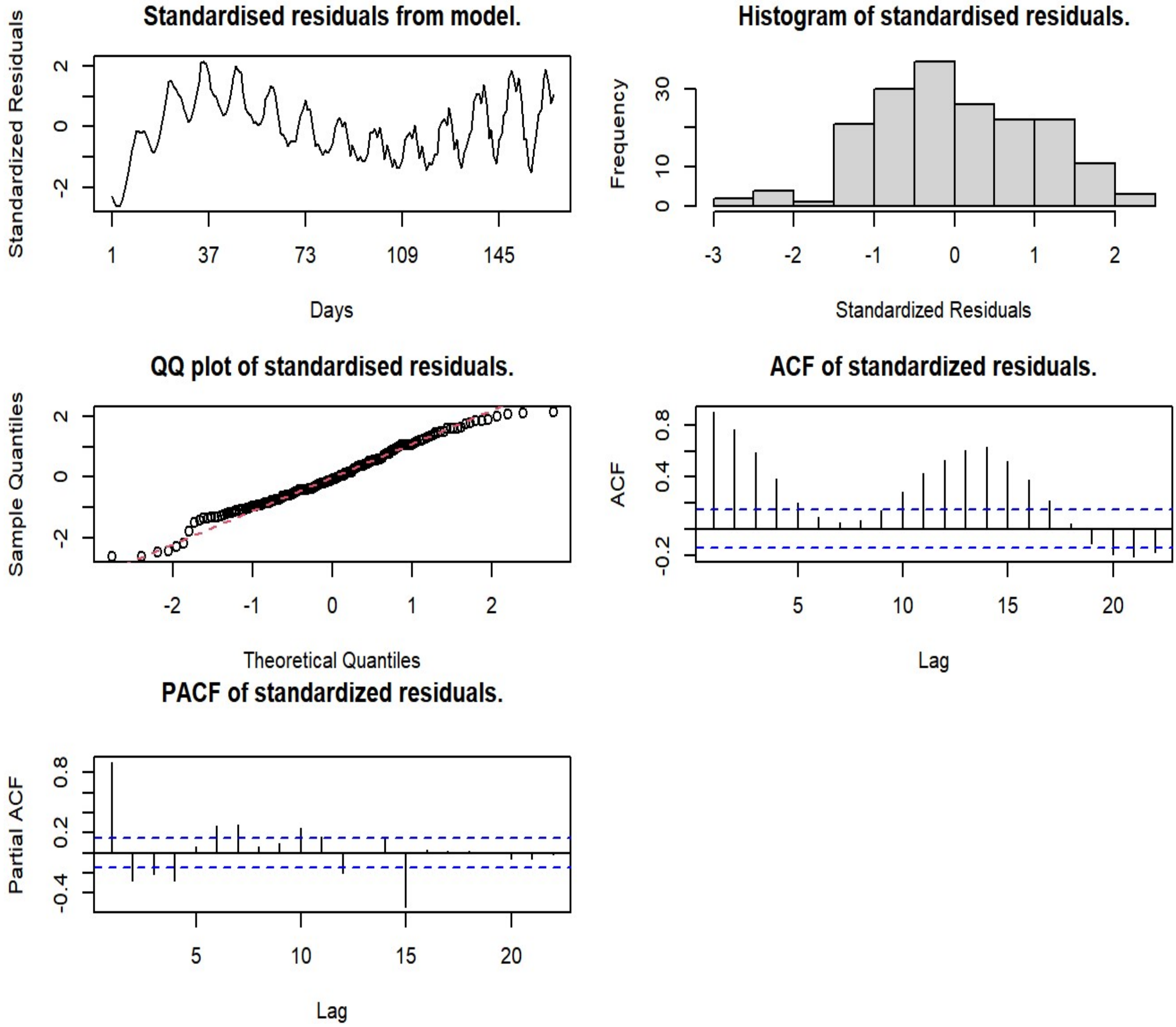

Figure 14: Plot of the residuals along with their histogram, QQ plot, ACF plot, and PACF plot

```
shapiro.test(res.seasquad.returns.ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res.seasquad.returns.ts
## W = 0.98577, p-value = 0.06655
```

Forecasting For Seasonal-Quadratic Model

A 5-day forecast is produced using the *predict()* function by providing it new data in the form of a dataframe which contains columns for the seasons, t and t^2. The resulting plot shows the original data in black, the fitted model in green, the forecast in red, and the upper and lower bounds of the confidence intervals in blue. The forecast predicts an increase in returns over the next 5 days. However, the range between the confidence interval bounds is quite significant which suggests that the model could be improved to be more accurate.

```
h = 5 # 5 steps ahead forecasts
t = time(returns.ts)
seas.=season(returns.ts)

lastTimePoint = t[length(t)]
dataFreq = frequency(returns.ts)
aheadTimes = data.frame(seas. = c("Season-11","Season-12","Season-13","Season-1","Season-2"),
                         t = seq(lastTimePoint+(1/dataFreq), lastTimePoint+h*(1/dataFreq), 1/dataFreq),
                         t2 =  seq(lastTimePoint+(1/dataFreq), lastTimePoint+h*(1/dataFreq), 1/dataFreq)^2)

frcquad = predict(seasquad.returns.ts, newdata = aheadTimes, interval = "prediction")

plot(returns.ts, xaxt = "n", xlim= c(t[1],aheadTimes$t[nrow(aheadTimes)]), ylim = c(-50,max(frcquad[,3])), xl
ab = "Days", ylab = "Returns in AUD100",
     main = "Forecasts from the seasonal-quadratic model fitted to the returns series")
axis(1, at=seq(1, 14, by=3), labels=seq(1, 179, by=36))
lines(ts(fitted(seasquad.returns.ts),start = t[1],frequency = dataFreq), col = "green")
lines(ts(as.vector(frcquad[,3]), start = aheadTimes$t[1],frequency = dataFreq), col="blue", type="l")
lines(ts(as.vector(frcquad[,1]), start = aheadTimes$t[1],frequency = dataFreq), col="red", type="l")
lines(ts(as.vector(frcquad[,2]), start = aheadTimes$t[1],frequency = dataFreq), col="blue", type="l")
legend("topleft", lty=1, pch=1, col=c("black","green","blue","red"),
       c("Data","Model","5% forecast limits", "Forecasts"))
```

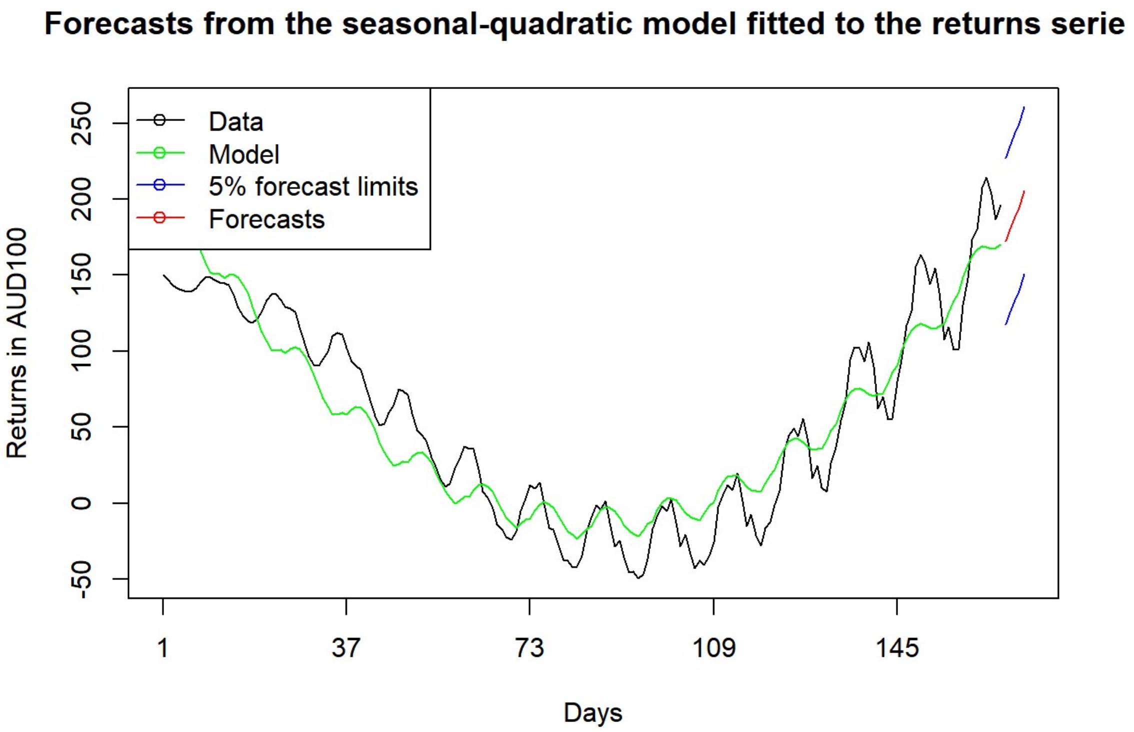

Figure 15: Combined plot of original returns series, fitted model, forecast, and CI bounds

Conclusion

During the investigation, a total of 6 models were fitted and tested for the time series data describing returns on a share market trader's investment portfolio over the course of 179 days. The series was seen to be non-stationary, with apparent seasonal and trend components. The linear model failed to sufficiently describe the trend and the seasonality seen in the series while the harmonic and seasonal models performed better at describing the seasonal components. However, they failed to grasp the trends seen in the series. The quadratic model performed well at describing the trend of the series, but did not catch the seasonal components. Finally, the harmonic-quadratic and seasonal-quadratic models both performed well by capturing the trend and seasonality of the data. However, seasonal components remained in the residuals for all models tested, suggesting that none of them could sufficiently capture all the seasonality in the data. Between the two combined models, the seasonal-quadratic had an improved R-squared value, while the harmonic-quadratic model produced a very slightly lower p-value for the Sharpiro-Wilk normality test. Given this information, it is concluded that the seasonal-quadratic model was the best at describing the returns data due to its significantly higher R-squared value compared to all other models.

References

- Canvas (2024) Module 1 - Online Notes.html.. Available at: https://rmit.instructure.com/courses/124176/files/36179161?module_item_id=5935305&fd_cookie_set=1module_item_id=5935330&fd_cookie_set=1 (https://rmit.instructure.com/courses/124176/files/36179161?module_item_id=5935305&fd_cookie_set=1module_item_id=5935330&fd_cookie_set=1) (accessed 26 March 2024).
- Canvas (2024) Module 2 - Online Notes.html.. Available at: https://rmit.instructure.com/courses/124176/files/36178912?module_item_id=5935330&fd_cookie_set=1 (https://rmit.instructure.com/courses/124176/files/36178912?module_item_id=5935330&fd_cookie_set=1) (accessed 26 March 2024).