

Homework 8

Benjamin Noland

1. a. The model specification can be written as

$$Y_t - Y_{t-4} = 0.5(Y_{t-1} - Y_{t-5}) + e_t - 0.3e_{t-1},$$

or equivalently as

$$\nabla_4 Y_t = 0.5 \nabla_4 Y_{t-1} + e_t - 0.3e_{t-1}.$$

The process $\{Y_t\}$ is therefore easily seen to be $\text{ARIMA}(1, 0, 1) \times (0, 1, 0)_4$.

- b. The model specification can be written as

$$(Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}) = e_t - 0.5e_{t-1} - 0.5e_{t-2} + 0.25e_{t-13}$$

or equivalently as

$$\nabla_{12} Y_t - \nabla_{12} Y_{t-1} = e_t - 0.5e_{t-1} - 0.5e_{t-2} + 0.25e_{t-13}$$

This in turn is equivalent to

$$\nabla \nabla_{12} Y_t = e_t - 0.5e_{t-1} - 0.5e_{t-2} + 0.25e_{t-13}.$$

Note that the MA characteristic polynomial of this differenced series can be factored as

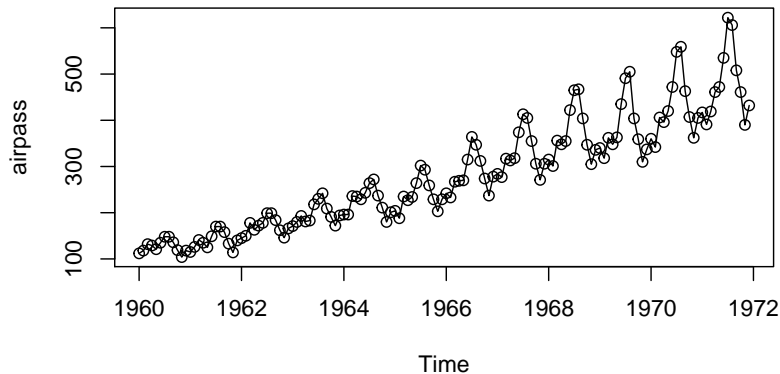
$$1 - 0.5x - 0.5x^{12} + 0.25x^{13} = (1 - 0.5x^{12})(1 - 0.5x) = \Theta(x)\theta(x).$$

The process $\{Y_t\}$ is therefore $\text{ARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$.

2. (Cryer & Chan, Exercise 10.9)

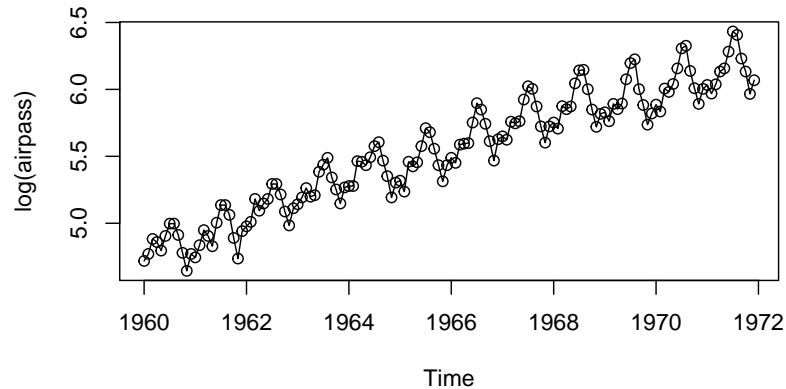
- a. A plot of the original time series is shown below:

```
data(airpass)
plot(airpass, type = "o")
```



This plot shows that the time series is clearly non-stationary. In particular, it has non-constant mean, and the series grows more variable with time. After taking logarithms, we get the following:

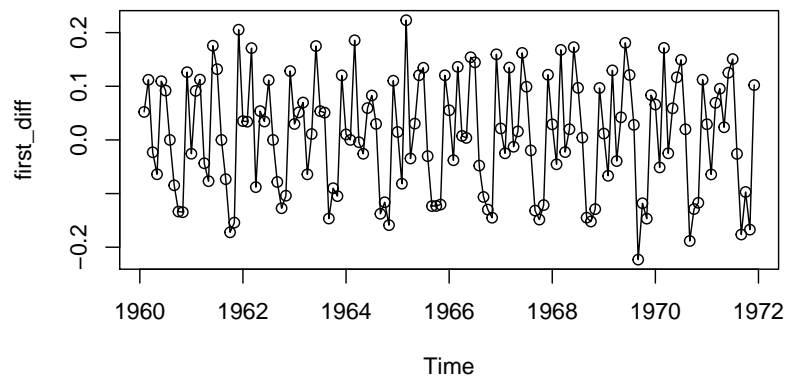
```
plot(log(airpass), type = "o")
```



Although this transformed series appears to still have non-constant mean, the series no longer appears to be getting more variable with time. Thus the transformation seems appropriate.

- b. After taking first differences of the log-transformed series, we get the following:

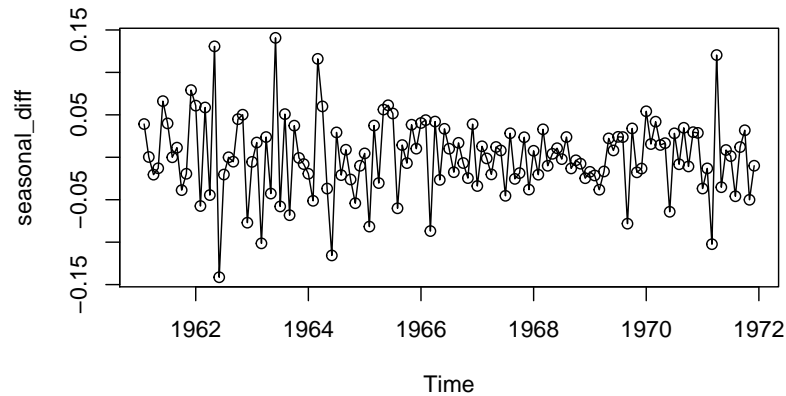
```
first_diff <- diff(log(airpass))
plot(first_diff, type = "o")
```



This new series appears to be stationary (in particular, it has taken care of the issue of non-constant mean in the undifferenced series). However, there still appears to be seasonality in the series.

- c. It seems reasonable to expect a seasonality of $s = 12$. Taking seasonal differences with $s = 12$, we get the following:

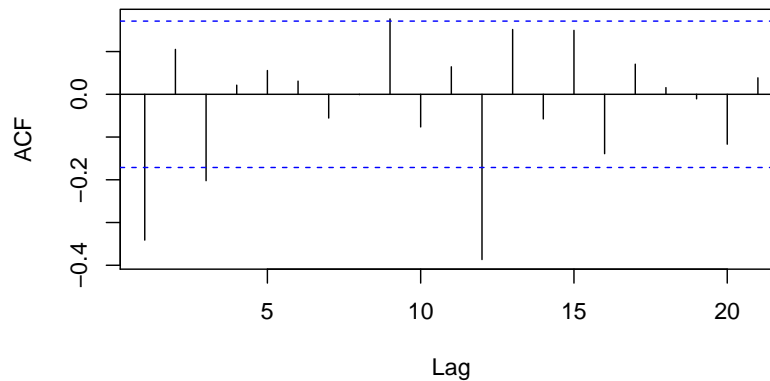
```
seasonal_diff <- diff(first_diff, lag = 12)
plot(seasonal_diff, type = "o")
```



The seasonal differencing appears to have removed the seasonality in the series.

- d. A plot of the sample ACF of the seasonal difference of the first difference of the log-transformed series is as follows:

```
acf(as.vector(seasonal_diff), main = NULL)
```



The ACF is statistically significant from zero at lag 1, and then (with a few exceptions), is zero for higher lags. This suggests that this transformed series *may* be suitably modeled by an MA(1) process.

- e. The following fits an $\text{ARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ model to the log-transformed series:

```
fit <- arima(log(airpass), order = c(0, 1, 1),
             seasonal = list(order = c(0, 1, 1), period = 12))
fit
```

```
##
```

```
## Call:
```

```
## arima(x = log(airpass), order = c(0, 1, 1), seasonal = list(order = c(0, 1,
##      1), period = 12))
```

```
##
```

```
## Coefficients:
```

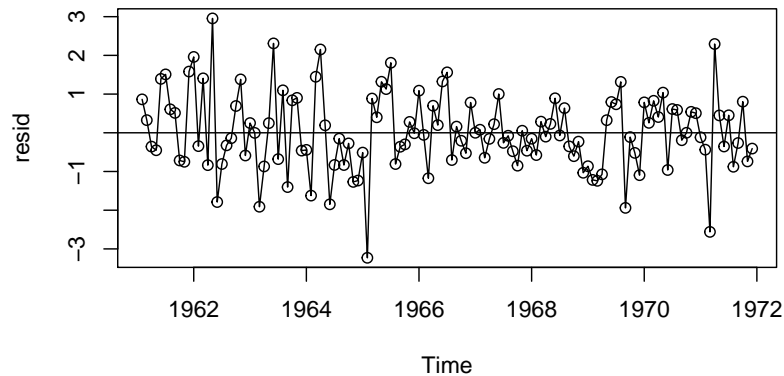
```
##      ma1      sma1
```

```
##    -0.402  -0.557
```

```
## s.e.    0.090    0.073
##
## sigma^2 estimated as 0.00135:  log likelihood = 245,  aic = -485
```

f. The following is a plot of the standardized residuals from the model:

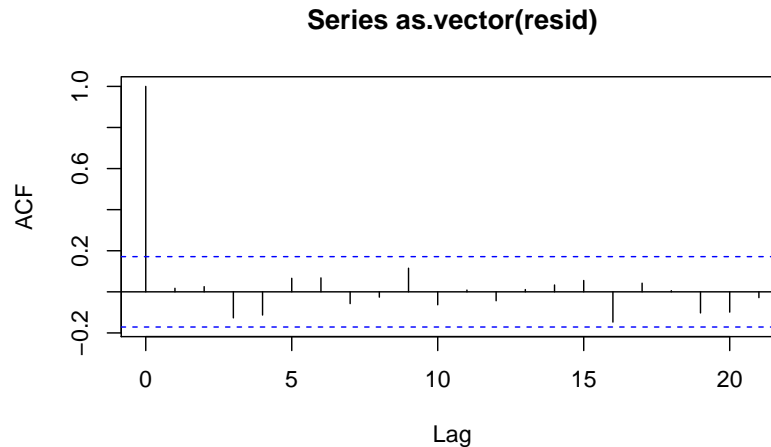
```
resid <- window(rstandard(fit), start = c(1961, 2))
plot(resid, type = "o")
abline(h = 0)
```



The residuals do not appear to be white noise. In particular, the variability in the series is very high from about 1961 to 1965, and then dwindles from about 1966 to 1970.

A plot of the sample ACF of the standardized residuals is as follows:

```
acf(as.vector(resid), drop.lag.0 = FALSE)
```

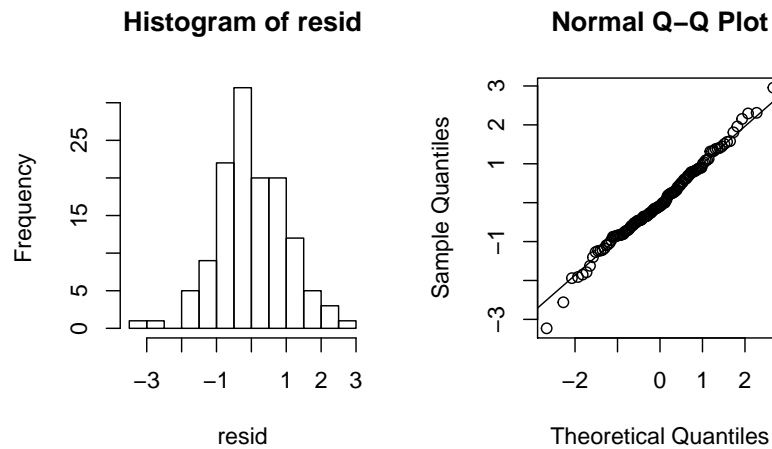


The values at all lags beyond lag 1 are zero, providing evidence that the residuals may in fact be white noise.

Plots for assessing normality of the residuals are shown below:

```
old_par <- par(mfrow = c(1, 2))
hist(resid)
qqnorm(resid)
```

```
qqline(resid)
par(old_par)
```



The normal Q-Q plot seems to deviate from linearity a bit in the tails, providing *some* evidence against normality of the residuals. The results of the Shapiro-Wilk test applied to the residuals are as follows:

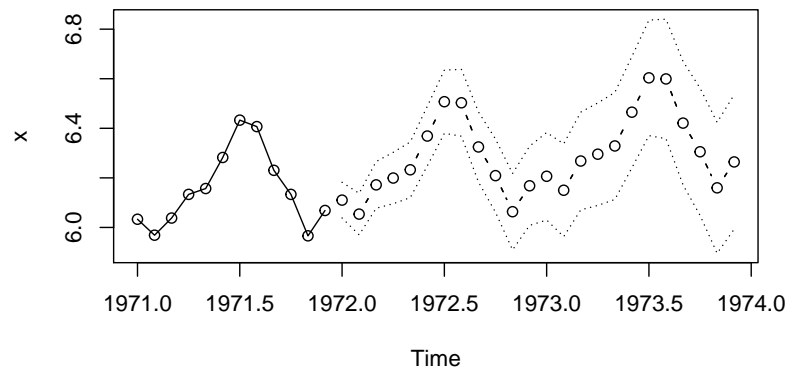
```
shapiro.test(resid)

##
##  Shapiro-Wilk normality test
##
## data:  resid
## W = 1, p-value = 0.6
```

Thus, according to Shapiro-Wilk, we fail to reject normality of the residuals at the usual levels.

- g. The following plots forecasts for the next two years after the series ends, along with pointwise forecast limits:

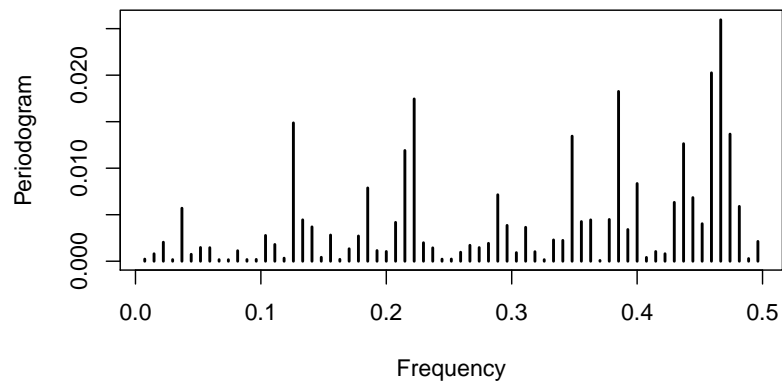
```
plot(fit, n1 = 1971, n.ahead = 24)
```



The forecasts continue the seasonal trend of the original time series. As should be expected, the pointwise confidence limits get larger as lead time increases.

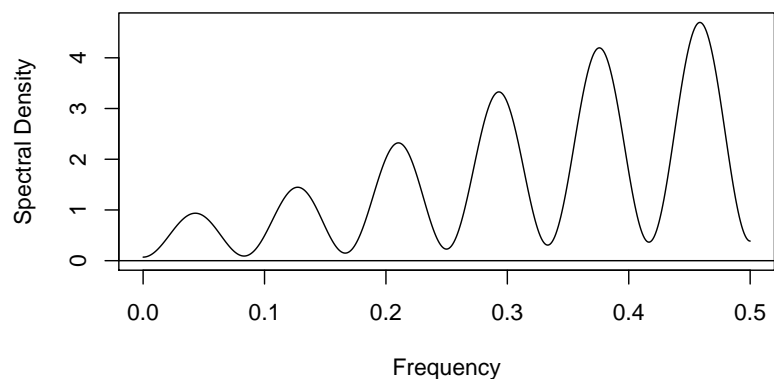
3. a. The following plots the periodogram of the series in question:

```
periodogram(seasonal_diff)
```



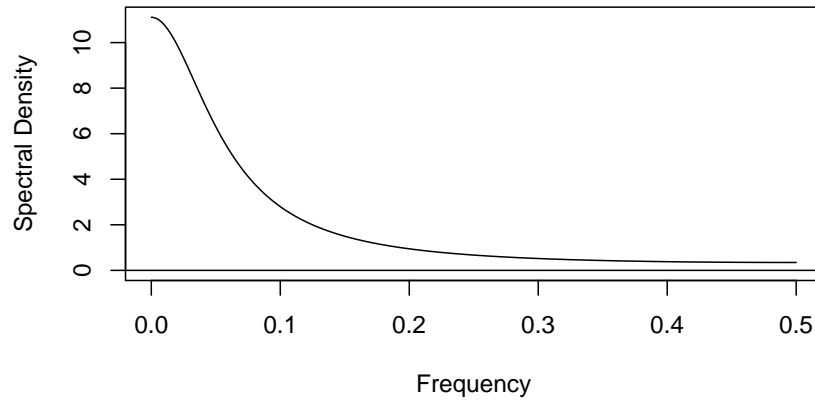
- b. Assuming the fitted model is correct, the series in question is $\text{ARMA}(0, 1) \times (0, 1)_{12}$. Using the coefficient estimates from the fitted model, the estimated spectral density of the series is as follows:

```
ma1 <- coef(fit)[["ma1"]]
sma1 <- coef(fit)[["sma1"]]
ARMAspec(model = list(ma = ma1,
                      seasonal = list(sma = sma1, period = 12)))
```



- c. The plots from part (a) and part (b) are both very similar, suggesting that the series in question does in fact follow an $\text{ARMA}(0, 1) \times (0, 1)_{12}$ with coefficients approximately equal to the estimates from the fitted model.
4. (Cryer & Chan, Exercise 13.13) The theoretical spectral density of an $\text{AR}(1)$ model with $\phi = 0.7$ is as follows:

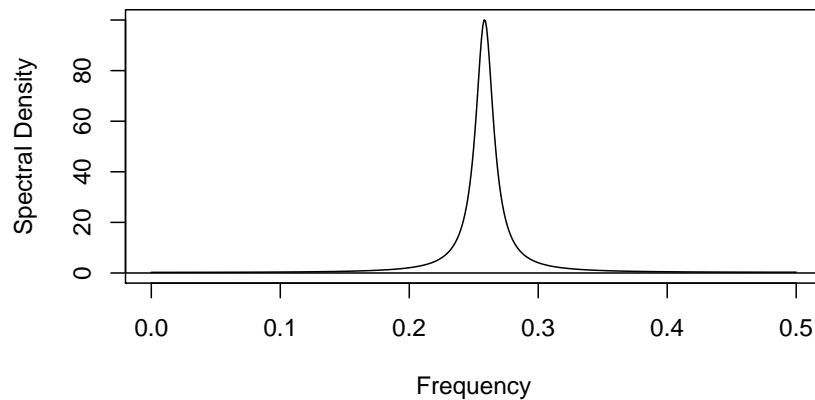
```
ARMAspec(model = list(ar = 0.7))
```



Thus the spectral representation consists primarily of low-frequency terms. Thus we would expect the series to consistently hang above or below its mean for a number of consecutive lags.

5. (Cryer & Chan, Exercise 13.17) The theoretical spectral density of an AR(2) model with $\phi_1 = -0.1$ and $\phi_2 = -0.9$ is as follows:

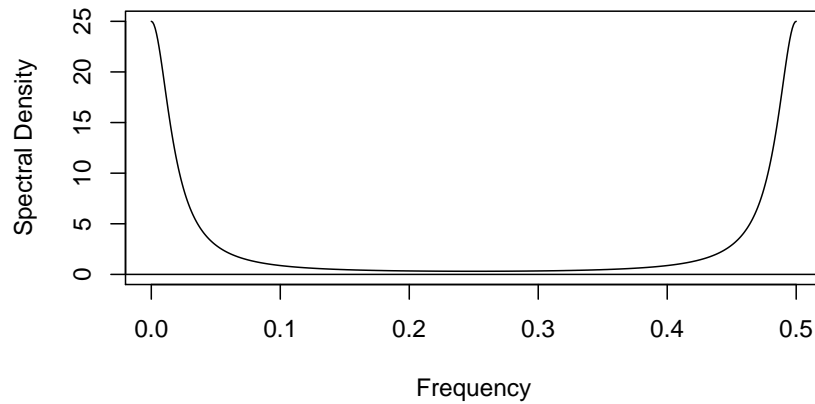
```
ARMAspec(model = list(ar = c(-0.1, -0.9)))
```



Most of the frequencies in the spectral representation of the process are concentrated around 0.25. Thus we expect the process to exhibit an approximately sinusoidal behavior with an approximate frequency of 0.25.

6. (Cryer & Chan, Exercise 13.21) The theoretical spectral density of an AR(2) model with $\phi_1 = 0$ and $\phi_2 = 0.8$ is as follows:

```
ARMAspec(model = list(ar = c(0, 0.8)))
```

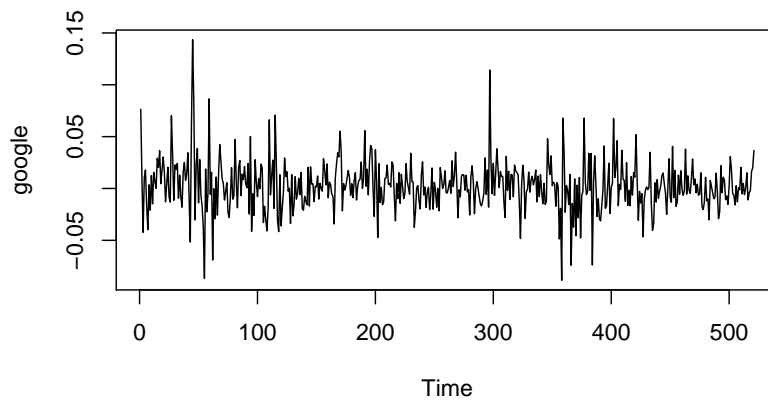


Thus the spectral representation of the process consists mainly of extremely low and extremely high frequencies. Thus we should expect the process to occasionally hang above or below its mean value, and occasionally make giant leaps across it.

7. (Cryer & Chan, Exercise 12.9)

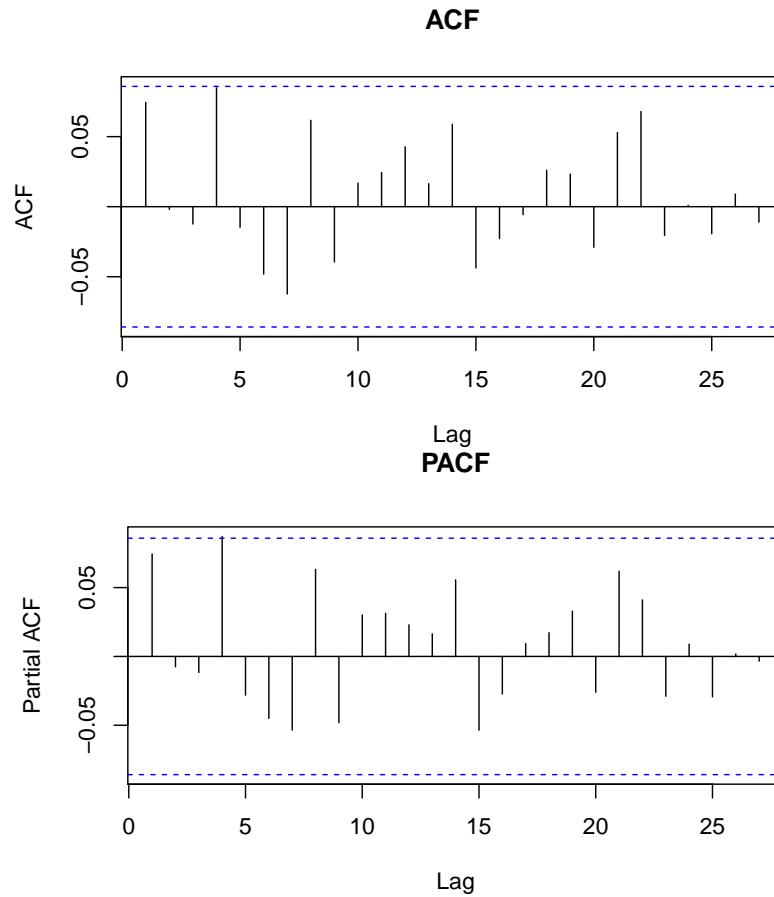
a. The time series is displayed below:

```
data(google)
plot(google, type = "l")
```



The series appears to be approximately stationary. The following are plots of the sample ACF and PACF of the series, respectively:

```
acf(google, main = "ACF")
pacf(google, main = "PACF")
```

These plots suggest that there is no significant autocorrelation in the series.

- b. The mean of the series is approximately 0.003, with approximate standard error 0.001, and thus the mean of the series is statistically different from zero.
- d. The following is the sample EACF for the squared returns:

```
eacf(google^2)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x o o o o o o o x o o o x
## 1 x o o o o o o o o x o o o x
## 2 x o o o o o o o o x o o o x
## 3 x x x o o o o o o x o o o x
## 4 x x x o o o o o o o o o o
## 5 x x x o o o o o o o o o o
## 6 x x x x o o o o o o o o o
## 7 o x x o o x o o o o o o o o
```

This suggests that a GARCH(1, 1) model may be appropriate. The following fits this model to the series of squared returns:

```
fit <- garch(google, order = c(1, 1))
```

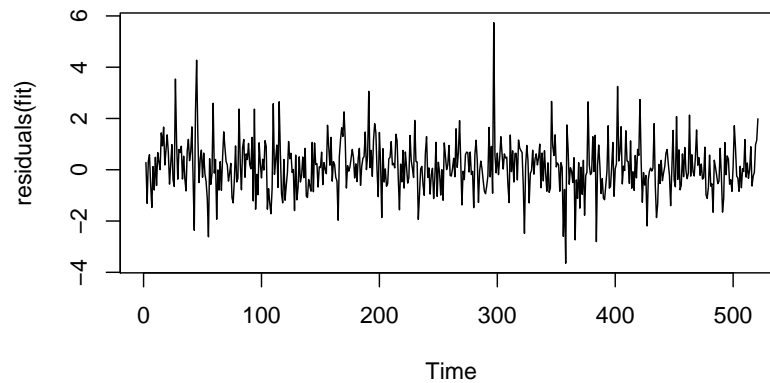
The model summary is as follows:

```
summary(fit)
```

```
##
## Call:
## garch(x = google, order = c(1, 1))
##
## Model:
## GARCH(1,1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6460 -0.4649  0.0823  0.6538  5.7394
##
## Coefficient(s):
##      Estimate Std. Error  t value Pr(>|t|)
## a0  5.06e-05    1.23e-05    4.11  4.0e-05 ***
## a1  1.26e-01    2.14e-02    5.92  3.2e-09 ***
## b1  7.87e-01    3.58e-02   21.98 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##  Jarque Bera Test
##
## data:  Residuals
## X-squared = 224, df = 2, p-value <2e-16
##
##
##  Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 7e-04, df = 1, p-value = 1
```

The following is a plot of the standardized residuals:

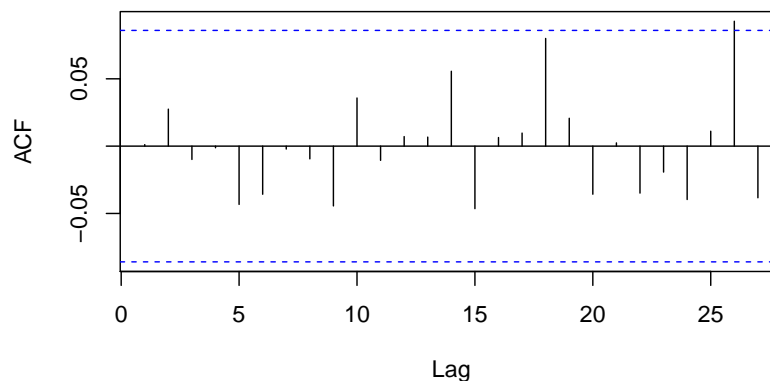
```
plot(residuals(fit))
```



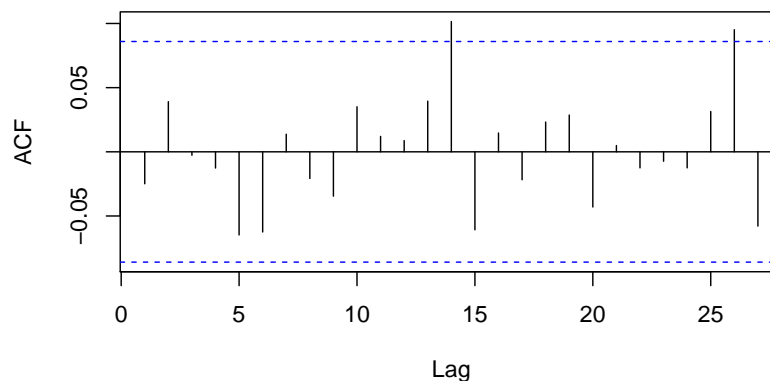
The residuals look very much like white noise. The following plots display the sample ACF for the squares and absolute values of the residuals, respectively:

```
acf(residuals(fit)^2, na.action = na.omit,
    main = "Squared residuals")
acf(abs(residuals(fit)), na.action = na.omit,
    main = "Absolute residuals")
```

Squared residuals

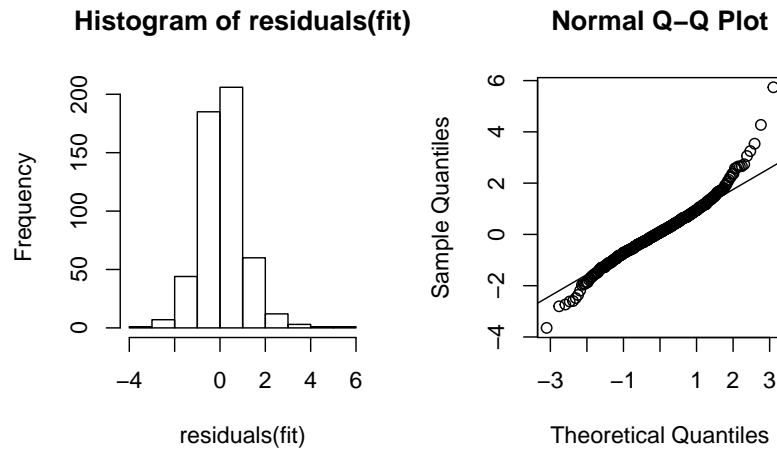


Absolute residuals



These plots together suggest that the residuals are independent. Finally, we assess normality of the residuals using the following plots:

```
old_par <- par(mfrow = c(1, 2))
hist(residuals(fit))
qqnorm(residuals(fit))
qqline(residuals(fit))
par(old_par)
```



The residuals appear to deviate from normality, as indicated by the non-linearity in the tails of the normal Q-Q plot. In addition, the Shapiro-Wilk test yields the following when applied to the residuals:

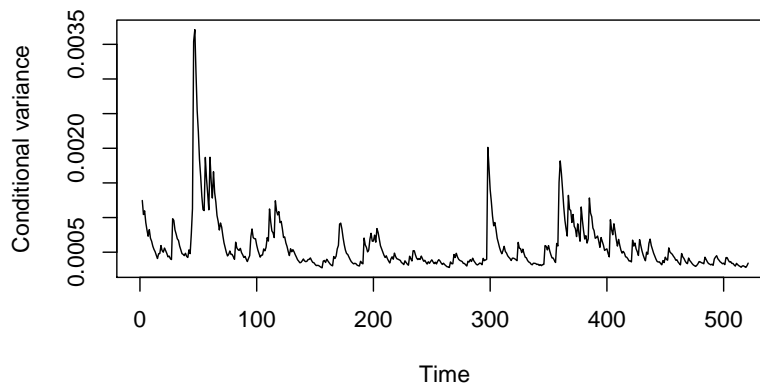
```
shapiro.test(residuals(fit))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(fit)
## W = 1, p-value = 3e-09
```

Thus we reject the null hypothesis of normality of the residuals at the usual levels.

e. The following is a plot of the conditional variances estimated by the model:

```
plot(fitted(fit)[,1]^2, type = "l", ylab = "Conditional variance")
```



Here we see several large spikes indicating periods of high-volatility in the returns. In addition, the estimated conditional variance remains relatively high for a modest time period after the third large spike (i.e., the time period roughly after day 360).

- f. The normal Q-Q plot of the standardized residuals was computed in part (d). The non-normality of the residuals may result in less reliable confidence intervals for the model coefficients.
- g. From the summary of the model fit, we have the estimate $\hat{\beta} \approx 0.787$ of β , with standard error $SE \approx 0.0358$. Thus an approximate 95% confidence interval for β is $\hat{\beta} \pm 2SE$, or approximately $(0.715, 0.859)$.
- h. The stationary mean is zero, which is very close to the sample mean of 0.003. The stationary variance is approximately

$$\frac{\hat{\omega}}{1 - \hat{\alpha} - \hat{\beta}} \approx 5.82 \times 10^{-4},$$

where $\hat{\omega}$, $\hat{\alpha}$, and $\hat{\beta}$ are the coefficient estimates from the model fit. This is very close to the sample variance of 5.694×10^{-4} .