APPLIED TIME SERIES - 36-728 - SPRING 2011 Mini 2

Assignment 3 - SOLUTIONS

1 Problem 5.1

(a)

This looks like an ARMA(2,1).
The characteristic polynomial is $\phi = 1 - x + 0.25x^2$.
To find the roots, we can either factorize the polynomial by hand or use the polyroot function in R (here, polyroot(z=c(1,-1,0.25)) for example.
We find that there is a unique root at $x = 2$, and since $|2| > 1$, this process is stationary.
So, it is an ARMA(2,1) with $\phi_1 = 1, \phi_2 = -0.25$ and $\theta_1 = 0.1$.

Note: To check stationarity, we can also use the conditions of Equation (4.3.11), page 72, because this is an AR model of order 2. Here $\phi_1 + \phi_2 = 0.75 < 1$, $\phi_2 - \phi_1 = 1.25 < 1$, and $|\phi_2| = 0.25 < 1$, so the process is a stationary. But this technique is not as general, so we recommend to practice using the characteristic polynomial instead.

(b)

This looks like an AR(2) model, with $\phi_1 = 2$ and $\phi_2 = -1$.
However, the characteristic polynomial is $1 - 2x + x^2$, which has root $x = 1$, so this process is not stationary. (Alternately, $\phi_1 + \phi_2 = 1$, not strictly less than 1, so the process is not stationary according to the conditions for AR(2) models on page 72.)
Consider then the differences:

$$\begin{aligned} Y_t - Y_{t-1} &= 2Y_{t-1} - Y_{t-2} + e_t - Y_{t-1} \\ &= (Y_{t-1} - Y_{t-2}) + e_t \end{aligned}$$

This looks like an AR(1) model. But then, the characteristic polynomial would be $\phi(x) = 1 - x$, with root $x = 1$, so it's not a stationary process.
We thus look at the second differences:

$$\begin{aligned} (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) &= Y_{t-1} - Y_{t-2} + e_t - (Y_{t-1} - Y_{t-2}) \\ &= e_t \end{aligned}$$

which is just white noise.
So, Y_t is an IMA(2,0) model, or ARIMA(0,2,0).

(c)

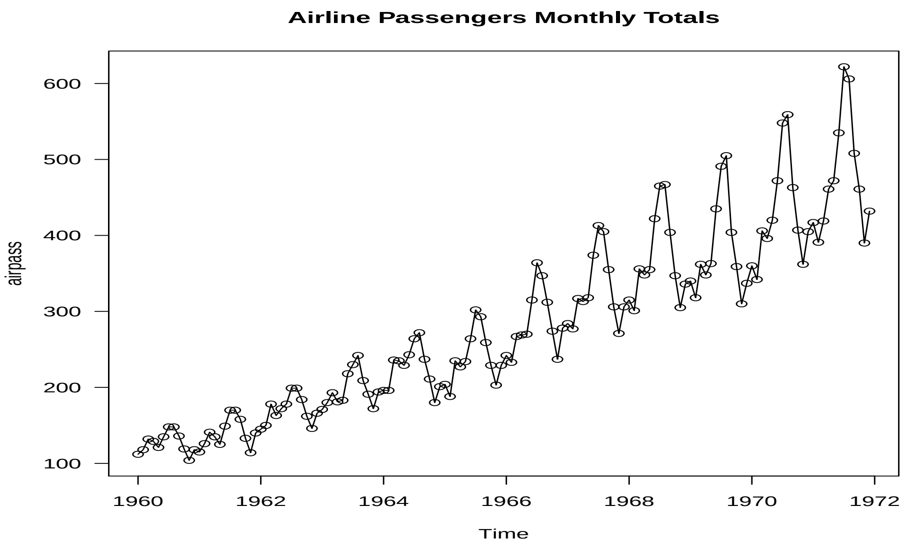
The characteristic polynomial for the AR part is $\phi(x) = 1 - 0.5x + 0.5x^2$.
The roots of this polynomial are $0.5 + 1.322876i$ and $0.5 - 1.322876i$.
Both have modulus $\sqrt{0.5^2 + 1.322876^2} = \sqrt{2}$, which is greater than 1.
Hence, this is a stationary process.
So, it is an ARMA(2,2) model with $\phi_1 = 0.5, \phi_2 = 0.5, \theta_1 = 0.5$, and $\theta_2 = 0.25$.

The answer above gives full credit, but technically we also need to verify that the MA part of the process is invertible.
The MA characteristic polynomial is $1 + 0.5x - 0.25x^2$.
It has root $-1.236068 - 0i$ and $3.236068 + 0i$, which both have modulus greater than 1, so the process indeed is invertible.

2 Problem 5.13

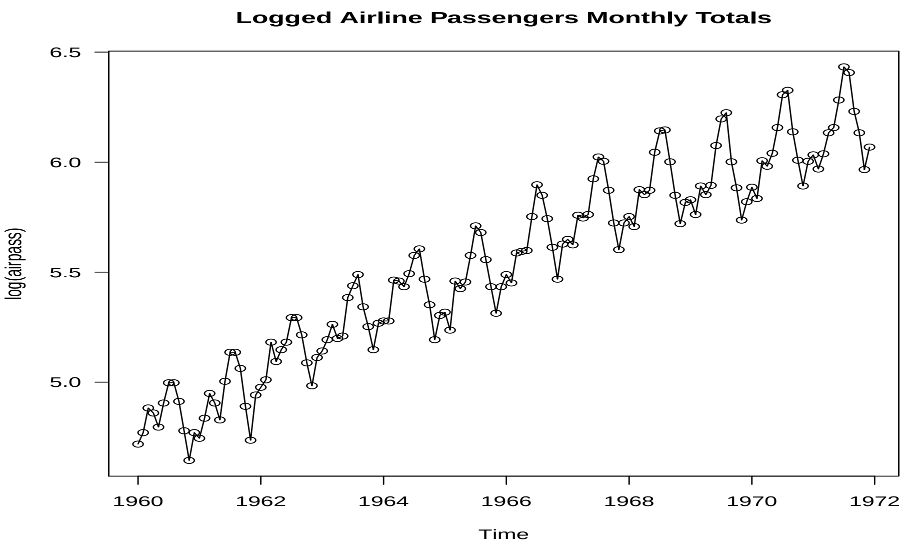
(a)

There is a clear general increasing trend in the monthly number of airline passengers. We observe as well a seasonal trend (more travel in summer). We also see an increased variation in the later years. Note that here, $\sigma_t^2 = \text{Var}(Y_t)$ is some increasing function g of $\mu_t = E(Y_t)$. Perhaps, $\sigma_t^2 \propto \mu_t$ or $\sigma_t^2 \propto \mu_t^2$, which would indicate the need for a square root or log transformation.

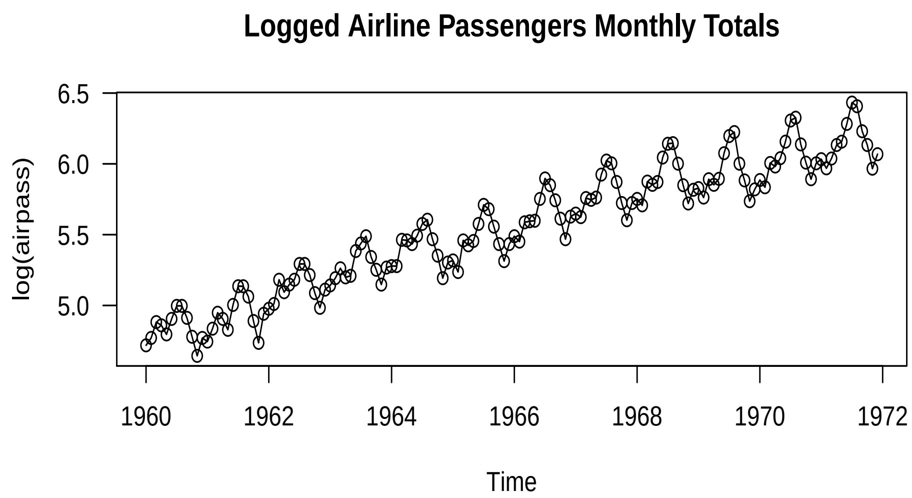


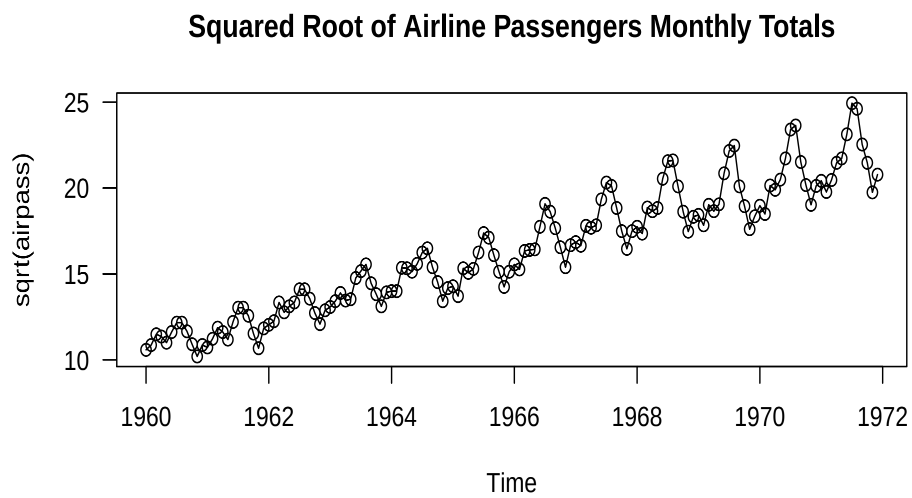
(b)

The log transformation makes the variation similar for all of the time series.



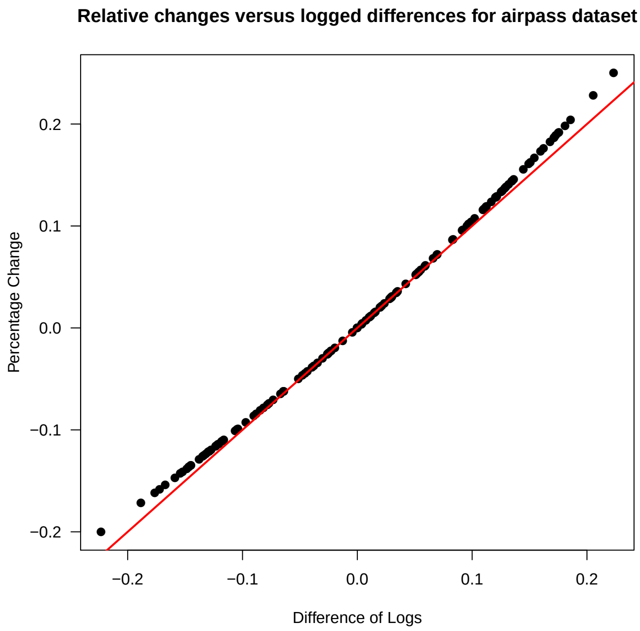
[The following is not graded.]
As we observed in part (a), a square root transformation might also be appropriate in this case. The figure below compares the effect of the log transformation and the square root transformation on the raw data. The log transformation indeed appears to stabilize the variance better.





(c)

There is a very strong linear relationship between the two quantities (we can check that $R^2 = 0.999$). The values are actually almost identical in most cases, as indicated by the identity line. Small departures from identity occur mostly at the extreme: percentage changes are larger than difference of logs for both very small and very large differences of logs. This is to be expected since the expansion $\log(1 + x) \approx x$ is best when $x \to 0$ (x is the percent change our case).

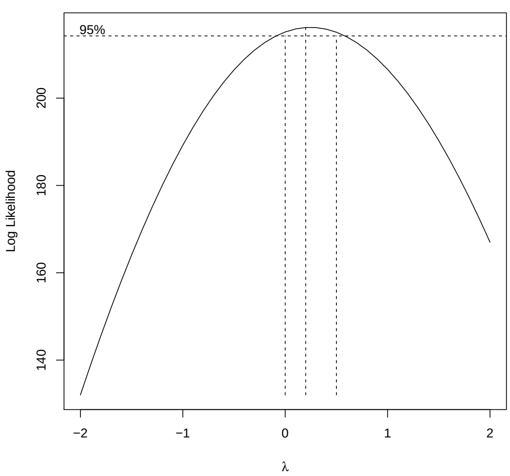


3 Problem 5.14

(a)

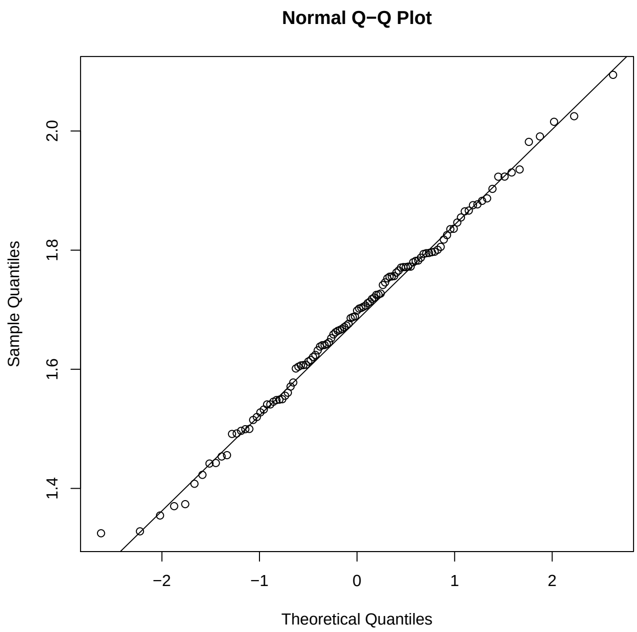
Simply calling BoxCox.ar on the dataset gives an MLE of 0.2, with a 95% C.I. from 0 to 0.5. We will use 0.2 for the remainder of this problem, but any value in the confidence interval will receive full credit.

Note that by default the BoxCox.ar function uses a grid with increments 0.1 to calculate the MLE. To get a more precise value, you can change the 'lambda' argument to the function call. We find that way that the MLE for this problem is closer to 0.25.



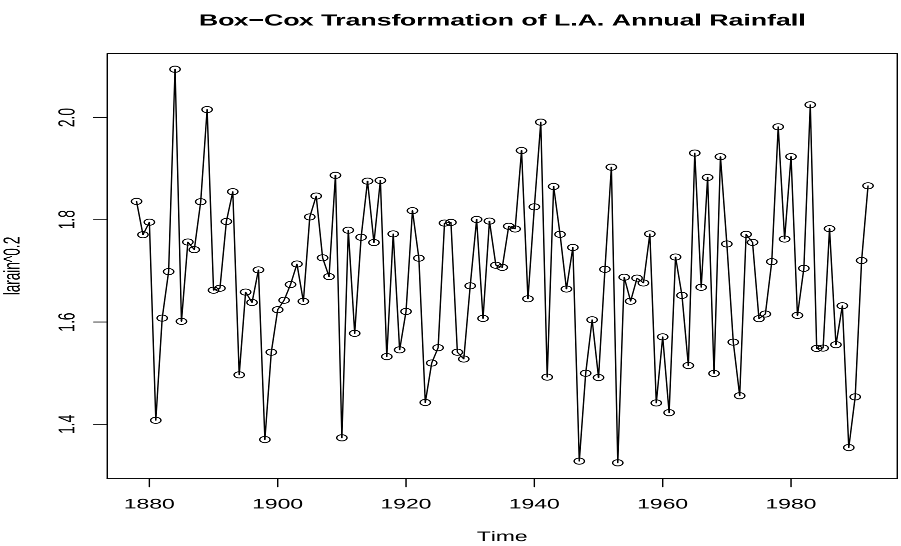
(b)

Comparing the QQplot for the transformed data (below) to the one of the raw data (textbook exhibit 3.17, p.50), we see that the Box-Cox transformation indeed made the data more normal.



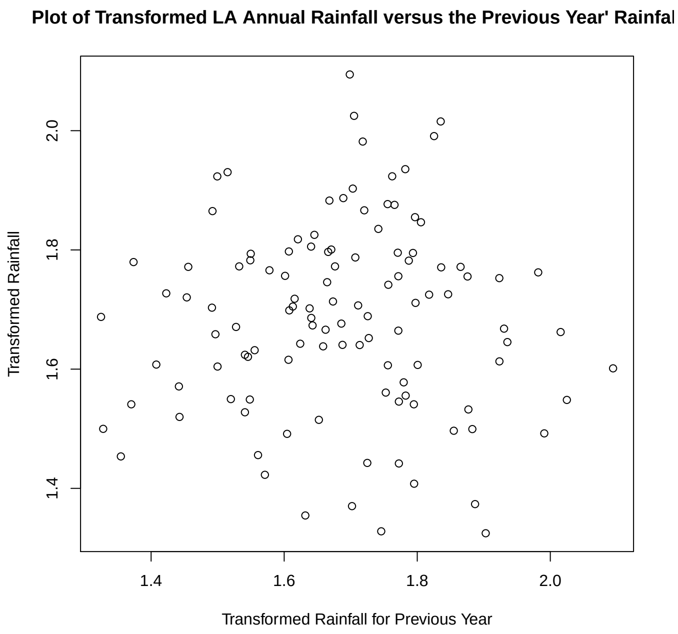
(c)

This transformed series could now be considered as normal white noise with a nonzero mean.



(d)

We still observe no relationship between the values for a year and that of the previous year. Transforming the data can modify the form the relationship between the variables, but it cannot induce correlation where none was present.

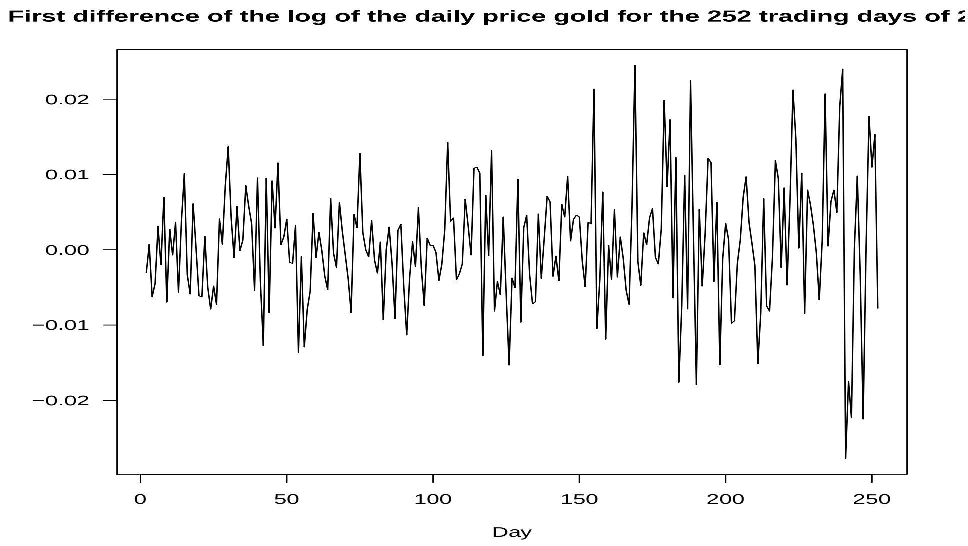


4 Problem 5.16

(a)

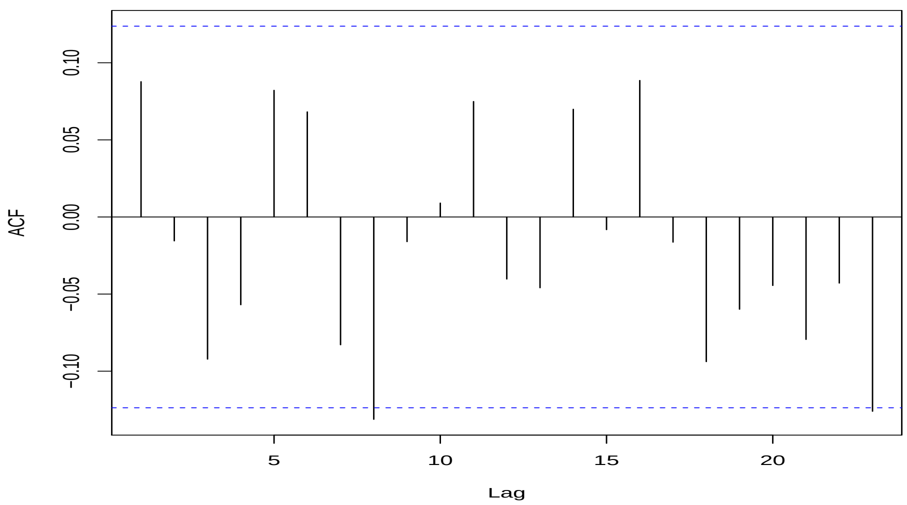
The daily price of gold in the first half of the year is relatively constant, but then there is a clear increase in the second part of the year.

(b)



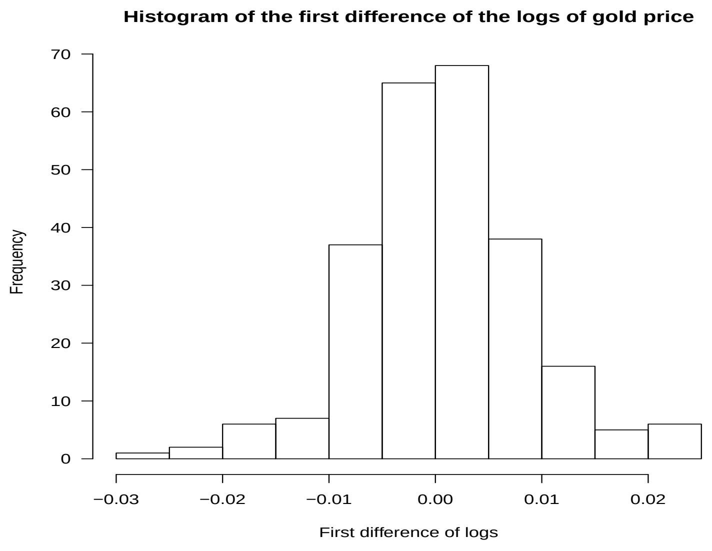
The series now almost looks stationary, except that the variance appears to be increasing with time.

(c)



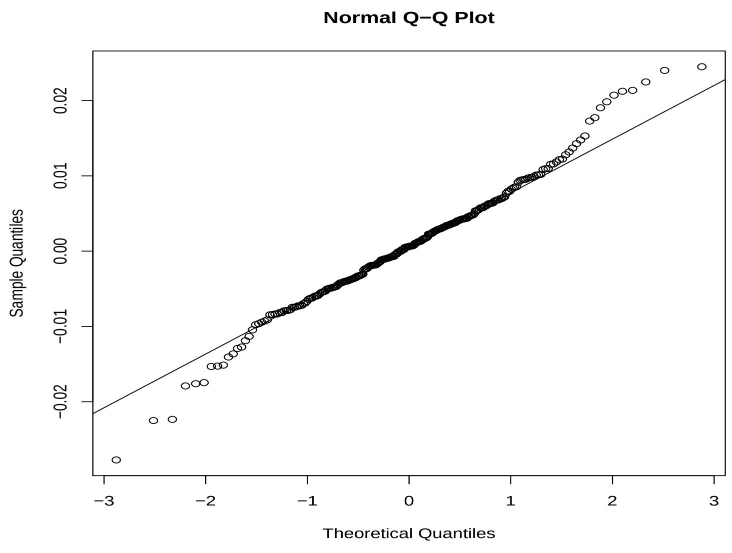
We notice that none of the lags show very significant ACF. Moreover, there does not seem to be any pattern in the ACF for different lag times. We thus conclude that the differences of the logarithms of gold prices display the autocorrelation structure of white noise. Therefore, the logarithms of gold prices could be considered as a random walk.

(d)



The distribution of the difference of logs is unimodal and reasonably symmetric, except for larger tails for large values than for small values. Overall, we might think that this distribution looks reasonably like a normal distribution.

(e)

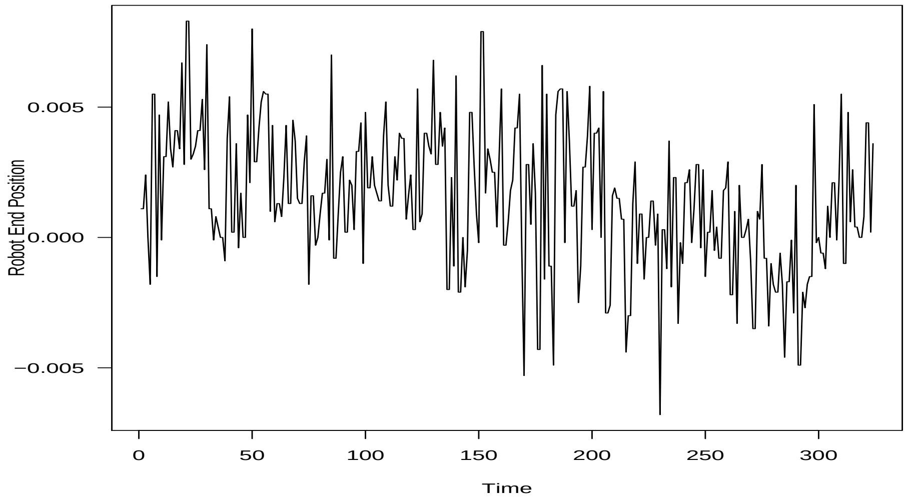


The Q-Q plot indicates that the distribution deviates from normality. In particular, the tails are lighter than those of a normal distribution. Note that this plot is much better than an histogram to verify normality; histograms should not be used for that purpose.

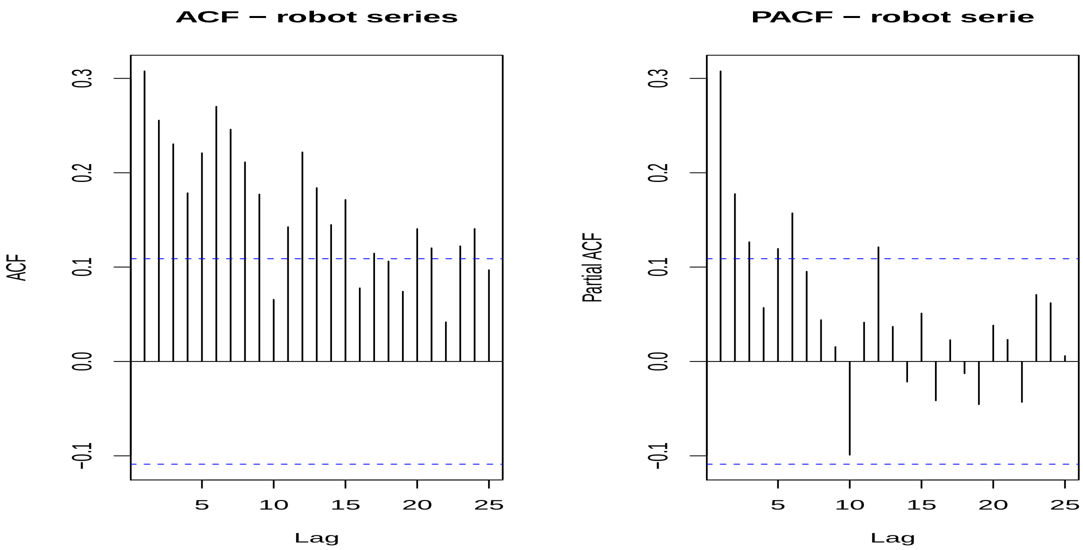
5 Problem 6.36

(a)

It is not completely clear from this plot alone whether the series is stationary. We might want to try a stationary model, but there is also enough drift that we might also suspect nonstationarity.



(b)



These plots are are not especially definitive. The ACF shows a decay that could be consistent with an AR model, but the decay is alarmingly slow, which may suggest non-stationarity. The pacf suggests possibly an AR(3) model for the series.

(c)

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

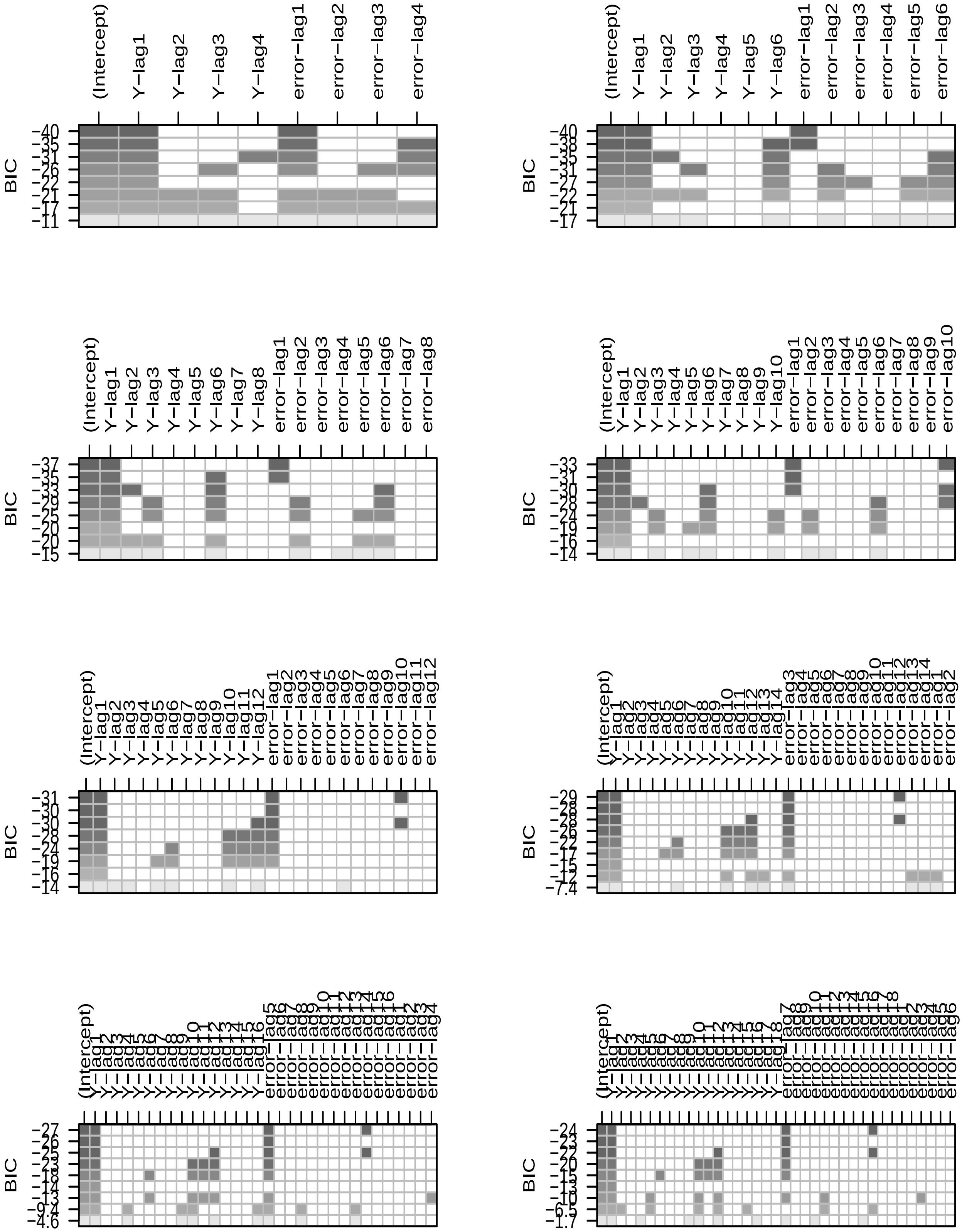
The EACF suggests an ARMA(1,1) model.

(d)

Unlike the hare data discussed in class, the sample size is large for this dataset, so it will be easier to search for a subset ARMA model within a large family of ARMA models. So, I ran the armasubsets function for several MA and AR orders, as follows:

```
par(mfrow=c(3,3))
plot(armasubsets(robot,nma=4,nar=4))
plot(armasubsets(robot,nma=6,nar=6))
plot(armasubsets(robot,nma=8,nar=8))
plot(armasubsets(robot,nma=10,nar=10))
plot(armasubsets(robot,nma=12,nar=12))
plot(armasubsets(robot,nma=14,nar=14))
plot(armasubsets(robot,nma=16,nar=16))
plot(armasubsets(robot,nma=18,nar=18))
```

The output can be found on the next page. We see that depending on the size of that ARMA family of models to be fitted, the best models are different. But the AR(1) model with some subsets of MA terms appear quite consistently. Possible models are subset ARMA(12,9, 7 or 5) with ϕ_2 to ϕ_9 set to zero. I don't see much scientific basis for such models, but would still retain them as potential models in case I do not find anything better.


BIC
(Intercept)
Y−lag1
Y−lag2
Y−lag3
Y−lag4
error−lag1
error−lag2
error−lag3
error−lag4
BIC
(Intercept)
Y−lag1
Y−lag2
Y−lag3
Y−lag4
Y−lag5
Y−lag6
error−lag1
error−lag2
error−lag3
error−lag4
error−lag5
error−lag6

6 Problem 6.37

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

This EACF indeed suggests that the logarithms of the L.A. annual rainfall follow a white noise process.

7 Problem 6.38

Recall that the PACF indicated that an AR(1) model was worthy of first consideration (textbook p-135).

Now, to obtain the EACF, we must restrict the order of the AR and MA parts considered by the R function. Otherwise, we get the following error:

```
Error in AA %*% t(X) : requires numeric/complex matrix/vector arguments
In addition: Warning message:
In ar.ols(z, order.max = i, aic = FALSE, demean = FALSE, intercept = FALSE) :
  model order:  18 singularities in the computation of the projection matrix results
```

Since we can only consider models up to order 17, we look at 9 AR lags with 7 MA lags. You can also consider 7 AR lags with 9 MA lags; mot of the outcome will be identical. The EACF indeed supports an AR(1) or AR(2) model for this series.

```
AR/MA
  0 1 2 3 4 5 6 7
0 x o o o o o o o
1 o o o o o o o o
2 o o o o o o o o
3 x o o o o o o o
4 o o o o o o o o
5 x o o o o o o o
6 x o o o o o o o
7 x o o o o o o o
8 o o o o o o o o
9 o o o o o o o o
```

8 Problem 8.1

From Equation (8.1.5), page 180, we have that

$$\sqrt{\mathrm{Var}(\hat{r}_1)} \approx \sqrt{\frac{\phi^2}{n}} = \sqrt{\frac{0.5^2}{100}} = 0.05.$$

So that would expect the lag 1 sample autocorrelation of the residuals to be within 0.1. The residual autocorrelation of 0.5 is very unusual.

9 Problem 8.2

For an MA model, we can use exactly the same equations are for an AR model by replacing ϕ by θ in the equations. Thus, we reach exactly the same conclusion as in 8.1

10 Problem 8.9

We want to compare the fit of an AR(1) model and an IMA(1,1) models for the robot data.

Fit Statistics

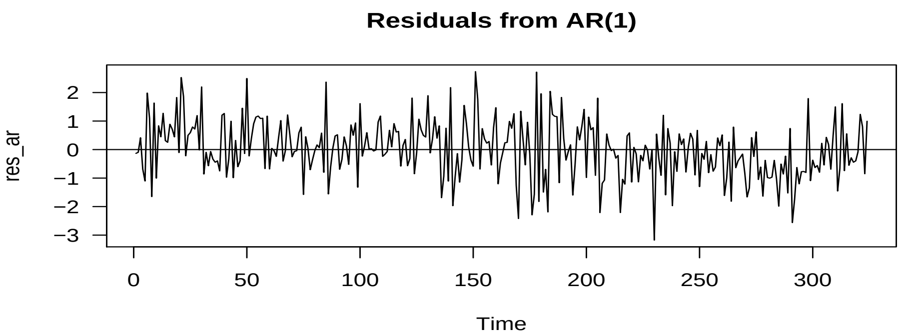
First, we fit both models and look at the model fit information returned by the function arima. We find that for both models the parameters are significantly different from zero. However, the log-likelihood and AIC values are slightly better for the IMA(1,1) model.

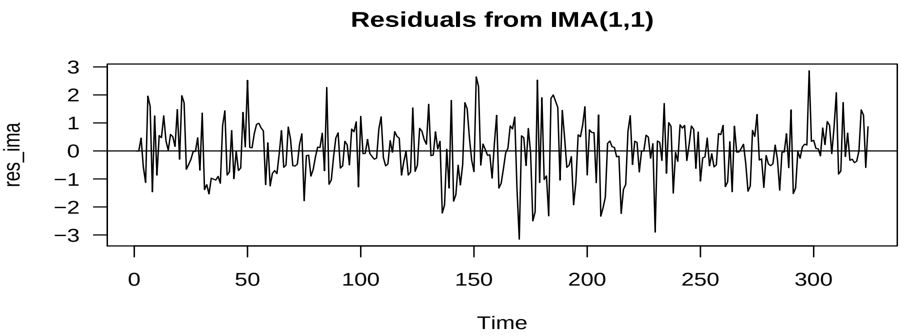
```
Call:
arima(x = robot, order = c(1, 0, 0))
Coefficients:
         ar1  intercept
      0.3074     0.0015
s.e.  0.0528     0.0002
sigma^2 estimated as 6.482e-06:  log likelihood = 1475.54,  aic = -2947.08

Call:
arima(x = robot, order = c(0, 1, 1))
Coefficients:
          ma1
      -0.8713
s.e.   0.0389
sigma^2 estimated as 6.069e-06:  log likelihood = 1480.95,  aic = -2959.9
```

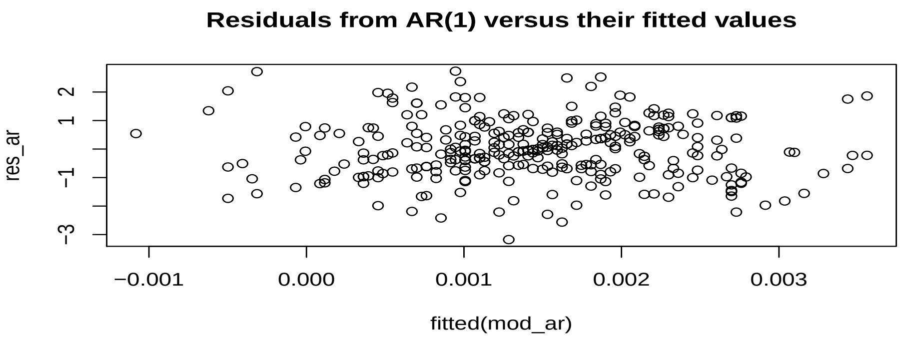
General behavior of the residuals

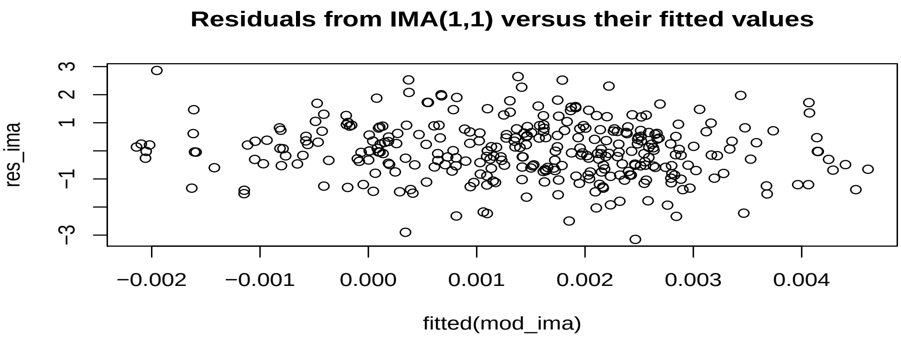
We now look at the residuals of both models. For the residuals from the AR(1) model, there seems to be more positive residuals in the first part of the series, and more negative residuals in the second half. We do not see such a drift for the IMA(1,1) model.

We also plot the residuals against the fitted values, so that we can check variance homogeneity. Once again, the IMA(1,1) seems to be a better fit than the AR(1) model.


Residuals from AR(1)
res_ar
2
1
0
−1
−2
−3
0
50
100
150
200
250
300
Time

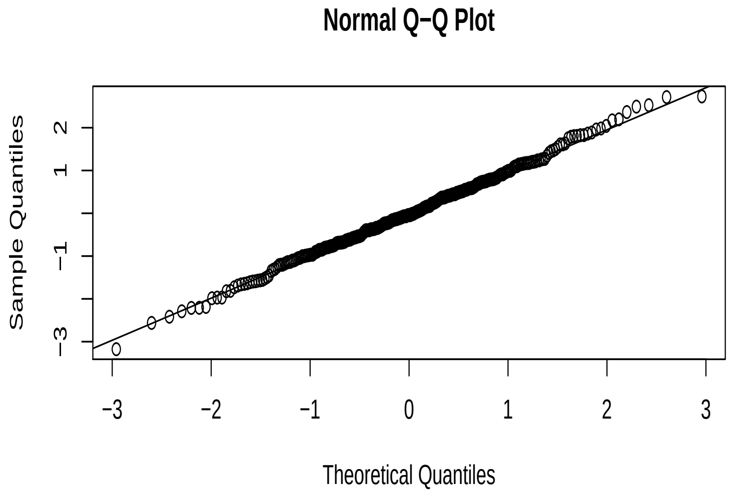

Residuals from IMA(1,1)
res_ima
3
2
1
0
−1
−2
−3
0
50
100
150
200
250
300
Time

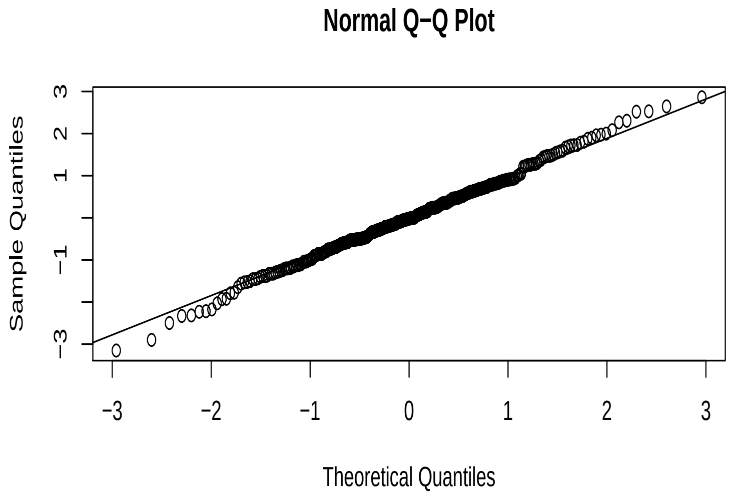

Residuals from AR(1) versus their fitted values
res_ar
2
1
−1
−3
−0.001
0.000
0.001
0.002
0.003
fitted(mod_ar)


Residuals from IMA(1,1) versus their fitted values
res_ima
3
2
1
−1
−3
−0.002
−0.001
0.000
0.001
0.002
0.003
0.004
fitted(mod_ima)

Normality of the residuals

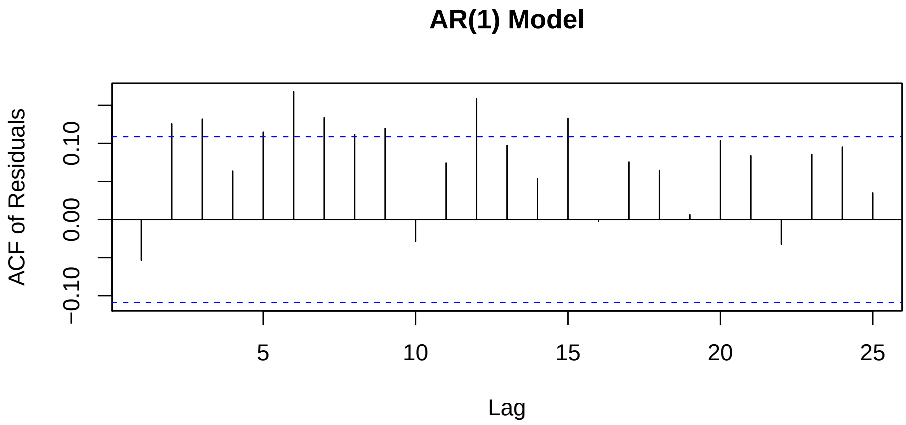
The QQ-norm plot for the AR(1) looks good, but for the IMA(1,1) model the residuals appear to deviate slightly from normality in the tails. We could test normality using the Shapiro-Wilks test, but recall that the test is sensitive to particular assumptions whereas the QQ-norm plot is not. Also, the plot suggests how the residuals might deviate from normality, and thus might suggest a way to fix the problem, for example via a transformation.

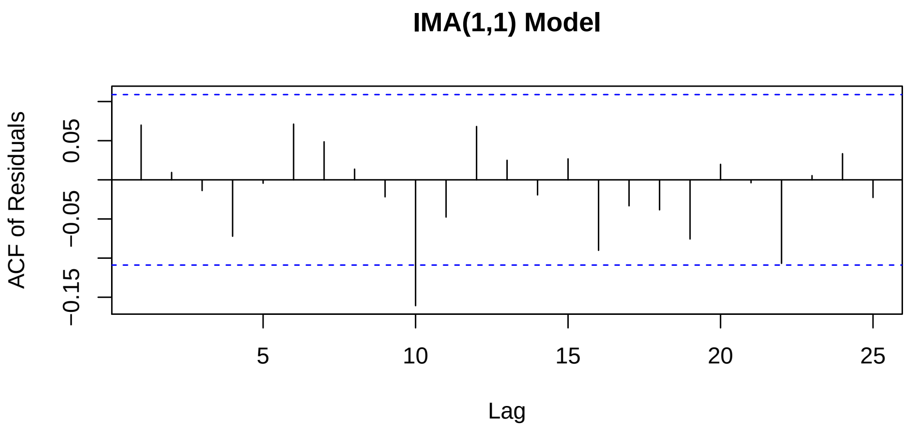




Independence of the residuals

We produce ACF plots for both series of residuals. The residuals from the AR(1) model appear to have too much autocorrelation. The residuals from the IMA(1,1) model are much less correlated with only one significant autocorrelation at lag 10. Also, the correlations at small lags are not close to the null confidence bands; otherwise we would have to check if they are significantly different from zero.





We confirm our conclusions using the Box-Ljung test:

```
Box-Ljung test - AR(1) Model

data:  residuals from  mod_ar
X-squared = 52.5123, df = 11, p-value = 2.201e-07

Box-Ljung test - IMA(1,1) Model

data:  residuals from  mod_ima
X-squared = 17.0808, df = 11, p-value = 0.1055
```

These results are also illustrated by the output of the tsdiag() function on the next page.
Finally, we also run the runs test on both series of residuals. The p-value for AR(1) model is 0.174 while the p-value for IMA(1,1) model is 0.00921. So, I would reject independence of the residuals in the IMA(1,1) model. It is however possible that this particular test may be wrong; in fact the probability that the test is wrong under the null hypothesis is 5% if the assumptions of the test are met.

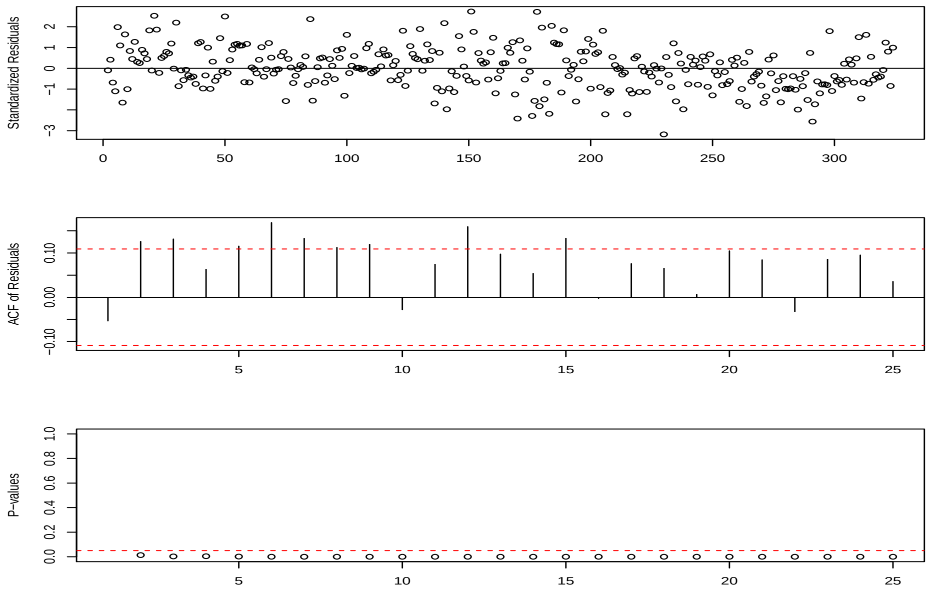


Figure 1: Diagnostics for AR(1) model

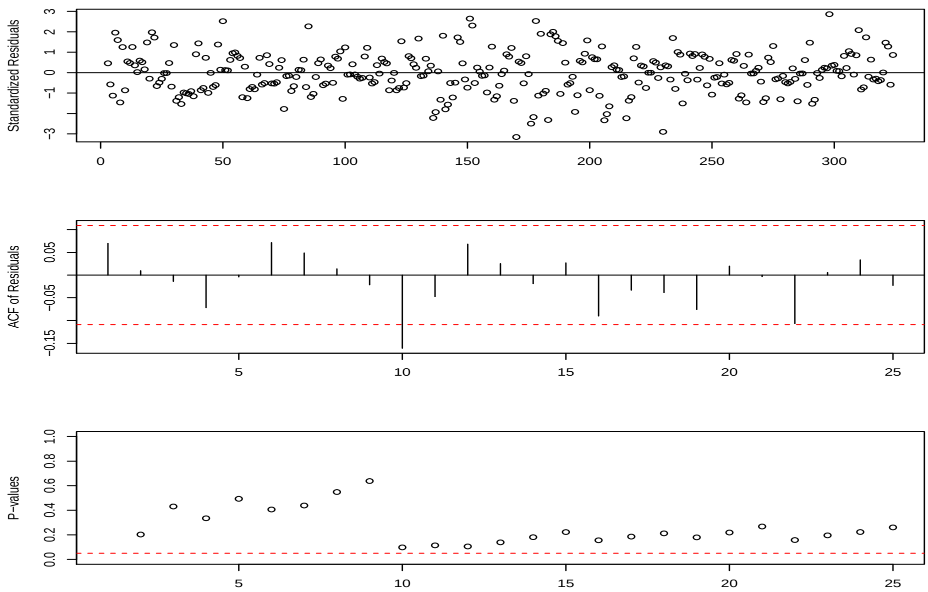


Figure 2: Diagnostics for IMA(1,1) model

Overfitting

Finally, we use overfitting to test the adequacy of the models.

First, we compare the fit of the AR(1) model to that of the AR(2) and ARMA(1,1) models:

```
Call:
arima(x = robot, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.3074     0.0015
s.e.  0.0528     0.0002

sigma^2 estimated as 6.482e-06:  log likelihood = 1475.54,  aic = -2947.08

Call:
arima(x = robot, order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.2525  0.1768     0.0015
s.e.  0.0546  0.0546     0.0002

sigma^2 estimated as 6.278e-06:  log likelihood = 1480.7,  aic = -2955.39


Call:
arima(x = robot, order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.9472  -0.8062     0.0015
s.e.  0.0309   0.0609     0.0005

sigma^2 estimated as 5.948e-06:  log likelihood = 1489.3,  aic = -2972.61
```

While $\hat{\phi}_1$ looks stable when we add another AR term, it changes a lot when we add an MA term. Moreover, both $\hat{\phi}_2$ and $\hat{\theta}_1$ differ significantly from zero. Overall, this indicates that the AR(1) is a poor fit to the data.

We now compare the fit from the IMA(1,1), that is ARIMA(0,1,1), to the ARIMA(1,1,1) and the ARIMA(0,1,2) models.

```
Call:
arima(x = robot, order = c(0, 1, 1))

Coefficients:
          ma1
      -0.8713
s.e.   0.0389

sigma^2 estimated as 6.069e-06:  log likelihood = 1480.95,  aic = -2959.9

Call:
arima(x = robot, order = c(1, 1, 1))

Coefficients:
         ar1      ma1
      0.1208  -0.9215
s.e.  0.0715   0.0429

sigma^2 estimated as 6.012e-06:  log likelihood = 1482.35,  aic = -2960.7

Call:
arima(x = robot, order = c(0, 1, 2))

Coefficients:
          ma1      ma2
      -0.8088  -0.0930
s.e.   0.0540   0.0594

sigma^2 estimated as 6.02e-06:  log likelihood = 1482.18,  aic = -2960.36
```

The estimate for ϕ_1 is similar under the three models, which indicates stability of the IMA(1,1) model. Also, neither the extra $\hat{\phi_1}$ nor $\hat{\theta_2}$ differ significantly from zero. The IMA(1,1) is thus a good fit to this series.

Conclusion

Overall, the IMA(1,1) model appears to a be a better fit to the series than the AR(1) model.