STAT 520 – Homework 3 -

1) Consider the MA(2) process, where all the $\{e_t\}$ values are independent white noise with variance σ^2_e:
$Y_t = e_t - 0.5\, e_{t-1} - 0.3\, e_{t-2}$
a) Find $\text{cov}(Y_t, Y_t) = \text{var}(Y_t)$.
b) Find $\text{cov}(Y_t, Y_{t-1})$ and, from this, find the lag-1 autocorrelation $\text{corr}(Y_t, Y_{t-1})$.
c) Find $\text{cov}(Y_t, Y_{t-2})$ and, from this, find the lag-2 autocorrelation $\text{corr}(Y_t, Y_{t-2})$.
d) Argue that $\text{cov}(Y_t, Y_{t-k}) = 0$ for all $k \geq 3$.

a) $\text{cov}(Y_t, Y_t) = \text{var}(Y_t) = \text{var}(e_t - 0.5e_{t-1} - 0.3e_{t-2})$
$= \text{var}(e_t) + 0.25\,\text{var}(e_{t-1}) + 0.09\,\text{var}(e_{t-2})$ by independence
$= \sigma_e^2 + 0.25\sigma_e^2 + 0.09\sigma_e^2 = 1.34\sigma_e^2$

b) $\text{cov}(Y_t, Y_{t-1}) = \text{cov}(e_t - 0.5e_{t-1} - 0.3e_{t-2},\ e_{t-1} - 0.5e_{t-2} - 0.3e_{t-3})$
$= 0 + 0 + 0 - 0.5\,\text{cov}(e_{t-1}, e_{t-1}) + 0 + 0 + 0$
$+ 0.15\,\text{cov}(e_{t-2}, e_{t-2}) + 0$
$= -0.5\,\text{var}(e_{t-1}) + 0.15\,\text{var}(e_{t-2})$
$= -0.35\sigma_e^2$

So $\text{corr}(Y_t, Y_{t-1}) = \dfrac{\text{cov}(Y_t, Y_{t-1})}{\sqrt{\text{var}(Y_t)\,\text{var}(Y_{t-1})}} = \dfrac{-0.35\sigma_e^2}{\sqrt{(1.34\sigma_e^2)(1.34\sigma_e^2)}}$

$= \dfrac{-0.35}{1.34} = \boxed{-0.26}$

c) $\text{cov}(Y_t, Y_{t-2}) = \text{cov}(e_t - 0.5e_{t-1} - 0.3e_{t-2},$
$e_{t-2} - 0.5e_{t-3} - 0.3e_{t-4})$
$= 0 + 0 + 0 + 0 + 0 + 0 - 0.3\,\text{var}(e_{t-2}) + 0 + 0$
$= -0.3\sigma_e^2$

$\Rightarrow \text{corr}(Y_t, Y_{t-2}) = \dfrac{-0.3\sigma_e^2}{\sqrt{(1.34\sigma_e^2)(1.34\sigma_e^2)}} = \dfrac{-0.3}{1.34} = \boxed{-0.22}$

d) For $cov(Y_t, Y_{t-k})$, there will be no overlapping terms, i.e, no noise terms with the same subscript, in the covariance expansion.
So all the covariances will be zero.

2) **[Mandatory for graduate students, extra credit for undergraduates]**
Repeat all parts of Problem 1, but for the process:
$Y_t = e_t - 1.2\, e_{t-1} + 0.7\, e_{t-2}$

a) $var(Y_t) = var(e_t) + 1.44\, var(e_{t-1}) + 0.49\, var(e_{t-2})$

$= \sigma_e^2 + 1.44\,\sigma_e^2 + 0.49\,\sigma_e^2 = 2.93\,\sigma_e^2$

b) $cov(Y_t, Y_{t-1}) = cov(e_t - 1.2e_{t-1} + 0.7e_{t-2},\ e_{t-1} - 1.2e_{t-2} + 0.7e_{t-3})$

$= 0+0+0-1.2\sigma_e^2+0+0+0-0.84\sigma_e^2+0 = -2.04\,\sigma_e^2$

$\Rightarrow corr(Y_t, Y_{t-1}) = \dfrac{-2.04\,\sigma_e^2}{\sqrt{(2.93\sigma_e^2)(2.93\sigma_e^2)}} = \dfrac{-2.04}{2.93} = -0.696$

c) $cov(Y_t, Y_{t-2}) = cov(e_t - 1.2e_{t-1} + 0.7e_{t-2},\ e_{t-2} - 1.2e_{t-3} + 0.7e_{t-4})$

$= 0+0+0+0+0+0+0.7\sigma_e^2+0+0 = 0.7\,\sigma_e^2$

$\Rightarrow corr(Y_t, Y_{t-2}) = \dfrac{0.7}{2.93} = 0.239$

d) Same answer as #1, part (d).

3) Consider the AR(1) process: $Y_t = \phi Y_{t-1} + e_t$
Show that if $|\phi| = 1$, the process cannot be stationary. [Hint: Take variances of both sides.]

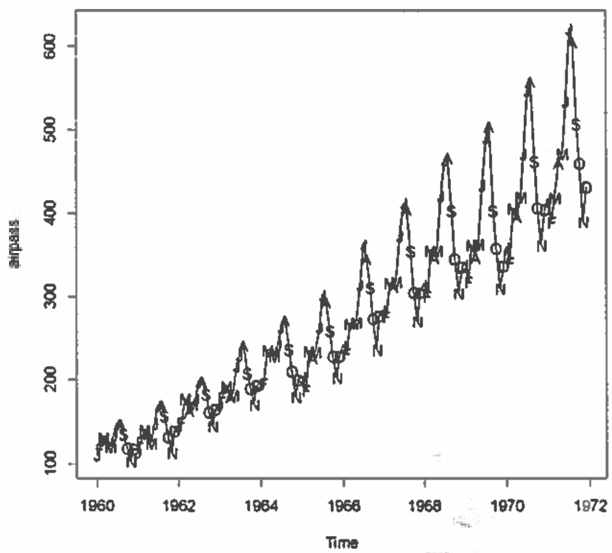
$var(Y_t) = \phi^2\, var(Y_{t-1}) + var(e_t)$ since e_t is indep of Y_{t-1}.

$\Rightarrow var(Y_t) = var(Y_{t-1}) + \sigma_e^2$ since $|\phi| = 1 \Rightarrow \phi^2 = 1$.

Since $\sigma_e^2 > 0$, then $var(Y_t) > var(Y_{t-1})$. If the series $\{Y_t\}$ were stationary, its variance would be the same at every time point.

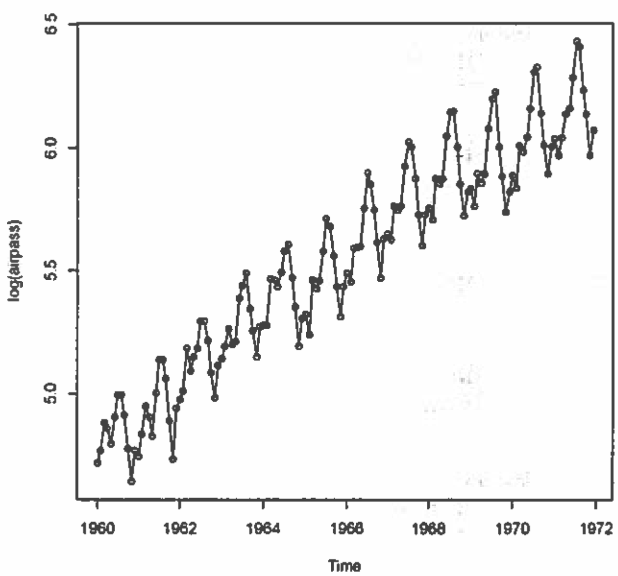
4) The monthly U.S. air passenger miles flown for January 1949 to December 1960 are in the `airpass` object in the TSA package. Type `library(TSA); data(airpass); print(airpass)` in R to see the data set.

(a) Plot the time series, using plotting symbols that allow you to check for seasonality. What basic pattern do you see from the plot?



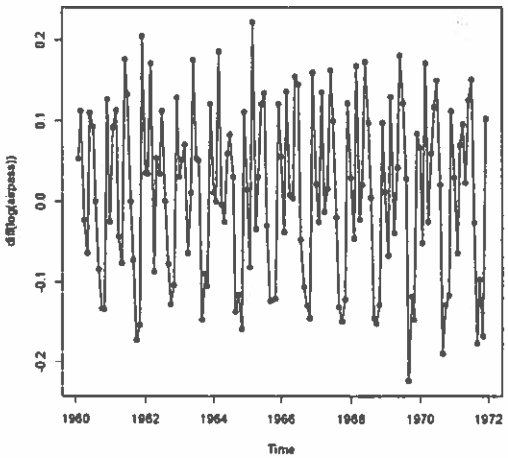
We can see a clear seasonal pattern, with air travel being most popular in the summer months, and least popular in the late fall (like November). Also, there is a clear increasing trend in air miles traveled over the period from 1960 to 1972. The variance of the measurements seems to grow over time as well.

(b) Plot the (natural) log-transformed time series. What basic pattern do you see from the plot? What effect has the log transformation had?



We see the same seasonality and increasing trend over time, but the key difference is that the variance appears more constant over time.

(c) Plot the differences of the natural logarithms. Does this plot suggest that a stationary model might be appropriate for the differences of the natural logarithms? Briefly explain.

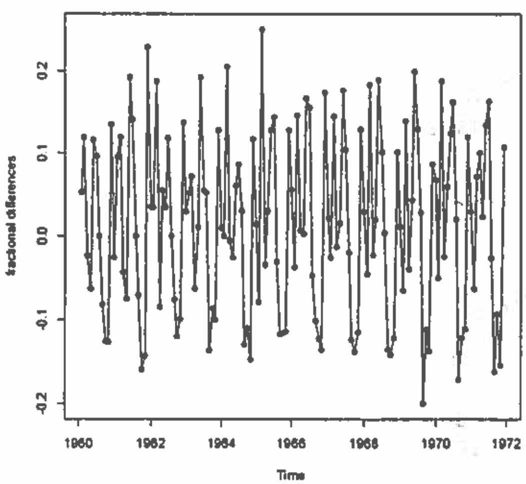


A stationary model could be appropriate for the differences of the logged data. The mean function and variance function both seem constant over time.

(d) Plot the fractional relative differences, $(Y_t - Y_{t-1})/Y_{t-1}$, which can be obtained in R with the code:

```
diff(airpass)/(zlag(airpass)[-1])
```

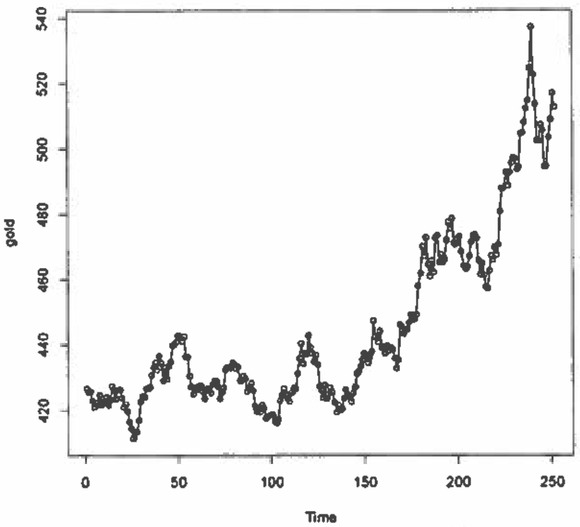
How do these values compare with the differences of the natural logarithms from part (c)?



These values look quite similar to those plotted in part (c), indicating that these transformations do similar things.

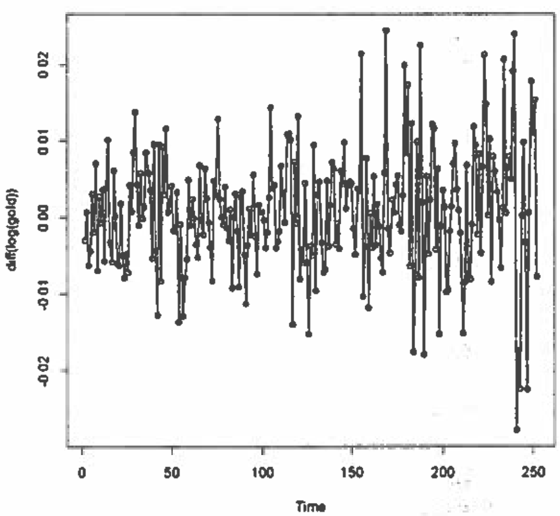
5) The daily price of gold over 252 trading days in 2005 are in the `gold` object in the `TSA` package. Type `library(TSA); data(gold); print(gold)` in R to see the data set.

(a) Plot the time series. What basic pattern do you see from the plot?



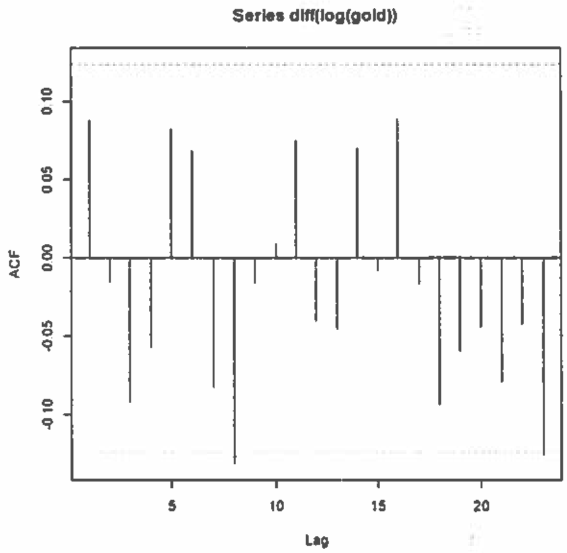
There is an overall increasing trend (perhaps nonlinear) over time.

(b) Plot the time series of the differences of the (natural) logarithms of these data. Does this plot suggest that a stationary model might be appropriate for the differences of the natural logarithms? Briefly explain.



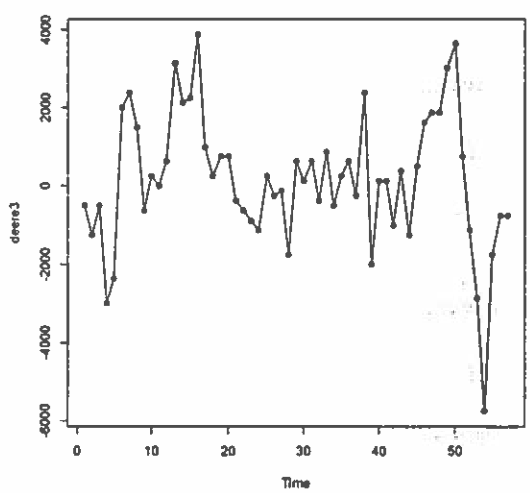
The mean function appears constant, although the variance of the differenced logged data seems to increase at the later time periods. Using a stationary model for these data is possible, but maybe not perfectly ideal.

(c) Plot the sample ACF for the differences of the logarithms of these data. Does this provide evidence that the log-transformed gold prices follow a random walk model? Why or why not?



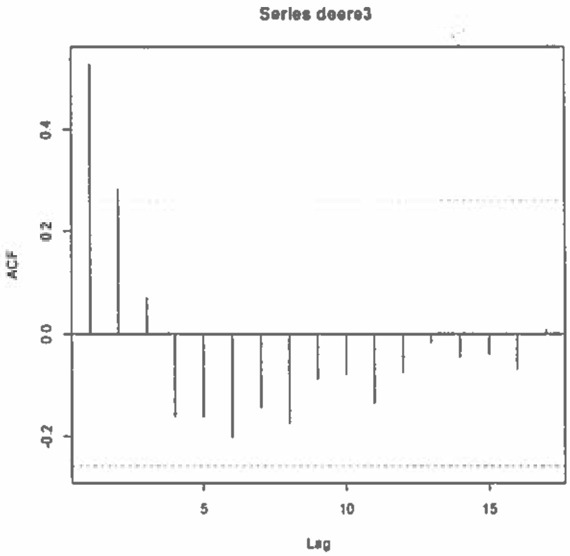
We see that there may be no real nonzero autocorrelations here. So we could reasonably model the differences of the logged data as white noise, and thus we could reasonably model the logged data themselves as a random walk.

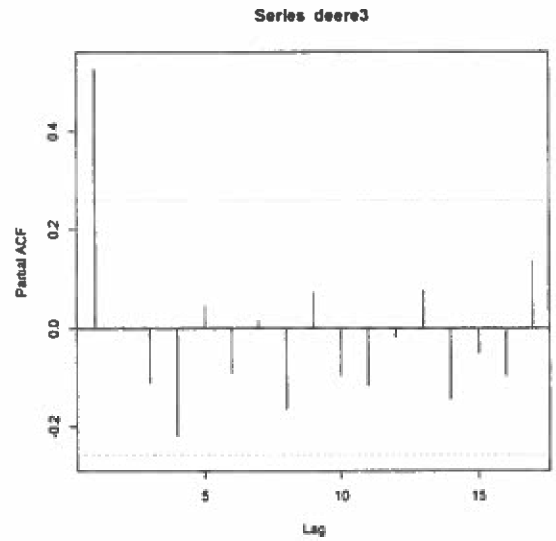
6) A data set of 57 consecutive measurements from a machine tool are in the `deere3` object in the `TSA` package. Type `library(TSA); data(deere3); print(deere3)` in R to see the data set.
(a) Plot the time series. What basic pattern do you see from the plot? Might a stationary model be appropriate for this plot?



There is not a strong increasing nor a strong decreasing trend. We note a low outlier around time 53. It is possible that a stationary model might be appropriate for these data.

(b) Using tools such as the ACF, PACF, and/or EACF, tentatively specify the type of model (AR, MA, or ARMA) as well as the order(s) of the model.



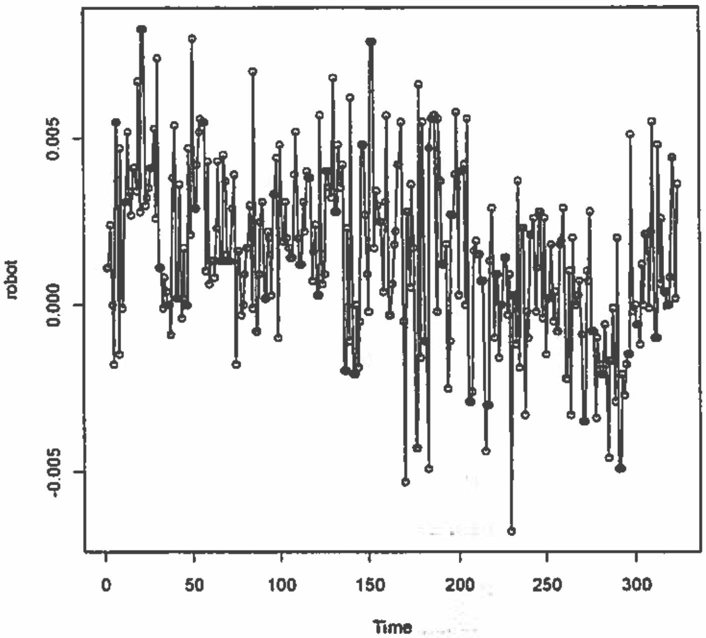



```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o o o o o o o o o  o  o  o
1 o o o o o o o o o o o  o  o  o
2 o o o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 o x o o o o o o o o o  o  o  o
5 o x o x o o o o o o o  o  o  o
6 o x o x o o o o o o o  o  o  o
7 o x x o o o o o o o o  o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

Based on the decaying, damped sine wave look of the ACF, and the fact that the PACF seems to cut off and become zero after lag 1, I would say that an AR(1) model is most appropriate. The EACF table agrees with this notion.

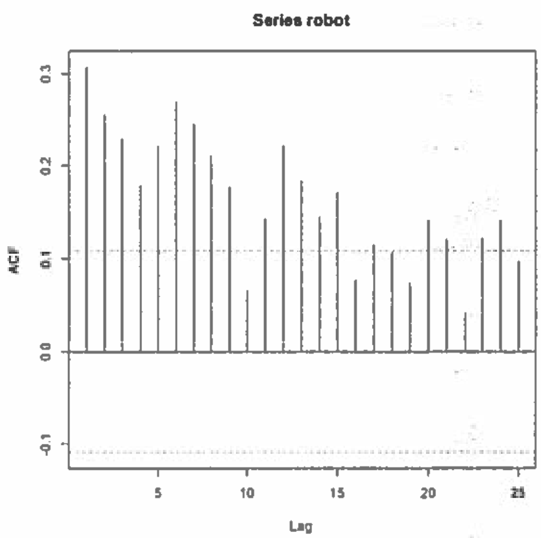
7) A data set of 324 measurements of an industrial robot's positions are in the `robot` object in the `TSA` package. Type `library(TSA); data(robot); print(robot)` in R to see the data set.

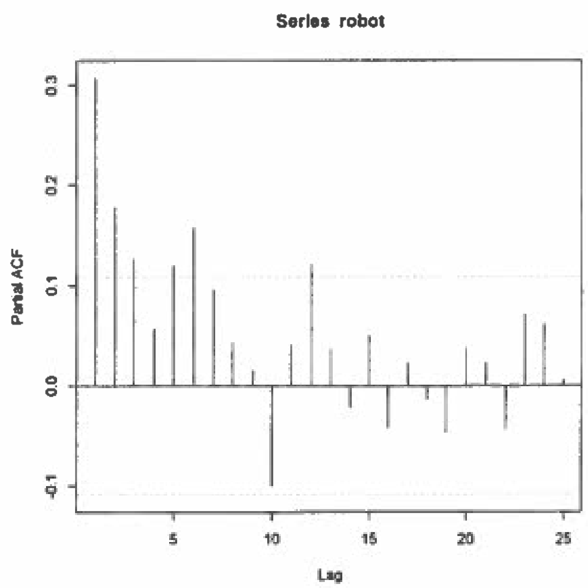
(a) Plot the time series. What basic pattern do you see from the plot? Might a stationary model be appropriate for this plot?



The time series shows pretty constant variance. The mean function might decrease slightly over time. A stationary model might be appropriate, although the slight decreasing trend is concerning.

(b) Using tools such as the ACF, PACF, and/or EACF, tentatively specify the type of model (AR, MA, or ARMA) as well as the order(s) of the model.





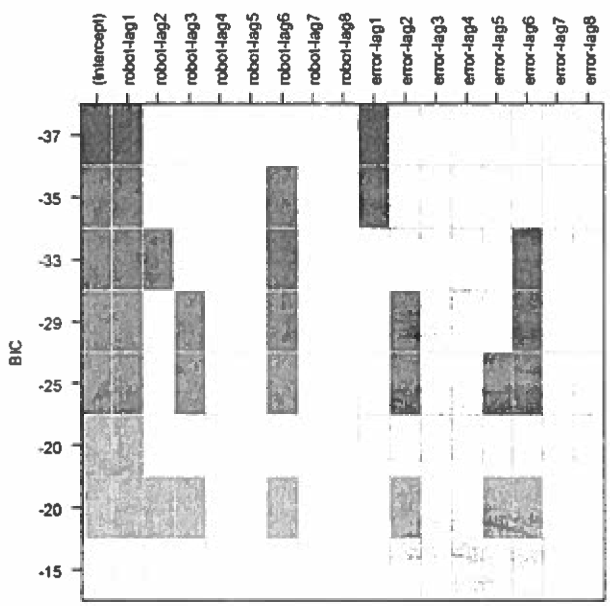
AR/MA

	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

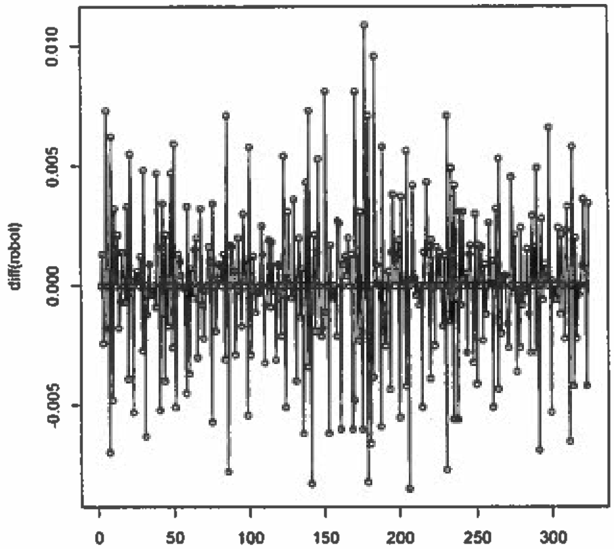
Neither the ACF nor the PACF cut off to zero after a certain number of lags; rather, they both seem to decay gradually. The EACF table indicates an ARMA(1, 1) model may be appropriate.

(c) Use the best subsets ARMA approach to specify a model. Consider up to 8 AR terms and up to 8 MA terms. Does the "best" subset ARMA model agree with the model you specified in part (b)?

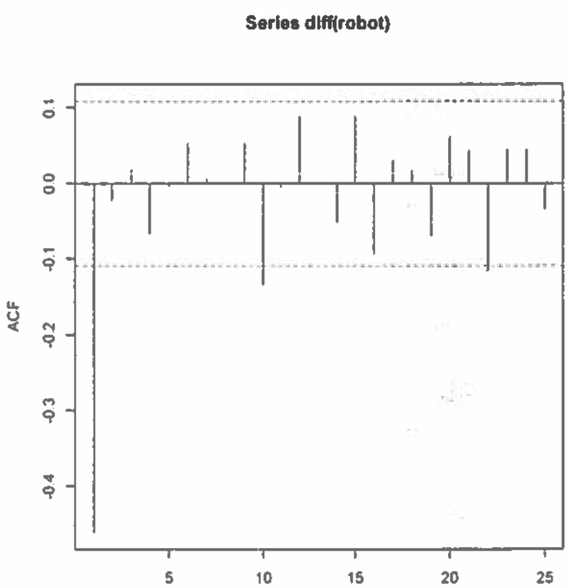


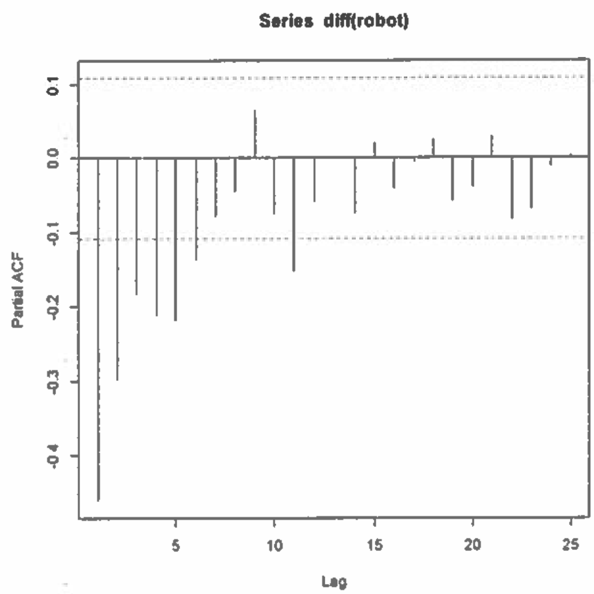
This indicates the a subset ARMA model with a Y_{t-1} term and an e_{t-1} term is the best; this indicates an ARMA(1,1) model.

(d) Repeat parts (a)-(c) on the **first differences** of the robot time series. Does this analysis suggest a particular model for the original robot data? Briefly explain.



The first differences of the robot series appear fairly stationary, with roughly constant variance and constant mean around zero.

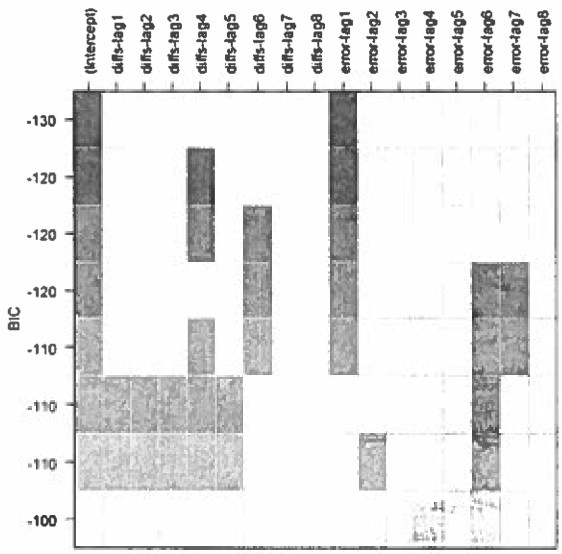




AR/MA

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

The ACF cuts off after lag 1, and the PACF gradually decays toward zero, so this indicates an MA(1) model for the differenced series. The EACF table agrees with this.



This indicates the a subset ARMA model with an e_{t-1} term is the best; this indicates an MA(1) model for the first differences. So this would indicate an ARIMA(0,1,1) model, also known as an IMA(1,1) model, for the original robot data.