STAT 520 – Homework 4 – Fall 2023

Note: All students should do all the problems for this homework set.

1) 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) Estimate the parameters of a (mean-centered) AR(1) model for this series. Use the least squares method and maximum likelihood, and report the estimated parameters from each of these methods. Comment on any similarities and differences.

Using least squares, the estimate of ϕ is 0.533 and the estimate of μ is 160.08. The estimate of the noise variance is 2100808. Using maximum likelihood, the estimate of ϕ is 0.526 and the estimate of μ is 124.35. The estimate of the noise variance is 2069354. We see that using the two methods, the estimates of ϕ and the noise variance are fairly similar, but there is a big difference in the estimate of the overall mean μ.

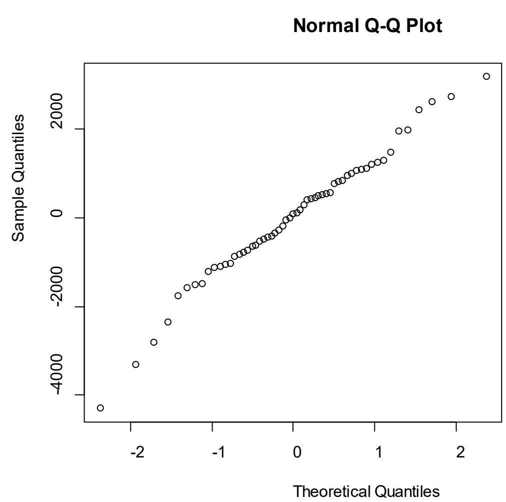
(b) Estimate the parameters of a (mean-centered) AR(2) model for this series. Use the least squares method and maximum likelihood, and report the estimated parameters from each of these methods. Comment on any similarities and differences.

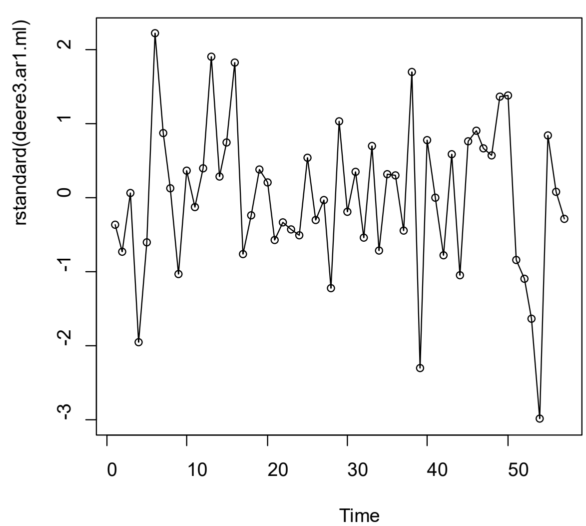
Using least squares, the estimate of ϕ_1 is 0.525, the estimate of ϕ_2 is 0.008, and the estimate of the mean μ is 201.2. The estimate of the noise variance is 2118086. Using maximum likelihood, the estimate of ϕ_1 is 0.521, the estimate of ϕ_2 is 0.008, and the estimate of the mean μ is 123.24. The estimate of the noise variance is 2069209. We see that using the two methods, the estimates of ϕ_1 and ϕ_2 and the noise variance are fairly similar, but there is a big difference in the estimate of the μ term.

(c) Compare the results of the ML fits from parts (a) and (b). Which model do you believe is preferable? Briefly explain your answer.

Comparing the ML fits, the AIC is better for the AR(1) model, at 995.02, compared to the AR(2) model at 997.02.

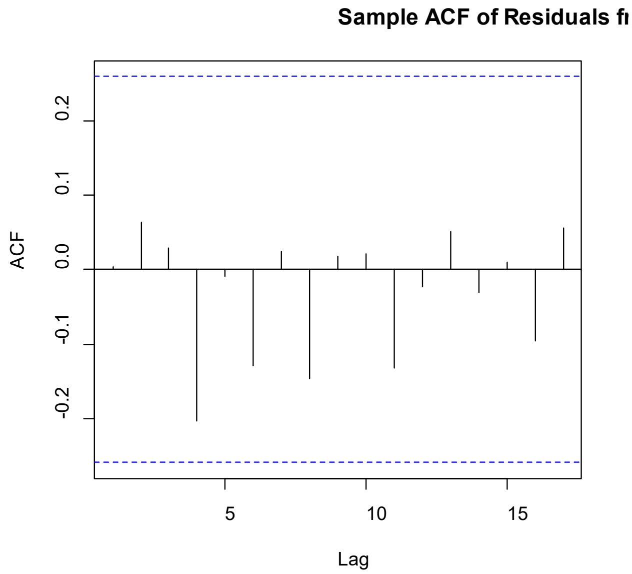
2) Consider the (mean-centered) AR(1) model for the `deere3` data in the `TSA` package, estimated using maximum likelihood.
(a) Give a basic plot of the standardized residuals over time and a Q-Q plot of the residuals. Comment on what these tell you about the adequacy of the model.





The residual plot does not have much pattern, but we may note one possible outlier around time 53. The Q-Q plot shows no evidence of nonnormality. From these plots we don't see any evidence against the AR(1) model.

(b) Give a plot of the sample autocorrelation function of the residuals. Also perform a runs test and a Ljung-Box test (with K = 8). Comment on what these tell you about whether the errors are independent in this model.



The ACF plot shows no significant autocorrelations for the residuals, so the errors can reasonably be assumed to follow white noise. The Ljung-Box test has p-value 0.5895 and the runs test has p-value 0.771, so again we see no evidence that the errors are not independent. The AR(1) model is supported by these diagnostics.

(c) Diagnose the fit of the AR(1) model by using the overfitting strategy.

Comparing the more general AR(2) fit with the AR(1) fit, the extra coefficient in the AR(2) model is NOT significantly different from zero, and the other parameter estimates do not change much between the two models. So the AR(1) model fit looks fine.

3) 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) Estimate the parameters of a (mean-centered) AR(1) model for these data, using maximum likelihood. Give the equation of the estimated model.

Using maximum likelihood, the estimate of ϕ is 0.308 and the estimate of μ is 0.0015. The estimate of the noise variance is 0.00000648. The equation of the estimated model is
$(Y_t - 0.0015) = 0.308\ (Y_{t-1} - 0.0015) + e_t$, with estimated noise variance 0.00000648.

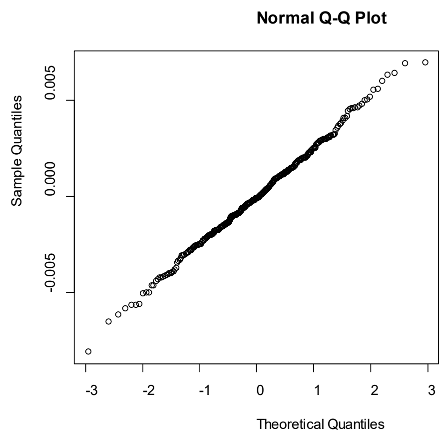
(b) Give an approximate 95% confidence interval for ϕ, the coefficient in the AR(1) model.
$0.308 \pm 1.96[(1 - 0.308^2)/324]^{1/2} = 0.308 \pm 1.96(0.0528) = (0.204, 0.411)$

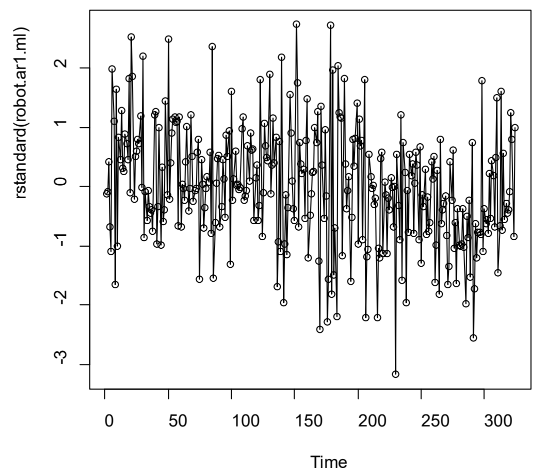
(c) Estimate the parameters of an IMA(1,1) model for these data. Give the equation of the estimated model.
Using maximum likelihood, the estimate of θ is –0.8713 in R's formulation, or 0.8713 in our book's formulation. The equation of the estimated model is
$Y_t = Y_{t-1} + e_t - 0.8713e_{t-1}$, with estimated noise variance 0.00000607.

(d) Compare the results from parts (a) and (c) using AIC.

The AR(1) model has an AIC of -2947.08. The IMA(1,1) model has an AIC of -2959.9. So the IMA(1,1) model is preferred by the AIC approach.

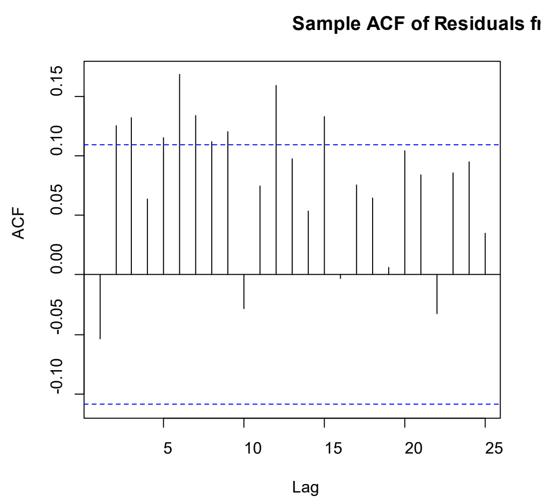
4) Consider the (mean-centered) AR(1) model for the `robot` data in the `TSA` package, estimated using maximum likelihood.
(a) Give a basic plot of the standardized residuals over time and a Q-Q plot of the residuals. Comment on what these tell you about the adequacy of the model.





There is not much pattern in the residual plot, but a slight downward trend over time is noticeable. The Q-Q plot shows no reason to doubt the normality assumption.

(b) Give a plot of the sample autocorrelation function of the residuals. Also perform a runs test and a Ljung-Box test (with K = 30). Comment on what these tell you about whether the errors are independent in this model.

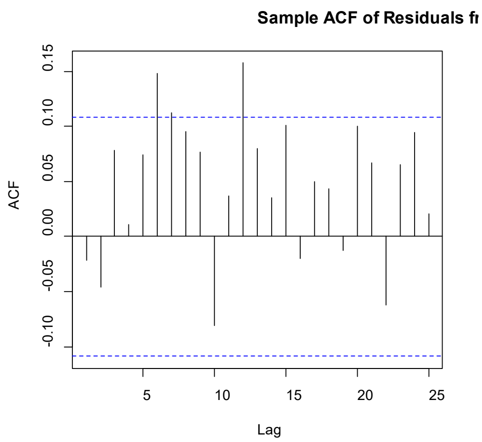


There is significant autocorrelation of the residuals at numerous lags. This is evidence that the errors are not white noise, and perhaps the model should be reconsidered. The Ljung-Box test has p-value near zero, giving evidence that the residual autocorrelations are excessively large. The runs test has p-value 0.17, so it fails to reject the hypothesis of independent errors. All in all, the AR(1) model is not well supported by these diagnostics.

(c) Diagnose the fit of the AR(1) model by using the overfitting strategy.

Overfitting with an AR(2) model, we see the extra AR coefficient in the AR(2) model IS significantly different from zero, and the other parameter estimates change a bit from the AR(1) model. Overfitting with an ARMA(1,1) model, we see the extra MA coefficient in the ARMA(1,1) model is highly significantly different from zero, and the AR coefficient's estimate changes a lot from the AR(1) model. This is evidence that the AR(1) model may be a poor fit.

(d) Repeat part (b), but with the residuals from a (mean-centered) AR(2) model for the robot data. Comment on whether your conclusions are any different.



There is still significant autocorrelation of the residuals at a few lags. This is evidence that the errors may not be white noise, and perhaps the AR(2) model is not the best choice. The Ljung-Box test has p-value near zero, giving evidence that the residual autocorrelations are excessively large. The runs test has p-value 0.07, so it fails to reject the hypothesis of independent errors, but barely. All in all, the AR(2) model is not completely supported by these diagnostics, although it seems a bit better than the AR(1) model.