Statistics 531/Econ 677
Winter, 2008
Ed Ionides

Let x_t be annual flow of the river Nile, measured at Aswan, between 1871 and 1970. x_t is graphed in Fig. 1.

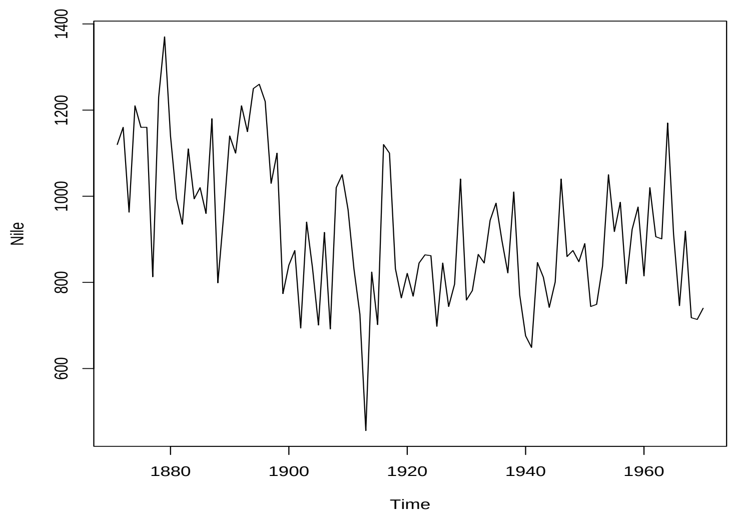

Figure 1: Annual flow of the river Nile, measured at Aswan

Section A [7 points]. A linear regression model $x_t = \beta_0 + \beta_1 t + \epsilon_t$ was fitted using least squares, via the R command `lm(Nile∼time(Nile))`. The output is

```
summary(lm(Nile~time(Nile)))
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6132.1736  1001.7578   6.121 1.92e-08 ***
time(Nile)    -2.7143     0.5216  -5.204 1.07e-06 ***
```

[A1, 1 pt]. What is the name of the time series model which makes the assumptions implicit in the above analysis?

Signal plus white noise.

[A2, 2 pts]. The estimate $\hat{\beta}_1 = -2.71$ is (a) too high; (b) too low; (c) about right. Choose (a), (b) or (c) and explain briefly.

(c) The least squares estimator $\hat{\beta}_1$ is unbiased even when the $\{\varepsilon_t\}$ are correlated.

[A3, 2 pts]. The standard error $SE(\hat{\beta}_1) = 0.53$ is (a) too high; (b) too low; (c) about right. Choose (a), (b) or (c) and explain briefly.

(b) The $\{\varepsilon_t\}$ are correlated (see Fig 1 or Fig 2(a)), so the usual least squares SEs are wrong. Since the correlation is positive, the SEs are too small.

[A4, 2 pts]. If possible from the information given, explain how to make an appropriate test of the hypothesis that $\beta_1 = 0$. If it is not possible, explain why.

It is not possible based on this output, since the SEs are wrong.

Section B [9 points]. An ARMA(1,1) model was fitted using the *R* command `arima(Nile, order = c(1,0,1))`. The output is given below. The sample ACF of x_t is shown in Fig. 2(a).

```
arima(Nile,order=c(1,0,1))
Coefficients:
         ar1      ma1  intercept
      0.8611  -0.5177   920.5567
s.e.  0.1067   0.1908    46.6736

sigma^2 estimated as 19892:  log likelihood = -637.04,  aic = 1282.08
```

[B1, 3 pts]. Write out the fitted model, being carful to specify all model assumptions.

$x_t = 921 + 0.86(x_{t1} - 921) + w_t - 0.52w_{t-1}$, where $w_{t-1} \sim$ i.i.d. $N(0, 19892)$

[B2, 2 pts]. The sample ACF of the residuals was plotted by `acf(resid(arma11))`. (see Fig. 2(b)). Explain how the residuals are defined.

The residuals $\hat{w_t}$ are the estimates of w_t from inverting the ARMA model: $\hat{w_t} = \frac{\phi(\hat{B})}{\theta(\hat{B})}x_t$ where $\phi(\hat{B}) = 1 - 0.86B$, $\theta(\hat{B}) = 1 - 0.52B$. In fact, R gives standardized residuals $\frac{\hat{w_t}}{\hat{\sigma}}$.

[B3, 2 pts]. Explain carefully what the dashed lines show in Fig. 2(a) and 2(b). Discuss briefly how these lines help us to understand the data.

The dashed lines are pointwise 95% confidence intervals for the sample ACF of Gaussian white noise. The sample ACF in 2(a) tells us to reject the white noise hypothesis for x_t. For the residuals, (b) is consistent with white noise.

[B4, 2 pts]. AR(1) and AR(2) models both have noticeably large AIC than ARMA(1,1). Explain why Fig. 2(a) suggests that an MA model is not suitable, and argue that ARMA(1,1)

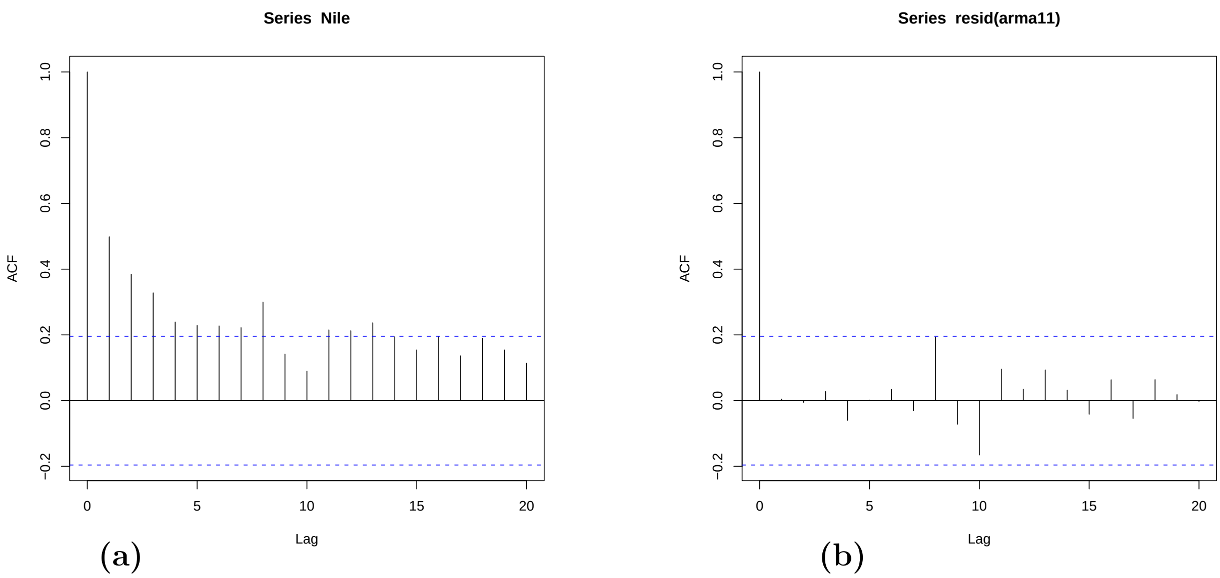

Figure 2: **(a)** Sample ACF for x_t. **(b)** Sample ACF for residuals from fitting ARMA(1,1) to x_t.

is a reasonable model for these data.

An MA model is not suitable because the ACF tails off. Fig 2(b) suggests that ARMA(1,1) fits well, so there is no need to look at larger, more complicated models. We could consider AR(1) or AR(2), but their AICs are less favorable, so ARMA(1,1) is reasonable.

Section C [6 pts]. The raw periodogram of x_t is shown in Fig. 3 with the spectral density of the fitted ARMA(1,1) model shown dashed.

[C1, 3 pts]. Give a formula that could be used to calculate a raw periodogram. Explain briefly how (if at all) the solid line in Fig. 3 is calculated differently from your formula.

$I(\nu) = \frac{1}{T}|\sum_{t=1}^{T} x_t e^{-2\pi i \nu t}|^2$, for observations $x_1, \ldots, x_T$.

R removes a trend and applies a taper when calculating the periodogram.

[C2, 3 pts]. An interesting question is whether the Nile data have a trend or can be modeled as stationary. Discuss this point. There is not necessarily a correct answer, but you should comment on relevant aspects of Sections A, B and C.

The regression in A suggests the presence of a trend, though we cannot tell if this is statistically significant.

In B, a stationary ARMA model model appears to fit the data (though $\hat{\phi}_1 = 0.86$ with

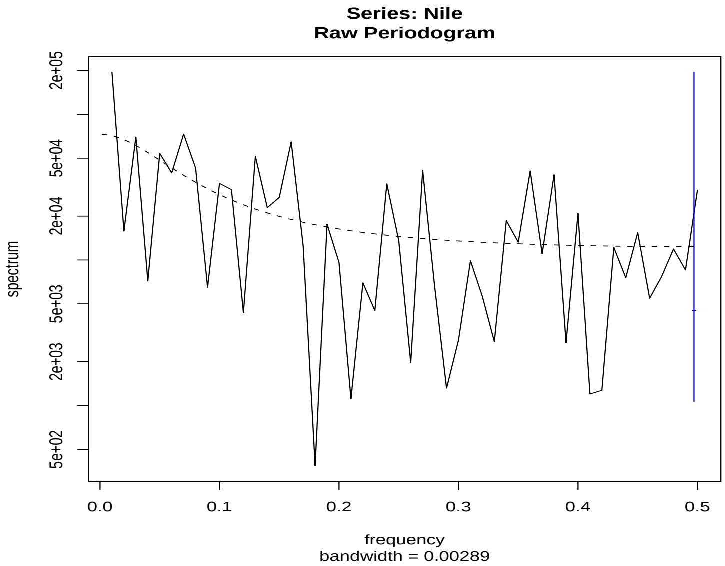

Figure 3: Periodogram, calculated by `spectrum(Nile)`, with the fitted ARMA(1,1) spectrum

SE 0.11 is arguably consistent with a random walk, $\phi = 1$).

The periodogram shows a fairly large residual at the lowest frequency, $\frac{1}{T}$, but the fitted stationary model is still within the 95% confidence interval at this frequency, given by the error bar.