Time Series Exercises 3: Solutions

1. Now we know that

$$\mathrm{Cov}\left(X\left(i\right), X\left(j\right)\right) = \sigma^2 \rho\left(i-j\right)$$

for all i, j, and so

$$\begin{aligned}
\mathrm{Var}\left(\bar{x}\right) &= \frac{1}{N^2}\mathrm{Var}\left(\sum_{t=1}^{N} X\left(t\right)\right) \\
&= \frac{1}{N^2}\left(\sum_{t=1}^{N}\mathrm{Var}\left(X\left(t\right)\right) + 2\sum\sum_{j<i}\mathrm{Cov}\left(X\left(i\right), X\left(j\right)\right)\right) \\
&= \frac{1}{N^2}\left(N\sigma^2 + 2\sum\sum_{j<i}\sigma^2\rho\left(i-j\right)\right) \\
&= \frac{\sigma^2}{N}\left(1 + \frac{2}{N}\sum\sum_{j<i}\rho\left(i-j\right)\right).
\end{aligned}$$

The double summation in the above can be written (taking each value for i in turn)

$$\begin{aligned}
\sum_{i=2}^{N}\sum_{j=1}^{i-1}\rho\left(i-j\right) &= \rho\left(1\right) + \left(\rho\left(2\right)+\rho\left(1\right)\right) + \left(\rho\left(3\right)+\rho\left(2\right)+\rho\left(1\right)\right) + \cdots \\
&\qquad + \left(\rho\left(N-1\right) + \cdots + \rho\left(1\right)\right) \\
&= \left(N-1\right)\rho\left(1\right) + \left(N-2\right)\rho\left(2\right) + \cdots + \left(N-\left(N-1\right)\right)\rho\left(N-1\right) \\
&= \sum_{k=1}^{N-1}\left(N-k\right)\rho\left(k\right)
\end{aligned}$$

from which the result follows.

2. As there is some evidence of a trend in the data, perhaps the series should be differenced to begin with. Examining the sample acf after differencing, the first few values are $r_1 = -0.48, r_2 = 0.033$, $r_3 = -0.049$, $r_4 = -0.106, \ldots$, suggesting a cut-off at lag 1. This points to an MA(1) process. So fitting an ARIMA$(0,1,1)$ to the data gives Res SS $= 0.6789$, χ^2 test statistic $= 4.5$ (on 11 d.o.f.), and fitted model is

$$X\left(t\right) = 0.015 + 1.0598 Z\left(t-1\right) + Z\left(t\right).$$

We can try higher order models, e.g., and ARIMA$(1,1,1)$ has $\chi^2 = 4.6$, ResSS $= 0.6779$, and $\hat{\alpha} = 0.0154$ is too close to zero to be considered worthy of inclusion. Similarly with higher order models, the reduction in Res SS is insufficient to warrant the extra parameters. Note that R provides both the log-likelihood function and AIC, and these numbers are worth examining, particularly as the latter penalises over–fitting. An obvious question to ask is whether the constant term should be included – i.e., is $E(X(t)) = \mu = 0$? R (and Minitab) includes a constant unless instructed not to, but as an aside be aware that we can test the hypothesis that $\mu = 0$ by examining the size of $\bar{x}/\text{e.s.e.}(\bar{x})$, where e.s.e.$(\bar{x})$ is our estimate of the standard deviation of our estimate of μ based on model. Typically we would reject that $\mu = 0$ if $|\bar{x}/\text{e.s.e.}(\bar{x})| > 2$. The e.s.e.$(\bar{x})$ depends on which model is fitted, however. As examples, the following are the estimates of the e.s.e.$(\bar{x})^2$ in low order models, where in each case N observations are taken:

Model	Estimate of var$(\bar{x})$
AR(1)	$\frac{c_0(1+r_1)}{N(1-r_1)}$
MA(1)	$\frac{c_0(1+2r_1)}{N}$
AR(2)	$\frac{c_0(1+r_1)\left(1-2r_1^2+r_2\right)}{N(1-r_1)(1-r_2)}$
MA(2)	$\frac{c_0(1+2r_1+2r_2)}{N}$
ARMA(1, 1)	$\frac{c_0}{N}\left(1+\frac{2r_1^2}{r_1-r_2}\right)$

In our example here we can legitimately set $\mu = 0$, giving rise to an invertible model.

3. Note that the sample acf r_k appears to decay here, whereas the pacf cuts–off after the first lag. This strongly suggests an AR(1) model. We would estimate the parameter α_1 by $r_1 = -0.8$, and using the model–based estimator of σ^2 we would take

$$\begin{aligned} \hat{\sigma}^2 &= c_0(1-\hat{\alpha}_1 r_1) \\ &= 3.34(1-0.64) \\ &= 1.2024. \end{aligned}$$

Is the mean of the underlying series zero? Well $\bar{x} = 0.03$, and in fact from above, the estimate of the standard deviation (i.e., the e.s.e.) for this estimate of μ is

$$\sqrt{\frac{c_0 (1 + r_1)}{N (1 - r_1)}} = 0.04.$$

Now since $|\bar{x}/0.04| < 2$, we would accept that μ is zero. So we would try as an initial model the AR(1)

$$X(t) = -0.8X(t-1) + Z(t)$$

where $Z(t) \sim N(0, 1.2024)$.

4. For an AR(2) process, recall that the Yule–Walker equations are

$$\begin{aligned} \rho(1) &= \alpha_1 \rho(0) + \alpha_2 \rho(1) \\ \rho(2) &= \alpha_1 \rho(1) + \alpha_2 \rho(0), \end{aligned}$$

so the "sample" Yule–Walker equations are (since $\rho(0) = r_0 = 1$)

$$\begin{aligned} r_1 &= \alpha_1 + \alpha_2 r_1 \\ r_2 &= \alpha_1 r_1 + \alpha_2. \end{aligned}$$

Rearranging, we find

$$\alpha_1 = r_1 (1 - \alpha_2)$$

and therefore

$$r_2 - r_1^2 = \alpha_2 \left(1 - r_1^2\right)$$

suggesting estimators based on the sample acf, namely

$$\begin{aligned} \hat{\alpha}_2 &= \frac{r_2 - r_1^2}{1 - r_1^2}, \\ \hat{\alpha}_1 &= \frac{r_1 (1 - r_2)}{1 - r_1^2}. \end{aligned}$$

Substituting $r_1 = 0.81$ and $r_2 = 0.43$ suggests

$$\begin{aligned} \hat{\alpha}_1 &= \frac{0.462}{0.349} = 1.343, \\ \hat{\alpha}_2 &= -\frac{0.226}{0.344} = -0.657. \end{aligned}$$

5. For this command, the syntax is

   ```
   > arima.sim(model, n, rand.gen = rnorm, innov = rand.gen(n,
   ...),n.start = NA, start.innov = rand.gen(n.start, ...), ...)
   ```

   The `model` argument requires the AR and MA coefficients in separate vectors, `n` is the length of the generated series, `rand.gen` stipulates the white noise distribution (Normal being the default), alternatively a series for the $Z(t)$ can be explicitly provided using the `innov` argument. The output is a vector of class `ts.` So to simulate from an AR(1) with $\alpha = 0.5$ and $\sigma^2 = 0.8$, enter

   ```
   > arima.sim(n = 100, list(ar = 0.5),sd = sqrt(0.8))
   ```

   For an AR(1), $\alpha_1 \approx r_1$, recalling that $\rho(1) = \alpha_1$ and r_1 estimates $\rho(1)$. The acf should just decay, and the pacf should "cut–off" after lag 1. For an AR(2), the acf decays and the pacf cuts–off at lag 2. The arguments can be modified to produce an MA(1) process, for instance, where the acf should of course cut–off after lag 1. The pacf should have no (or certainly only a spurious few) "significant" values after the first. Experimenting will demonstrate that the estimator of β found by solving

   $$r_1 = \frac{\beta}{1+\beta^2},$$

   subject to $|\beta| < 1$ is relatively inefficient. However, it is a reasonable starting point, and the best we can find if an MA(1) is to be fitted without the aid of a computer.

6. Consider the process

   $$X(t) = \alpha X(t-1) + Z(t) + \beta Z(t-1).$$

   This can be written as a "pure" MA process by repeated back substitution:

   $$\begin{aligned} X(t) &= \alpha^2 X(t-2) + Z(t) + (\alpha+\beta) Z(t-1) + \alpha\beta Z(t-2) \\ &= \alpha^3 X(t-3) + Z(t) + (\alpha+\beta) Z(t-1) + \alpha(\alpha+\beta) Z(t-2) + \alpha^2\beta Z(t-3) \\ &\vdots \\ &= Z(t) + (\alpha+\beta) Z(t-1) + \alpha(\alpha+\beta) Z(t-2) + \alpha^2(\alpha+\beta) Z(t-3) + \cdots . \end{aligned}$$

   Note that the general coefficient in the above is of the form $\alpha^j(\alpha+\beta)$; for stationarity, we need their sum, $(\alpha+\beta)\sum_{j=0}^{\infty}\alpha^j$, to be finite (i.e., to converge). This can only happen when $|\alpha| < 1$.

7. For the MA(1) model, based on data $x(1), \ldots, x(N)$, we would obviously forecast the next value to be $\hat{x}(N,1) = \hat{\theta} z(N)$, where $\hat{\theta}$ is a suitable estimate of θ and

$$z(N) = e(N) = x(N) - \hat{x}(N)$$

is the residual at time N, being an estimate of $Z(N)$. Clearly for this model

$$\hat{x}(N,l) = 0$$

for $l > 0$.

8. Now here

$$\begin{aligned}\hat{X}(N,1) &= -\beta_1 Z(N) - \beta_2 Z(N-1) \\ \hat{X}(N,2) &= -\beta_2 Z(N)\end{aligned}$$

and

$$\hat{X}(N,l) = 0$$

for $l = 3, 4, \ldots$. Note that for any stationary process with mean μ, using the Box–Jenkins procedure to find forecasts, we have that $\hat{X}(N,l) \to \mu$ as $l \to \infty$. Of course $\mu = 0$ in the model here.

9. Taking unobserved values of $Z(t)$ to be zero, we have

$$\begin{aligned}\hat{x}(360,2) &= -0.3085(-4.0406) - 0.1312 \times 0.77119 - 0.2022 \times 0.442 \\ &= 1.055 \\ \hat{x}(360,3) &= -0.1312(-4.0406) - 0.2022 \times 0.77119 \\ &= 0.374.\end{aligned}$$

10. From the information given, we see

$$\begin{aligned}\hat{x}(N,1) &= 0.5 \times 3.24 + 0 - 0.8 \times 0.64 + 0.4 \times 0.95 \\ &= 1.488. \\ \hat{x}(N,2) &= 0.5\hat{x}(N,1) + 0 + 0 + 0.4 \times 0.64 = 1.00 \\ \hat{x}(N,3) &= 0.5\hat{x}(N,2) + 0 = 0.5.\end{aligned}$$

Given that $x(N+1) = 1.6$, we can deduce the updated forecasts

$$\begin{aligned}
\hat{x}((N+1,1) &= 0.5 \times 1.6 + 0 - 0.8 \times 0.112 + 0.4 \times 0.64 \\
&= 0.9664 \\
\hat{x}(N+1,2) &= 0.5 \times 0.9664 + 0 + 0 + 0.4 \times 0.112 = 0.528.
\end{aligned}$$

11. We know that

$$\hat{X}(N,l) = \hat{\alpha}^l X(N)$$

for each $l = 1, 2, \ldots$. To construct the confidence interval, let us write the $\{\psi_i\}$ weights in the MA representation in terms of α. Here

$$\phi(B) = 1 - \alpha B$$

and since $\theta(B) = 1$ we have

$$\phi(B)\psi(B) = (1 - \alpha B)\left(1 + \psi_1 B + \psi_2^2 B^2 + \cdots\right) = 1.$$

Now we can equate coefficients in B :

$$\begin{aligned}
\psi_1 - \alpha &= 0, \\
\psi_2 - \alpha\psi_1 &= 0 \\
&\vdots
\end{aligned}$$

From this we see

$$\begin{aligned}
\psi_1 &= \alpha, \\
\psi_2 &= \alpha^2 \\
&\vdots \\
\psi_l &= \alpha^l.
\end{aligned}$$

Thus

$$\begin{aligned}
\sum_{j=0}^{l-1} \psi_j^2 &= 1 + \alpha^2 + \cdots + \alpha^{2l-2} \\
&= \frac{1 - \alpha^{2l}}{1 - \alpha^2},
\end{aligned}$$

summing the g.p., assuming $|\alpha| < 1$. So an approximate 95% confidence interval for $\hat{X}(N,l)$ is

$$\hat{x}(N,l) \pm 1.96\hat{\sigma}\sqrt{\frac{1-\hat{\alpha}^{2l}}{1-\hat{\alpha}^{2}}}.$$

12. In general

$$\begin{aligned} e(N,l) &= Z(N+l)+\psi_1 Z(N+l-1)+\cdots+\psi_{l-1}Z(N+1) \\ e(N,l+1) &= Z(N+l+1)+\psi_1 Z(N+l)+\cdots+\psi_l Z(N+1) \end{aligned}$$

hence

$$\begin{aligned} \operatorname{Cov}(e(N,l), e(N,l+1)) &= E(e(N,l)\,e(N,l+1)) \\ &= \sigma^2 \sum_{j=0}^{l-1} \psi_j \psi_{j+1}. \end{aligned}$$

Therefore the correlation between $e(N,l)$ and $e(N,l+1)$ is

$$\frac{\sum_{j=0}^{l-1}\psi_j\psi_{j+1}}{\sqrt{\sum_{j=0}^{l-1}\psi_j^2 \sum_{j=0}^{l}\psi_j^2}}.$$

Now in this example a 95% confidence interval is of the form

$$\hat{x}(60,l) \pm 1.96 \times 0.8\sqrt{\sum_{j=0}^{l}\psi_j^2}.$$

To find some $\{\psi_j\}$ values, note that

$$\theta(B) = \phi(B)\,\psi(B)$$

and here this implies

$$1 = \left(1-0.9B+0.4B^2\right)\left(1+\psi_1 B+\psi_2 B^2+\cdots\right).$$

Now equating coefficients in powers of B, we see

$$\begin{aligned}\psi_1 &= 0.9,\\ \psi_2 - 0.9\psi_1 + 0.4 &= 0\end{aligned}$$

and so $\psi_2 = 0.41$. In general

$$0.4\psi_{j-2} - 0.9\psi_{j-1} + \psi_j = 0.$$

The confidence interval for $x(61)$ is

$$0.9 \times 0.8 - 0.4 \times 1.2 \pm 1.96 \times 0.8\sqrt{1},$$

i.e., 0.24 ± 1.568, or $(-1.328, 1.808)$. For $x(62)$ we have

$$0.9 \times 0.24 - 0.4 \times 0.8 \pm 1.96 \times 0.8\sqrt{1 + 0.9^2},$$

i.e., -0.104 ± 2.10, or $(-2.20, 2.00)$. Similarly for $x(63)$,

$$0.9 \times (-0.104) - 0.4 \times 0.24 \pm 1.96 \times 0.8\sqrt{1 + 0.9^2 + 0.41^2},$$

i.e., -0.1896 ± 2.20, or $(-2.39, 2.01)$. For the estimated correlations,

$$\begin{aligned}\hat{\rho}(e(60,1), e(60,2)) &= \frac{1 \times 0.9}{\sqrt{1 \times (1 + 0.9^2)}} = 0.669,\\ \hat{\rho}(e(60,2), e(60,3)) &= \frac{1 \times 0.9 + 0.9 \times 0.41}{\sqrt{(1 + 0.9^2)(1 + 0.9^2 + 0.41^2)}} = 0.671.\end{aligned}$$

13. .

(a) The function `rev` reverses a vector.

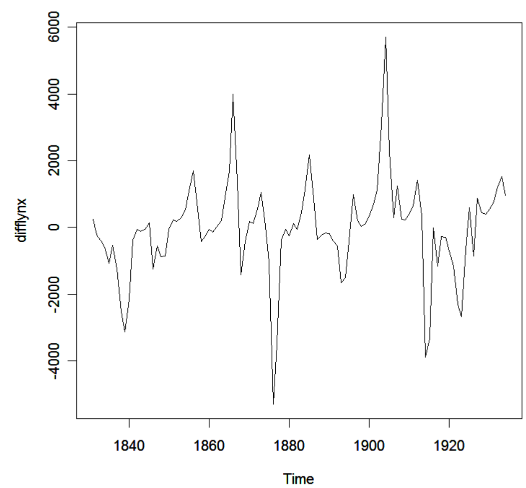
(b) Note that the final value of `lynx` is 3396.

(c) The best value of α here is 1, either found by trial and error or by plotting the estimated value against α. The prediction of 2657 is poor, however. The problem is that the series has some cyclical effects, which the simple exponential smoothing is ignoring.

(d) Use the command

```
difflynx <- diff(newlynx, lag=10)
```

The plot is below:



(e) Since

$$d(t) = x(t) - x(t-10),$$

for $t = 11, \ldots, 113$, then rearranging we find

$$x(t+1) = d(t) - x(t-9).$$

So for prediction we might use

$$\hat{x}(N,1) = \hat{d}(N,1) + x(N-9).$$

Here $N = 113$, so we have

$$\hat{x}(113,1) = \hat{d}(113,1) + x(104),$$

with $x(104) = 2432$. So we need to forecast the value of $d(114)$. Following the approach above, we might use `es` on the values up to $t = 112$, and see which value of α best predict $d(113) = 1525$. Again $\alpha = 1$ seems best, giving and estimate of 1191. This leads to $\hat{x}(113,1) = 2432 + 1191 = 3623$, a rather better forecast than before.

14. .

(a) The acf is decaying slowing, indicating possibly a non-stationary model, or else an AR (or maybe ARMA) model with long–term dependence. The pacf cuts off noticably at lag 2, suggesting an AR(2) could be suitable. We can examine this model via
$>$ `ar2Huron <- ar(LakeHuron, order.max=2, method=''ols'', demean=T)`
The model fitted to $Y(t) = X(t) - \hat{\mu}$ is

$$Y(t) = 1.0217Y(t-1) - 0.2376Y(t-2) + Z(t),$$

with $\hat{\sigma} = 0.6738$. Checking this model is stationary, we find the roots of

$$0.2376x^2 - 1.0217x + 1 = 0$$

are 1.5067 and 2.7934, and both are outside the unit circle.
An alternative method is via
$>$ `ar2Huron <- arima(LakeHuron, order=c(2,0,0))`
This fits

$$X(t) - 597.047 = 1.0436(X(t-1) - 597.047) - 0.2495(X(t-2) - 597.047) + Z(t)$$

with $\hat{\sigma} = 0.69197$. This is fitted via a hybrid approach – conditional least squares giving starting values for m.l.e.; hence the model is slightly different to the one given earlier. Probably this model is preferable.

(b) To examine the residuals, note the first two values are `na`'s in the fit, since the `ols` used above fixes the first $p = 2$ values. We can still plot the residuals, and note that few fall outside the range $(-1.5, 1.5)$. To view the acf, we can use
$>$ `acf(ar2Huron$resid[3:98])`
Pleasingly, there are no significant lags after lag 0. The goodness–of–fit tests will be non-significant. The `tsdiag` command works on the model fitted by the `arima` command, and indicates a reasonable fit.

(c) The `predict` command on the model fitted via `arima` gives

$$\begin{aligned} \hat{x}(98,1) &= 579.7896, \\ \hat{x}(98,2) &= 579.5942, \\ \hat{x}(98,3) &= 579.4329 \end{aligned}$$

with e.s.e.s being 0.69197, 1.00016, 1.1567. Consequently 95% prediction intervals are

$$\begin{aligned} 579.7896 \pm 1.96 \times 0.69197 &= (578.43, 581.15)\,, \\ 579.5942 \pm 1.96 \times 1.00016 &= (577.63, 581.55)\,, \\ 579.4329 \pm 1.96 \times 1.1567 &= (577.17, 581.70)\,, \end{aligned}$$

with all figures in feet. Note of course these are progressively wider. The forecasts from the model fitted by least squares are slightly different.

(d) Easiest is via
`> ARMAtoMA(ar=c(1.0436136,-0.249497), lag.max=3)`
which gives output
`1.0436136 0.8396323 0.6158733`
These figures estimate ψ_1, ψ_2 and ψ_3 in turn. Now since $\psi_0 = 1$ we check

$$\begin{aligned} 0.69197\sqrt{1} &= 0.69197, \\ 0.69197\sqrt{1 + 1.0436^2} &= 1.00015, \\ 0.69197\sqrt{1 + 1.0436^2 + 0.8396^2} &= 1.1566 \end{aligned}$$

and the confidence intervals above are attained, more or less.

BD