Activity Solution: Box–Jenkins Forecasting

This activity aims to help you understand how to apply Box–Jenkins forecasting methods once a model has been fitted to a time series. The following model

$$X(t) = 0.5X(t-1) + Z(t) - 0.8Z(t-1) + 0.4Z(t-2)$$

has been fitted to a series of 100 observations, for which $x(100) = 3.24$, $z(100) = 0.64$, and $z(99) = 0.95$. The residual sum of squares for the fit is 44.89.

1. *Clickers question:* Forecast the values of the process at times 101, 102, and 103.
   From the information given, we see

$$\begin{aligned}
\hat{x}(100,1) &= 0.5 \times 3.24 + 0 - 0.8 \times 0.64 + 0.4 \times 0.95 \\
&= 1.488. \\
\hat{x}(100,2) &= 0.5\hat{x}(100,1) + 0 + 0 + 0.4 \times 0.64 = 1.00 \\
\hat{x}(100,3) &= 0.5\hat{x}(100,2) + 0 = 0.5.
\end{aligned}$$

2. *Clickers question:* In order to create prediction intervals here, we require an estimate of the unknown variance σ^2. We can use an estimator based on the residual sum of squares, as in linear regression: the residual mean square can be used as an estimator of σ^2, defined as

$$\frac{\text{Res SS}}{N-p}$$

   where here $N = 100$ and p is the number of parameters estimated in fitting the model, excluding σ. Compute your estimate of σ for this model fit.
   The number of parameters estimated is 3. This suggests $\hat{\sigma} = \sqrt{44.89/97} = 0.6803$.

3. *Clickers question:* Find the (approximate) 95% prediction intervals for your forecasts in part 1.

Note here that

$$\begin{aligned}
X(t) &= (1-0.5B)^{-1}\left(1-0.8B+0.4B^2\right)Z(t) \\
&= \left(1+0.5B+0.5^2B^2+0.5^3B^3+\cdots\right)\left(1-0.8B+0.4B^2\right)Z(t) \\
&= (1+(0.5-0.8)B+\left(0.5^2+0.4-0.5\times 0.8\right)B^2+ \\
&\quad \left(0.5^3+0.5\times 0.4-0.8\times 0.5^3\right)B^3+\cdots)Z(t) \\
&= \left(1-0.30B+0.25B^2+0.225B^3+\cdots\right)Z(t),
\end{aligned}$$

so we have $\hat{\psi}_0 = 1$, $\hat{\psi}_1 = -0.30$, $\hat{\psi}_2 = 0.25$. *This leads to 95% prediction intervals here based on*

$$\hat{x}(100,l) \pm 1.96\hat{\sigma}\sqrt{1+\hat{\psi}_1^2+\hat{\psi}_2^2+\cdots+\hat{\psi}_{l-1}^2}.$$

Using $\hat{\sigma} = \sqrt{44.89/97} = 0.6803$, *we have for* $X(101)$ *an interval*

$$1.488 \pm 1.96\times 0.6803\sqrt{1} = 1.488 \pm 1.333,$$

for $X(102)$ *the interval is*

$$1.0 \pm 1.96\times 0.6803\sqrt{1+(-0.30)^2} = 1.0 \pm 1.392,$$

and for $X(103)$,

$$0.5 \pm 1.96\times 0.6803\sqrt{1+(-0.30)^2+0.25^2} = 0.5 \pm 1.4315\,.$$

4. *Clickers question:* If $x(101) = 1.60$, find $\hat{x}(101,1)$ and $\hat{x}(101,2)$.
   Given that $x(101) = 1.60$, *we can deduce the updated forecasts*

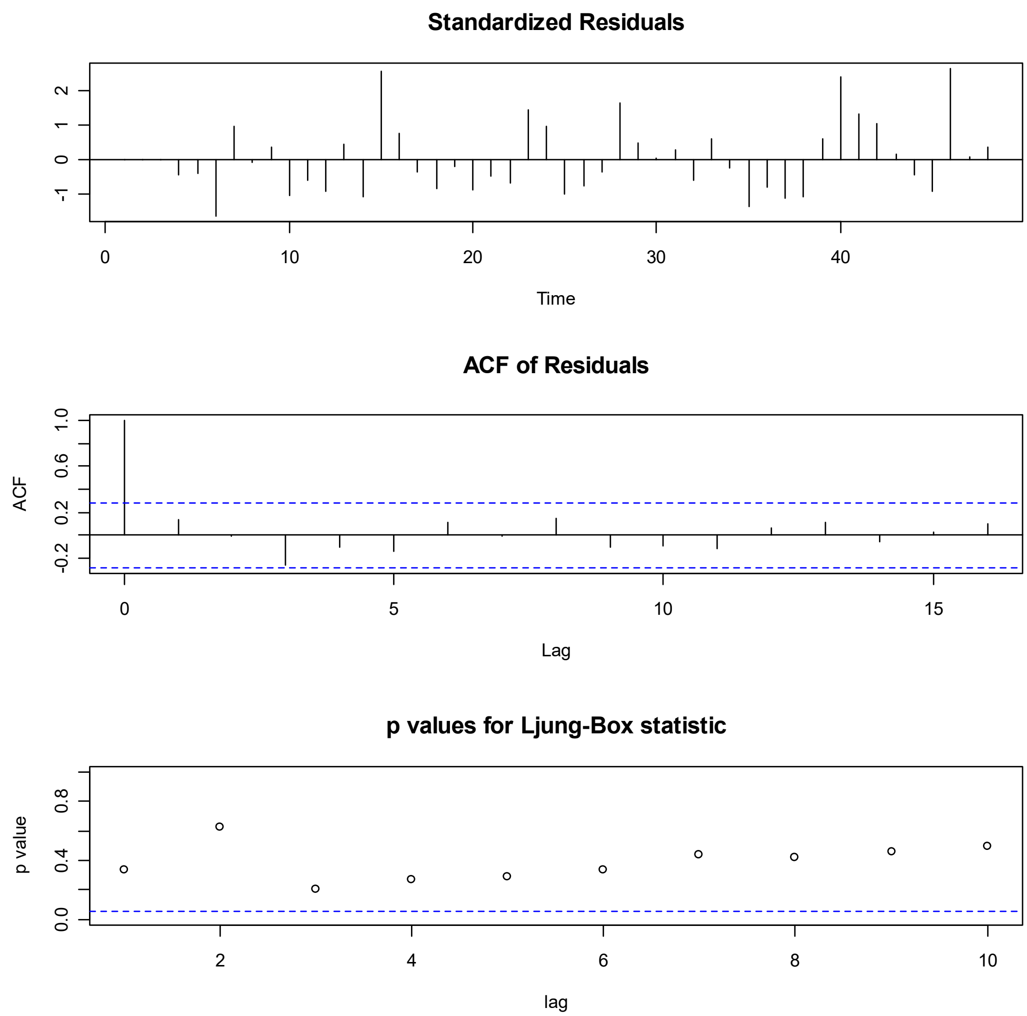
$$\begin{aligned}
\hat{x}(101,1) &= 0.5\times 1.60+0-0.8\times 0.112+0.4\times 0.64 \\
&= 0.9664 \\
\hat{x}(101,2) &= 0.5\times 0.9664+0+0+0.4\times 0.112 = 0.528.
\end{aligned}$$

5. What would happen here to our forecast value $\hat{x}(101,l)$ as l grows large?
   When the lead time increases the forecast value will tend to zero, the mean of $X(t)$.

Recall the lh data, comprising 48 observations of the level of luteinizing hormone in the blood of a woman, taken at ten minute intervals. Earlier analysis suggested an AR(1) model, though cases could be made for both MA(1) and ARMA(1, 1) models. Fitting an AR(1) by m.l.e., the model is

$$X(t) - 2.413 = 0.574\,(X(t-1) - 2.413) + Z(t)\,,$$

for which the residual MS is 0.197. This differs slightly from the model found by least squares. Diagnostic plots for the model are plotted below:



1. Comment on each of the plots above with regards to the support provided for the model fitted.

For the standardised residual plot, we would expect to see no more than a few outside the range ± 2. There appears to be three large standardised residuals, though none are much above +2. Hence this plot gives no major cause for concern. More tellingly, the acf for the residuals has no significant values, suggesting all time–dependent behaviour has been accounted for in the model. Finally, none of the P-values for the Box-Ljung tests are below 0.05, confirming the evidence from the acf. Overall, the model fit looks satisfactory.

2. *Clickers question:* Now $x(48) = 2.9$ is the final observation. Provide forecasts for the next two values in the series.
   So

$$\begin{aligned}\hat{x}(48,1) &= \hat{\alpha}(x(48) - \hat{\mu}) + \hat{\mu} \\ &= 0.574(2.9 - 2.413) + 2.413 \\ &= 2.692\end{aligned}$$

   is the one–step forecast. The two–step forecast is

$$\begin{aligned}\hat{x}(48,2) &= \hat{\alpha}(\hat{x}(48,1) - \hat{\mu}) + \hat{\mu} \\ &= 0.574(4.0776) + 2.413 \\ &= 2.573.\end{aligned}$$

3. *Clickers question:* Estimate the standard deviations of the forecasts you found above.
   To find the standard errors, which for $\hat{x}(48,l)$ is

$$\hat{\sigma}\left(\sum_{i=0}^{l-1}\hat{\psi}_i^2\right)^{\frac{1}{2}},$$

   note that

$$\begin{aligned}(1-\hat{\alpha})^{-1} &= (1-0.574)^{-1} \\ &= 1 + 0.574 + 0.574^2 + \cdots \\ &= \hat{\psi}_0 + \hat{\psi}_1 + \hat{\psi}_2 + \cdots.\end{aligned}$$

   Hence

$$\hat{\sigma} = (0.197)^{\frac{1}{2}} = 0.444$$

gives the e.s.d. of $\hat{x}(48, 1)$. *In the same manner*

$$\hat{\sigma}\left(1 + \hat{\psi}_1^2\right)^{\frac{1}{2}} = 0.444\left(1 + 0.574^2\right)^{\frac{1}{2}} = 0.512$$

is the e.s.d. of $\hat{x}(48, 2)$, *and so on.*

4. *Clickers question:* Hence find 95% prediction intervals for your forecasts.
   The prediction intervals are of the form

   $$\hat{x}(48, l) \pm 1.96 \times e.s.e.\left(\hat{x}(48, l)\right).$$

   For $l = 1$, *this is*

   $$2.692 \pm 0.870\,2$$

   and for lead time 2 the interval is

   $$2.\,573 \pm 1.0035$$

5. What would happen here to our forecast value $\hat{x}(48, l)$ as l grows large?
   When the lead time increases the forecast value will tend to 2.413, *the estimated mean of* $X(t)$.

6. Re-cap what you have done during this activity. Why do you think you were asked to do this activity? What have you learned?
   Box–Jenkins forecasting was applied using a model fitted to (simulated) data. The model provided forecasts up to lead time of three time units, an estimate of the standard deviation of the white noise process underlying the model fitted, and prediction intervals. We also saw how to update forecasts in the light of new data.
   We then examined the diagnostics from a fitted time series model and evaluated the fit of the model based on this information. The fitted model has been used to provide forecasts, estimated standard deviations for these forecasts, and corresponding prediction intervals. Finally, we appreciated that the forecast values will converge to the estimated mean as the lead time increases. The case study provided an overview of model diagnostics and Box–Jenkins forecasting. Although in practice such calculations would be performed using software, it is reassuring to appreciate what the software does.
   Learning outcomes associated with this activity include:

(a) *Describe and implement the Box–Jenkins approach to forecasting.*

(b) *Apply Box–Jenkins method to forecast future values of a time series, given necessary parameter estimates.*

(c) *Compute prediction intervals for a Box–Jenkins forecast.*

(d) *Describe the behaviour of Box–Jenkins forecasts as the lead time increases.*

(e) *Fit appropriate ARMA models to time series, considering diagnostic checks where relevant.*

(f) *Interpret results of model diagnostic tests based on the residuals of a fitted time series model.*