1. Analyze the following residuals analysis plots. Do the residuals look like white noise?

   Source: Section 5.11 Exercise 4 from *Forecasting: Principles and Practice* (3rd ed).

   (a) Figure 1 showcasing the residuals of the seasonal naïve forecast for the Australian brick production.

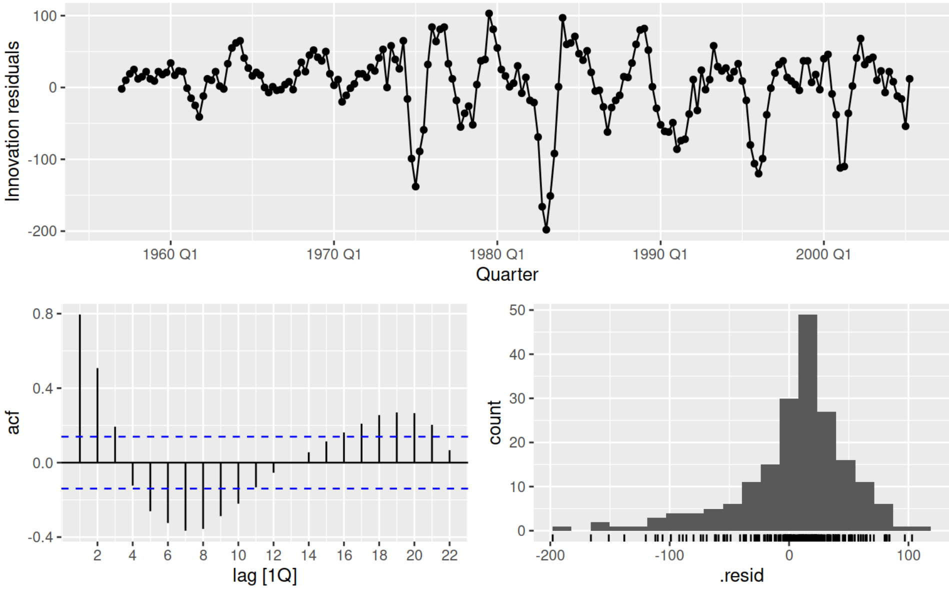


Figure 1: Residuals of the seasonal naïve forecast for the Australian brick production.

**Solution:**

- The residuals do not appear random. Periods of low production and high production are evident, leading to autocorrelation in the residuals.
- The residuals from this model are not white noise. The ACF plot shows a strong sinusoidal pattern of decay, indicating that the residuals are autocorrelated.
- The histogram is also not normally distributed, as it has a long left tail.

(b) Figure 2 showcasing the residuals of the naïve forecast for the Australian Exports series.

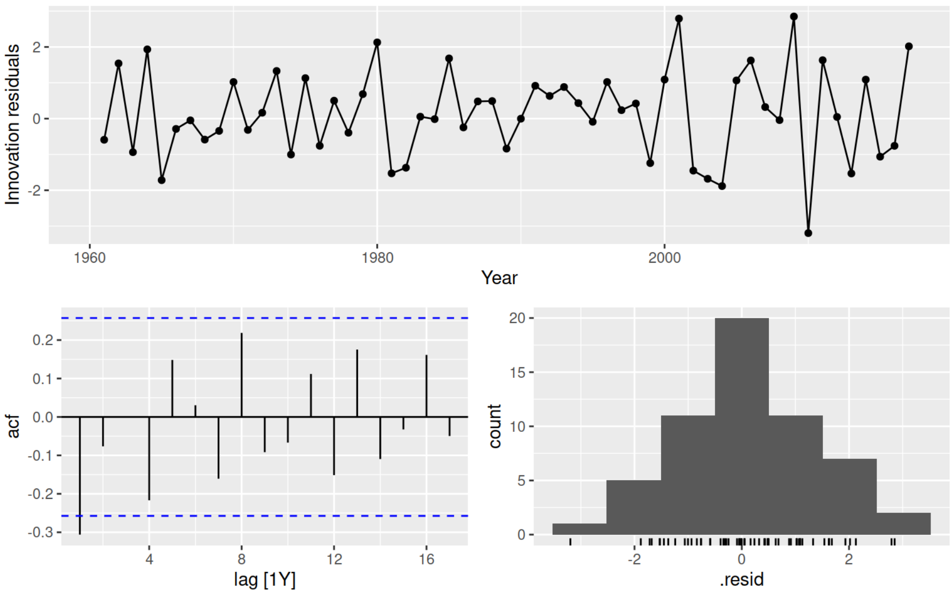


Figure 2: Residuals of the naïve forecast for the Australian Exports series.

**Solution:**

- The ACF plot reveals that the first lag exceeds the significance threshold.
- This data may still be white noise, as it is the only lag that exceeds the blue dashed lines (5% of the lines are expected to exceed it). However as it is the first lag, it is probable that there exists some real auto-correlation in the residuals that can be modelled.
- The distribution appears normal.
- The residual plot appears mostly random, however more observations appear to be above zero. This again, is due to the model not capturing the trend.

2. Are the following statements true or false? Explain your answer.

   Source: Section 5.11 Exercise 6 from *Forecasting: Principles and Practice* (3rd ed).

   (a) Good forecast methods should have normally distributed residuals.

   **Solution: False.** Although many good forecasting methods produce normally distributed residuals this is not required to produce good forecasts. Other forecasting methods may use other distributions, it is just less common as they can be more difficult to work with.

   (b) A model with small residuals will give good forecasts.

**Solution: False.** It is possible to produce a model with small residuals by making a highly complicated (overfitted) model that fits the data extremely well. This highly complicated model will often perform very poorly when forecasting new data.

(c) The best measure of forecast accuracy is MAPE.

**Solution: False.** There is no single best measure of accuracy - often you would want to see a collection of accuracy measures as they can reveal different things about your residuals. MAPE in particular has some substantial disadvantages - extreme values can result when $x_t$ is close to zero, and it assumes that the unit being measured has a meaningful zero.

(d) If your model doesn't forecast well, you should make it more complicated.

**Solution: False.** There are many reasons why a model may not forecast well, and making the model more complicated can make the forecasts worse. The model specified should capture structures that are evident in the data. Although adding terms that are unrelated to the structures found in the data will improve the model's residuals, the forecasting performance of the model will not improve. Adding unnecessary predictors that are not related to the structure of the data can lead to overfitting, which means the model will perform well on the training data but may fail to generalize and make accurate forecasts on new data. This can also reduce the accuracy of coefficient estimates because the unnecessary predictor increases the variance in the model's estimates. Adding missing features relevant to the data (such as including a seasonal pattern that exists in the data) should improve forecast performance.

(e) Always choose the model with the best forecast accuracy as measured on the test set.

**Solution: False.** There are many measures of forecast accuracy, and the appropriate model is the one which is best suited to the forecasting task. For instance, you may be interested in choosing a model which forecasts well for predictions exactly one year ahead. In this case, using cross-validated accuracy could be a more useful approach to evaluating accuracy.

3. The following time series data represents quarterly sales (in units) for a product over the past fours years: $\{10, 12, 20, 22, 12, 14, 22, 24, 14, 16, 24, 26, 16, 18, 26, 28\}$.

(a) Split the data into training (75%) and evaluation (25%) sets.

**Solution:** The 75-25% split implies using the first 12 data points for training and the remaining 4 for evaluation:

- Training split: $\{10, 12, 20, 22, 12, 14, 22, 24, 14, 16, 24, 26\}$
- Evaluation split: $\{16, 18, 26, 28\}$

(b) Provide forecasts for the evaluation set with two baselines of your choice.

**Solution:**

- Naïve forecast: $\{26, 26, 26, 26\}$
- Seasonal naïve forecast with $P = 4$: $\{14, 16, 24, 26\}$

(c) Compute the baselines' mean absolute error.

**Solution:**

- The naïve forecast errors are $\{-10, -8, 0, -2\} \Rightarrow MAE = 5$
- The seasonal naïve forecast errors $\{2, 2, 2, 2\} \Rightarrow MAE = 2$

(d) The company model forecasted $\{25, 22, 28, 31\}$ for the evaluation set. Compute the mean absolute scaled error.

**Solution:** The model errors are $\{-9, -4, -2, -3\}$.
The training forecasts for the naïve baselines are:

$$\{2, 8, 2, -10, 2, 8, 2, -10, 2, 8, 2\}$$

Thus, the naïve baseline training MAE is $\frac{56}{11}$.
The scaled model errors are

$$\left\{\frac{-9}{\frac{56}{11}}, \frac{-4}{\frac{56}{11}}, \frac{-2}{\frac{56}{11}}, \frac{-3}{\frac{56}{11}}\right\} = \left\{\frac{-99}{56}, \frac{-44}{56}, \frac{-22}{56}, \frac{-33}{56}\right\}.$$

Finally, we can compute the model mean absolute scaled error:

$$MASE = \frac{1}{4}(\frac{99}{56} + \frac{44}{56} + \frac{22}{56} + \frac{33}{56}) = \frac{1}{4} \cdot \frac{198}{56} = \frac{99}{112}$$

(e) The forecasts produced by the company model have a horizon of half a year. Compute the rolling mean absolute error with a step size of one quarter.

**Solution:** We calculate the errors and rolling MAE for each step:

- Step 1 errors: $\{-9, -4\} \Rightarrow MAE = \frac{13}{2}$
- Step 2 errors: $\{-4, -2\} \Rightarrow MAE = 3$
- Step 3 errors: $\{-2, -3\} \Rightarrow MAE = \frac{5}{2}$

Thus, the overall rolling MAE is 4.
Note that the rolling forecast acts similarly to a weighted smoothing moving average on the forecast errors. Choosing a step-size smaller than the forecast horizon introduces redundancy, reducing the independence of each evaluation. The typical choice for the step size in rolling forecasts is usually equal to the forecast horizon.

4. A company is comparing two different time series models to forecast an economic indicator based on a dataset containing 500 observations. The models have the following properties:

   - AR(2) with intercept ($c \neq 0$), Log-likelihood = -150
   - AR(3) with intercept ($c \neq 0$), Log-likelihood = -145

   (a) Compute the AIC and BIC for both models.

   **Solution:** The AR(2) model has 3 parameters while the AR(3) has 4 parameters.

$$\begin{aligned}
\text{AIC}_{\text{AR(2)}} &= 2k - 2\log(L) = 2 \times 3 - 2(-150) = 306 \\
\text{AIC}_{\text{AR(3)}} &= 2 \times 4 - 2(-145) = 298 \\
\text{BIC}_{\text{AR(2)}} &= k\log(n) - 2\log(L) = 3\log(500) - 2(-150) \approx 318.64 \\
\text{BIC}_{\text{AR(3)}} &= 4\log(500) - 2(-145) \approx 314.86
\end{aligned}$$

   (b) Based on the AIC and BIC values, which model should be preferred? Explain your answer.

   **Solution:** The AR(3) has lower AIC and BIC values than the AR(2) indicating it is preferable in terms of goodness of fit, penalized for the number of parameters. Thus, the AR(3) should be preferred over the AR(2).

5. A company is forecasting its quarterly sales for the next year. The forecast for each quarter is 50, 55, 53, and 60 units respectively. The variance of the forecasted residuals is estimated to be 16. Compute the 95% confidence interval for the forecasted sales in the first quarter, assuming that the residuals follow a normal distribution.

   **Solution:** For a normal distribution, the 95% confidence interval for the forecast is given by:

$$\text{CI} = \hat{x}_i \pm 1.96 \times \hat{\sigma}_i$$

   Substituting the given values:

$$\text{CI} = 50 \pm 1.96 \times 4 = 50 \pm 7.84$$

   Thus, the 95% confidence interval for the first quarter forecast is $[42.16, 57.84]$.

6. Consider the process $x_t = 6 + w_t + w_{t-1}$ with $w_t \sim \mathcal{N}(0, \sigma^2)$, and the observed realization

$$(x_1, x_2, x_3, x_4) = (5, 7, 5, 7).$$

   (a) Compute the past one-step-ahead forecast errors.

**Solution: Residual recursion.** Using $\hat{w}_t = x_t - 6 - \hat{w}_{t-1}$ with $\hat{w}_0 = 0$:

$$\begin{aligned}
\hat{w}_1 &= x_1 - 6 - \hat{w}_0 = 5 - 6 - 0 = -1, \\
\hat{w}_2 &= x_2 - 6 - \hat{w}_1 = 7 - 6 - (-1) = 2, \\
\hat{w}_3 &= x_3 - 6 - \hat{w}_2 = 5 - 6 - 2 = -3, \\
\hat{w}_4 &= x_4 - 6 - \hat{w}_3 = 7 - 6 - (-3) = 4.
\end{aligned}$$

(b) Derive the forecasts $\hat{x}_{5|4}$, $\hat{x}_{6|4}$, $\hat{x}_{7|4}$.

**Solution:** Consider the general MA(1) model

$$x_t = c + w_t + \theta w_{t-1}$$

**One-step-ahead forecast** $h = 1$.
We are at time $t$, and want to determine the forecast of $x_{t+1}$.

$$x_{t+1} = c + w_{t+1} + \theta w_t$$

We know $w_t$ from the fitted residual $\hat{w}_t$, while $w_{t+1}$ is future noise:

$$\hat{x}_{t+1|t} = E[x_{t+1} \mid x_{1:t}] = E[c \mid x_{1:t}] + E[w_{t+1} \mid x_{1:t}] + E[\theta w_t \mid x_{1:t}] = c + \theta\, \hat{w}_t$$

**Multi-steps-ahead forecast** $h \geq 2$.

$$x_{t+h} = c + w_{t+h} + \theta w_{t+h-1}$$

At time $t$ and for $h \geq 2$, both $w_{t+h}$ and $w_{t+h-1}$ are future noise:

$$\hat{x}_{t+h|t} = E[x_{t+h} \mid x_{1:t}] = c \qquad (h \geq 2)$$

Therefore, the requested forecasts are:

$$\hat{x}_{5|4} = c + \theta \hat{w}_4 = 6 + 1 \cdot 4 = 10, \qquad \hat{x}_{6|4} = \hat{x}_{7|4} = c = 6.$$

(c) Derive the forecast error and variance.

**Solution: One-step-ahead forecast** $h = 1$.

$$x_{t+1} = c + w_{t+1} + \theta w_t, \qquad \hat{x}_{t+1|t} = c + \theta w_t$$

Error:

$$e_{t+1} = x_{t+1} - \hat{x}_{t+1|t} = (c + w_{t+1} + \theta w_t) - (c + \theta w_t) = w_{t+1}$$

Variance:

$$\mathrm{Var}(e_{t+1}) = \mathrm{Var}(w_{t+1}) = \sigma^2$$

**Multi-steps-ahead forecast** $h \geq 2$.

$$x_{t+h} = c + w_{t+h} + \theta w_{t+h-1}, \qquad \hat{x}_{t+h|t} = c$$

Error:

$$e_{t+h} = x_{t+h} - \hat{x}_{t+h|t} = (c + w_{t+h} + \theta w_{t+h-1}) - c = w_{t+h} + \theta w_{t+h-1}$$

Variance (sum of independent terms):

$$\text{Var}(e_{t+h}) = \text{Var}(w_{t+h}) + \theta^2 \text{Var}(w_{t+h-1}) = \sigma^2 + \theta^2\sigma^2 = \sigma^2(1+\theta^2)$$

(d) Compute 95% confidence intervals for $\hat{x}_{5|4}$, $\hat{x}_{6|4}$, and $\hat{x}_{7|4}$.

**Solution:** Estimate the innovation variance:

$$\sigma_e^2 = \frac{1}{T-k-m} \sum_{i=m+1}^{T} \hat{w}_i^2$$

with $m = 1$ because the first residuals $\hat{w}_1$ depends on the assumed initial value $\hat{w}_0 = 0$, and $t = 2$ because the MA(1) model has two parameters, the mean $c$ and the coefficient $\theta$:

$$\hat{\sigma}_e^2 = \frac{2^2 + (-3)^2 + 4^2}{4-1-2} = 29.$$

Thus, the forecast error variances are:

$$\sigma_{5|4}^2 = 29, \qquad \sigma_{6|4}^2 = \sigma_{7|4}^2 = 29 \cdot (1+\theta^2).$$

With $z_{0.975} = 1.96$:

$$CI(\hat{x}_{5|4}) = 10 \pm 1.96\sqrt{29}, \qquad CI(\hat{x}_{6|4}) = CI(\hat{x}_{7|4}) = 6 \pm 1.96\hat{\sigma}_e\sqrt{58}.$$

For a moving-average process, the forecast variance increases up to the model's order and then remains constant; in an MA(1), it rises from $\sigma^2$ at one step ahead to $\sigma^2(1+\theta^2)$ for two or more steps ahead, because beyond one step all future values depend only on new, independent shocks.