Final

Aaron Banlao

2023-06-13

```
library(pacman)
p_load(tidyverse, fpp3, tsibbledata)
```

```
data("souvenirs")
souvenirs
```

```
## # A tsibble: 84 x 2 [1M]
##       Month Sales
##       <mth> <dbl>
##  1 1987 Jan 1665.
##  2 1987 Feb 2398.
##  3 1987 Mar 2841.
##  4 1987 Apr 3547.
##  5 1987 May 3753.
##  6 1987 Jun 3715.
##  7 1987 Jul 4350.
##  8 1987 Aug 3566.
##  9 1987 Sep 5022.
## 10 1987 Oct 6423.
## # ... with 74 more rows
```

Exercise 7.10 Problem 4

The data set souvenirs concerns the monthly sales figures of a shop which opened in January 1987 and sells gifts, souvenirs, and novelties. The shop is situated on the wharf at a beach resort town in Queensland, Australia. The sales volume varies with the seasonal population of tourists. There is a large influx of visitors to the town at Christmas and for the local surfing festival, held every March since 1988. Over time, the shop has expanded its premises, range of products, and staff.

Produce a time plot of the data and describe the patterns in the graph. Identify any unusual or unexpected fluctuations in the time series. Explain why it is necessary to take logarithms of these data before fitting a model. Fit a regression model to the logarithms of these sales data with a linear trend, seasonal dummies and a "surfing festival" dummy variable. Plot the residuals against time and against the fitted values. Do these plots reveal any problems with the model? Do boxplots of the residuals for each month. Does this reveal any problems with the model? What do the values of the coefficients tell you about each variable? What does the Ljung-Box test tell you about your model? Regardless of your answers to the above questions, use your regression model to predict the monthly sales for 1994, 1995, and 1996. Produce prediction intervals for each of your forecasts. How could you improve these predictions by modifying the model?

Answer

a)

As the description of the problem states, there is an influx of visitors around Christmas time and in March, which is reflected in the plot. What is portrayed is that there is a growth in the trend that occurs every year, with the exception of the year 1991.

b)

A logarithm is necessary because the beginning of the data starts out with smaller values while the data points end with extreme values.

d)

When plotting the time against the fitted values, there seems to be no apparent pattern, indicating this is a good model.

e)

The residual plot shows that the variation of the residuals are not constant

f)

The coefficients of each variable measure the magnitude of the growth trend of sales of each month.

g)

We can conclude that the results are statistically significant and that the residuals are distinguishable from white noise

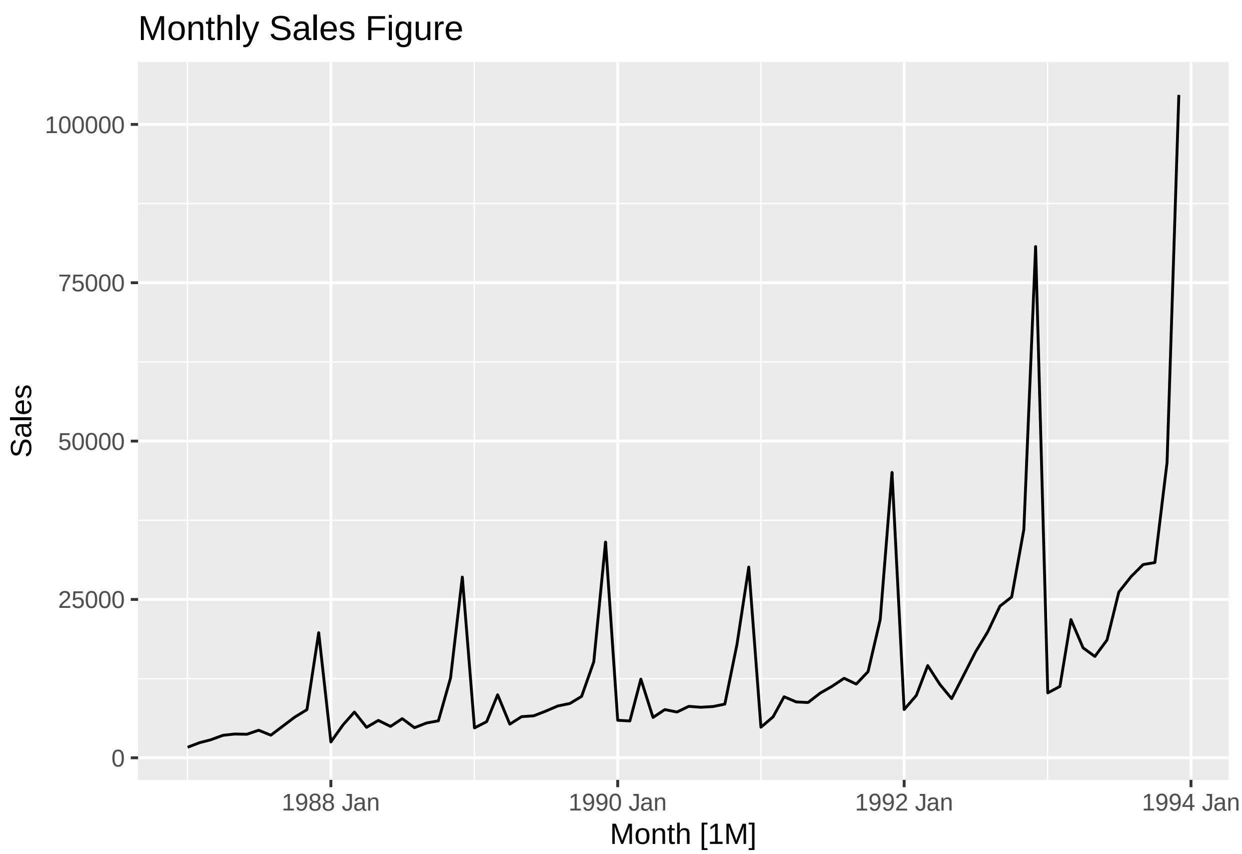
i)

Since our model shows autocorrelation, one way of dealing with it is to add more meaningful variables.

Code

a)

```r
souvenirs %>%
  autoplot(Sales) +
  labs(title = "Monthly Sales Figure")
```



b)

```
souvenirs$Sales <- log(souvenirs$Sales)
souvenirs
```

```
## # A tsibble: 84 x 2 [1M]
##       Month Sales
##       <mth> <dbl>
##  1 1987 Jan  7.42
##  2 1987 Feb  7.78
##  3 1987 Mar  7.95
##  4 1987 Apr  8.17
##  5 1987 May  8.23
##  6 1987 Jun  8.22
##  7 1987 Jul  8.38
##  8 1987 Aug  8.18
##  9 1987 Sep  8.52
## 10 1987 Oct  8.77
## # ... with 74 more rows
```

c)

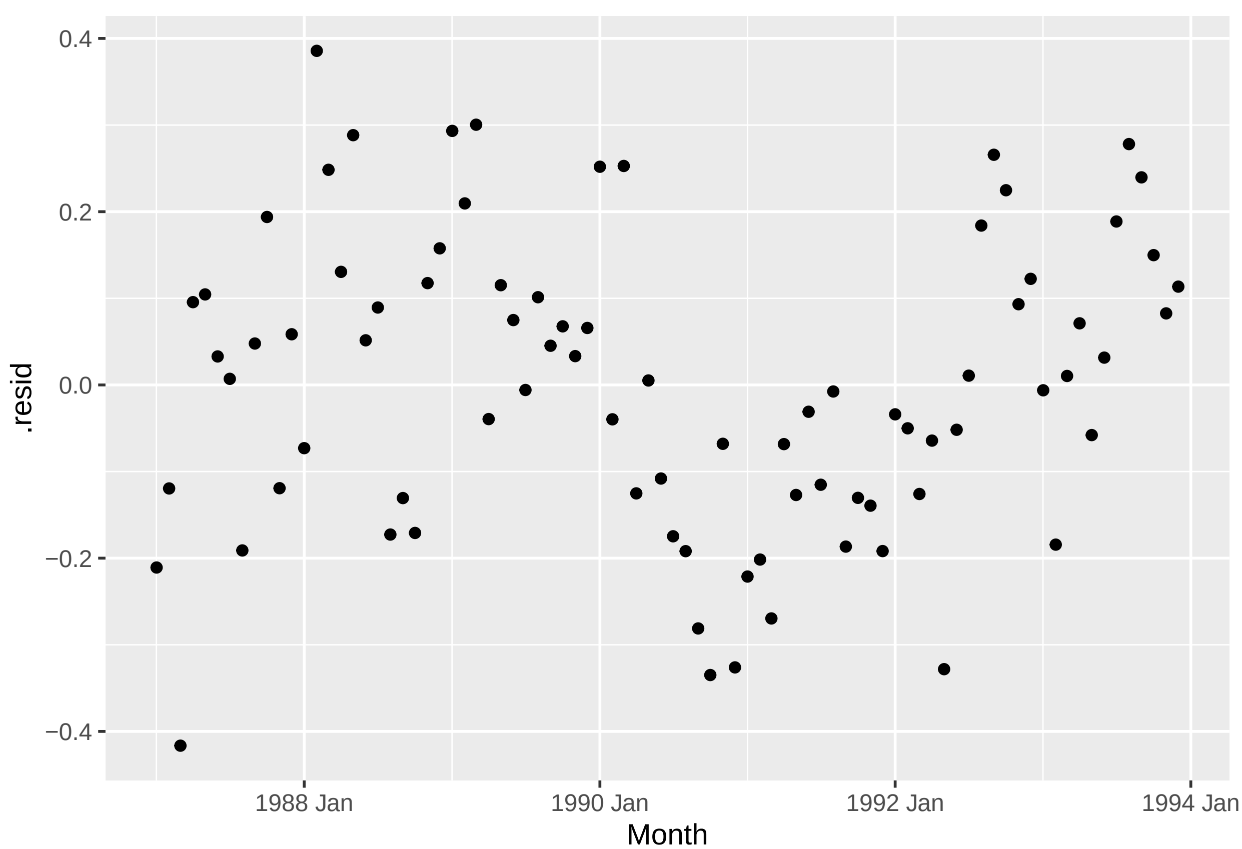
```
souvenirs$surf <- ifelse(month(souvenirs$Month) == 3, 1, 0)
head(souvenirs)
```

```
## # A tsibble: 6 x 3 [1M]
##      Month Sales  surf
##      <mth> <dbl> <dbl>
## 1 1987 Jan  7.42     0
## 2 1987 Feb  7.78     0
## 3 1987 Mar  7.95     1
## 4 1987 Apr  8.17     0
## 5 1987 May  8.23     0
## 6 1987 Jun  8.22     0
```

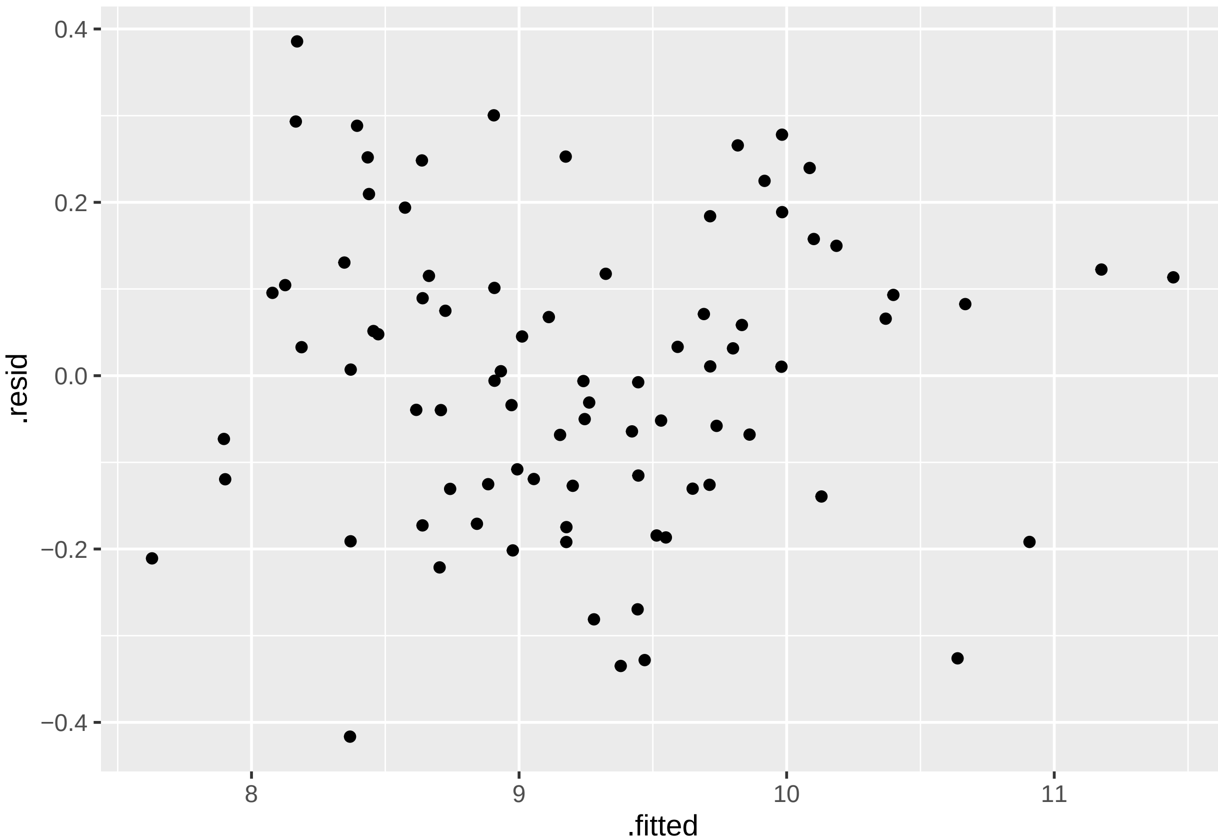
```
fit <- souvenirs %>%
  model(TSLM(Sales ~ trend() + season()))
```

d)

```
augment(fit) %>%
  ggplot(aes(x = Month, y = .resid)) +
  geom_point()
```

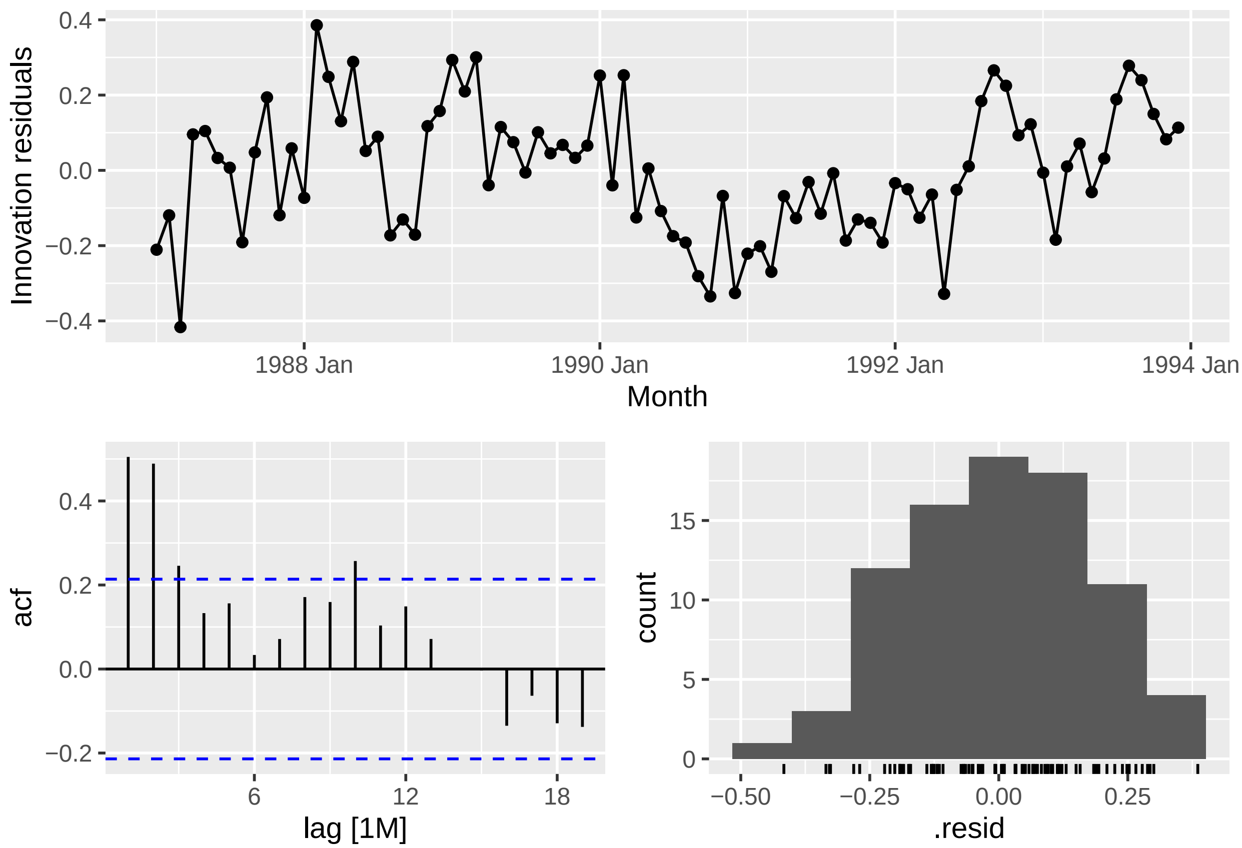


```
augment(fit) %>%
  ggplot(aes(x = .fitted, y = .resid)) +
  geom_point()
```



e)

```
fit %>%
  gg_tsresiduals()
```



f)

```
tidy(fit)
```

```
## # A tibble: 13 x 6
##    .model                           term        estim~1 std.e~2 stati~3  p.value
##    <chr>                            <chr>         <dbl>   <dbl>   <dbl>    <dbl>
##  1 TSLM(Sales ~ trend() + season()) (Intercept)  7.61   7.69e-2   98.9  8.19e-78
##  2 TSLM(Sales ~ trend() + season()) trend()      0.0224 8.45e-4   26.5  1.48e-38
##  3 TSLM(Sales ~ trend() + season()) season()ye~  0.251  9.93e-2    2.53 1.37e- 2
##  4 TSLM(Sales ~ trend() + season()) season()ye~  0.695  9.93e-2    7.00 1.18e- 9
##  5 TSLM(Sales ~ trend() + season()) season()ye~  0.383  9.94e-2    3.85 2.52e- 4
##  6 TSLM(Sales ~ trend() + season()) season()ye~  0.408  9.94e-2    4.11 1.06e- 4
##  7 TSLM(Sales ~ trend() + season()) season()ye~  0.447  9.94e-2    4.50 2.63e- 5
##  8 TSLM(Sales ~ trend() + season()) season()ye~  0.608  9.95e-2    6.12 4.69e- 8
##  9 TSLM(Sales ~ trend() + season()) season()ye~  0.585  9.95e-2    5.88 1.21e- 7
## 10 TSLM(Sales ~ trend() + season()) season()ye~  0.666  9.96e-2    6.69 4.27e- 9
## 11 TSLM(Sales ~ trend() + season()) season()ye~  0.744  9.96e-2    7.47 1.61e-10
## 12 TSLM(Sales ~ trend() + season()) season()ye~  1.20   9.97e-2   12.1  7.22e-19
## 13 TSLM(Sales ~ trend() + season()) season()ye~  1.96   9.98e-2   19.6  2.12e-30
## # ... with abbreviated variable names 1: estimate, 2: std.error, 3: statistic
```

g)

```
augment(fit) %>%
  features(.innov, ljung_box, lag = 10)
```

```
## # A tibble: 1 x 3
##   .model                           lb_stat lb_pvalue
##   <chr>                              <dbl>     <dbl>
## 1 TSLM(Sales ~ trend() + season())    64.7  4.60e-10
```

h)

```
fit %>%
  forecast(h = 36) %>%
  hilo(level = 95) %>%
  mutate(
    lower = `95%`$lower,
    upper = `95%`$upper
  )
```

```
## # A tsibble: 36 x 7 [1M]
## # Key:       .model [1]
##    .model         Month         Sales .mean                        `95%` lower upper
##    <chr>          <mth>        <dist> <dbl>                       <hilo> <dbl> <dbl>
##  1 TSLM(Sales~ 1994 Jan N(9.5, 0.041)  9.51 [ 9.111891,  9.906636]95  9.11  9.91
##  2 TSLM(Sales~ 1994 Feb N(9.8, 0.041)  9.78 [ 9.385328, 10.180073]95  9.39 10.2
##  3 TSLM(Sales~ 1994 Mar  N(10, 0.041) 10.2  [ 9.851884, 10.646628]95  9.85 10.6
##  4 TSLM(Sales~ 1994 Apr  N(10, 0.041)  9.96 [ 9.562004, 10.356749]95  9.56 10.4
##  5 TSLM(Sales~ 1994 May  N(10, 0.041) 10.0  [ 9.609458, 10.404202]95  9.61 10.4
##  6 TSLM(Sales~ 1994 Jun  N(10, 0.041) 10.1  [ 9.670819, 10.465563]95  9.67 10.5
##  7 TSLM(Sales~ 1994 Jul  N(10, 0.041) 10.3  [ 9.854465, 10.649209]95  9.85 10.6
##  8 TSLM(Sales~ 1994 Aug  N(10, 0.041) 10.3  [ 9.853995, 10.648739]95  9.85 10.6
##  9 TSLM(Sales~ 1994 Sep  N(10, 0.041) 10.4  [ 9.957380, 10.752124]95  9.96 10.8
## 10 TSLM(Sales~ 1994 Oct  N(10, 0.041) 10.5  [10.057462, 10.852206]95 10.1  10.9
## # ... with 26 more rows
```

Exercise 9.11 Problem 7

Consider aus_airpassengers, the total number of passengers (in millions) from Australian air carriers for the period 1970-2011.

Use ARIMA() to find an appropriate ARIMA model. What model was selected. Check that the residuals look like white noise. Plot forecasts for the next 10 periods. Write the model in terms of the backshift operator.

Answer

a)

The model chosen is ARIMA(0, 2, 1). According to the ACF plot, the residuals look like white noise. However, the histogram of residuals are skewed to the right meaning that the prediction intervals would be inaccurate.

b)

Yt = -0.8963 * Et-1 +Et

Code

a)

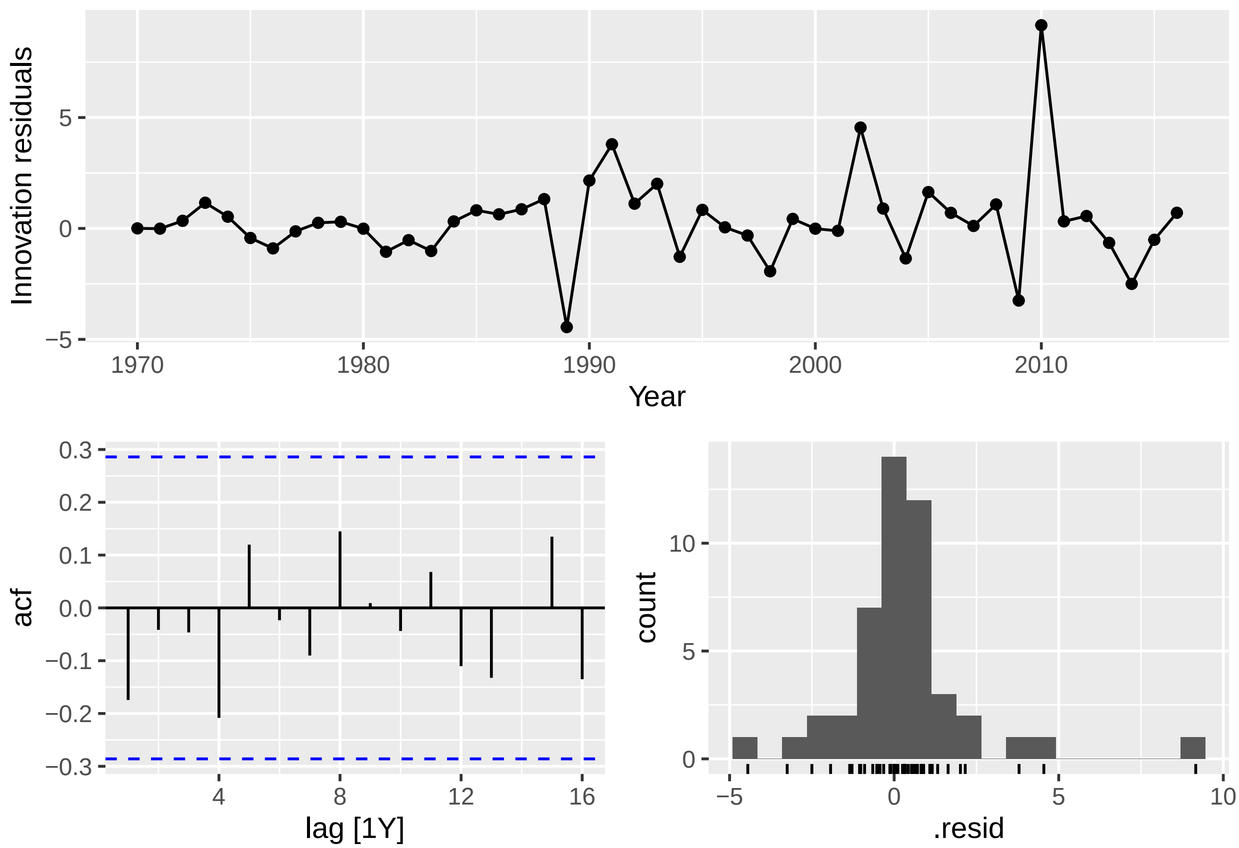
```
head(aus_airpassengers)
```

```
## # A tsibble: 6 x 2 [1Y]
##    Year Passengers
##   <dbl>      <dbl>
## 1  1970       7.32
## 2  1971       7.33
## 3  1972       7.80
## 4  1973       9.38
## 5  1974      10.7
## 6  1975      11.1
```

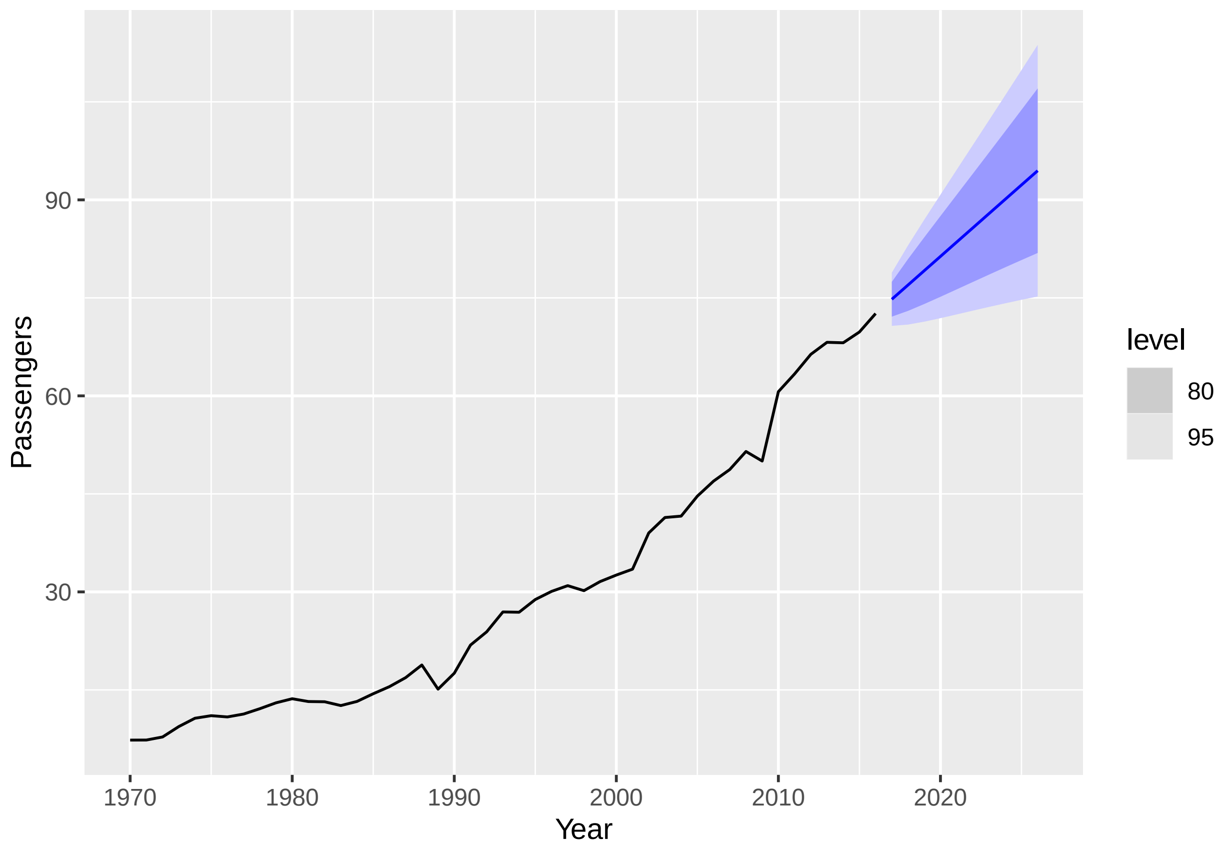
```
fit <- aus_airpassengers %>%
  model(ARIMA(Passengers))

report(fit)
```

```
## Series: Passengers
## Model: ARIMA(0,2,1)
##
## Coefficients:
##           ma1
##       -0.8963
## s.e.   0.0594
##
## sigma^2 estimated as 4.308:  log likelihood=-97.02
## AIC=198.04   AICc=198.32   BIC=201.65
```

```
fit %>%
  gg_tsresiduals()
```




```
fit %>%
  forecast(h = 10) %>%
  autoplot(aus_airpassengers)
```



90
60
30
Passengers
1970
1980
1990
2000
2010
2020
Year
level
80
95