Solutions to Homework Assignment #3

   (a) You may specify one of the following two models. The first model is multiplicative seasonal model
   \[ r_t = 0.0122 + (1 + 0.213B)(1 + 0.125B^{12})a_t, \quad \sigma_a^2 = 0.00415. \]
   This model is adequate; see model checking in Figure 1.
   Another model is
   \[ r_t = 0.0122 + (1 + 0.206B + 0.120B^{12})a_t, \quad \sigma_a^2 = 0.00415. \]
   This MA(12) model is also adequate.
   (b) The fitted model is
   \[ r_t = 0.0066 + 0.066I_{t^{(jan)}} + a_t + 0.211a_{t-1}, \quad \sigma_a^2 = 0.00387. \]
   (c) Yes, the model is adequate; see model checking in Figure 2.
   (d) The regression model is preferred based on the AIC criterion.
   (e) The two models are essentially the same. The RMSE slightly prefers the multiplicative seasonal model, yet the MAE selects the regression model.

   (a) Using the `auto.arima` command, an ARIMA(1,1,1) model specified.
   (b) The fitted model is
   \[ (1 - 0.603B)(1 - B)x_t = (1 - 0.757B)a_t, \quad \sigma_a^2 = 3.58. \]
   Model checking indicates the model is adequate, except for some potential outliers. See Figure 3
   (c) The predictions are

(a) The model selected in ARIMA(1,1,2). The model checking is shown in Figure 4. There appear to have some outliers in the data.

(b) Let \( r_t \) denote the bond return. The fitted model is

\[
(1 + 0.445B)(1 - B)r_t = (1 - 0.281B - 0.588B^2)a_t, \quad \sigma_a^2 = 0.136.
\]

(c) The two largest outliers, based on iterated procedure, are at \( t = 232 \) and 251, respectively. With the two outliers, the AR coefficient becomes insignificant. Therefore, the refined model

\[
(1 - B)(r_t - 3.537I_t^{(232)} - 2.549I_t^{(251)}) = (1 - 0.757B - 0.141B^2)a_t,
\]

where \( \sigma_a^2 = 0.107 \).

(d) The predictions are

\[
> \text{predict(m4c,newxreg=xx,4)}
\]

\$pred
Time Series:
Start = 661
End = 664
Frequency = 1
[1] 0.009943769 0.009899984 0.009899984 0.009899984

\$se
Time Series:
Start = 661
End = 664


(a) An ARIMA(3,0,2) model is selected. The fitted model is

$$(1 - 0.439B - 0.245B^2 + 0.155B^3)y_t = (1 - 0.291B - 0.304B^2)a_t,$$

where $\sigma_a^2 = 0.00189$.

(b) The model is not adequate as it contains some large outliers. See Figure 5.

(c) The four largest outliers are at $t = 1105, 955, 1085,$ and $1839$. The fitted model becomes

$$(1 - 1.29B + 0.61B^2 + 0.02B^3)(y_t + 0.64I_t^{(1105)} + 0.55I_t^{(955)} - 0.39I_t^{(1085)} - 0.39I_t^{(1839)})$$

$$= (1 - 1.13B + 0.46B^2)a_t, \quad \sigma_a^2 = 0.0015.$$ 

The AR-3 coefficient is not statistically significant, but dropping it marks substantial increase is the AIC value. Thus, the coefficient is kept.

(d) The Ljung-Box statistics show $Q(10) = 17.49$ with $p$-value 0.06. There is no sufficient evidence to reject the null hypothesis of zero serial correlations of the 5% level. [Strictly speaking, one should adjusted degrees of freedom of $Q(10)$ based on the number of parameters used. For simplicity, I did not make any adjustment.]

5. Log return series

(a) An ARIMA (3,0,1) model is selected. The model is

$$(1 - 0.72B + 0.01B + 0.07B^3)d_t = (1 - 0.66B)a_t, \quad \sigma_a^2 = 0.0032.$$ 

(b) The model also contains some large outliers. Thus, it is not adequate. See Figure 6.

(c) The model with two largest outliers is

$$(1 - 0.67B - 0.02B^2 + 0.09B^3)(d_t + 0.66I_t^{(1105)} + 0.42I_t^{(955)})$$

$$= (1 - 0.62B)a_t, \quad \sigma_a^2 = 0.0030.$$ 

(d) The Ljung-Box statistics show $Q(10) = 14.65$ with $p$-value 0.15. Thus, there is not evidence to reject the null hypothesis of no serial correlations.

(e) The log transformation seems to reduce the number of outliers in the data.
Figure 1: Model checking of the seasonal model in problem 1, part (a).

Figure 2: Model checking of the regression model of problem 1.
Figure 3: Model checking of the ARIMA(1,1,1) model for VIX index.

Figure 4: Model checking of the model of part (a) of problem 3.
Figure 5: Model checking of the model of part (a) of problem 4.

Figure 6: Model checking of the model of part (a) of problem 5.