Solutions to Homework Assignment #3

You may read the R output file that provides more details on the empirical data analysis.

   (a) Build a pure MA model for the return series. Remove any parameter estimate with \( t \)-ratio less than 1.645 in absolute value. Write down the fitted MA model.
   \textbf{Answer:} Sample ACF suggests an MA(5) model. Since the \( t \)-ratio of the sample mean is 1.75, which is greater than 1.645, so constant is kept. The fitted model is
   \[ r_t = 5 \times 10^{-4} + a_t - 0.06a_{t-1} - 0.09a_{t-3} - 0.12a_{t-5}, \quad \sigma^2 = 9.18 \times 10^{-5}. \]
   (b) Is the MA model adequate? Why?
   \textbf{Answer:} Yes, model checking indicates the model is adequate. See Figure 1.
   (c) Build a pure AR model for the return series. Remove any parameter estimate with \( t \)-ratio less than 1.645 in absolute value. Write down the fitted AR model.
   \textbf{Answer:} AIC selects an AR(5) model. The fitted model is
   \[ (1 + 0.06B + 0.09B^3 + 0.14B^5)(r_t - 5 \times 10^{-4}) = a_t, \quad \sigma^2 = 9.15 \times 10^{-5}. \]
   (d) Is the AR model adequate? Why?
   \textbf{Answer:} Yes, the model is also adequate. See Figure 2.
   (e) Which model do you prefer using the in-sample fit? Why?
   \textbf{Answer:} The AR model is preferred by the AIC. \((-6489.1 \text{ versus } -6486.24)\).
   (f) Which model do you prefer using out-of-sample predictions? You may answer the question by using the last 50 data points for prediction.
   \textbf{Answer:} The MA(5) model is preferred as it produces lower RMSE and MAE.

2. Monthly market liquidity measure of Professors Pastor and Stambaugh.
   (a) Build a time series model for \( x_t \) (the mean equation). Remove any estimates with \( t \)-ratio less than 1.96 in absolute value and re-fit the model. Write down the fitted model.
   \textbf{Answer:} AIC selects an AR(10) model. The fitted model is
   \[ (1 + 0.08B - 0.12B^9 - 0.11B^{10})(r_t - 0.0041) = a_t, \quad \sigma^2 = 0.0012. \]
   Note that: I use an AR model. You can use an MA(10) model too.
(b) Is the model adequate? Why?

**Answer:** The model fits the data well, except for possible outliers. See Figure 3.

(c) Identify the largest outlier in the series. Refine the fitted model by using an indicator for the outlier. Write down the refined model.

**Answer:** The largest outlier occurred at $t = 493$. Let $I_t^{(493)}$ be the dummy variable for the time index $t = 493$, i.e.

$$I_t^{(493)} = \begin{cases} 
1 & \text{if } t = 493, \\
0 & \text{otherwise.}
\end{cases}$$

The model with outlier is

$$(1 - 0.10B^9 - 0.12B^{10})(r_t - 0.0038 - 0.19I_t^{(493)}) = a_t, \quad \sigma^2 = 0.0011. \quad \tag{1}$$

Model checking indicates no other outlier in the model. See Figure 4.

If the MA(10) model is used, the model becomes

$$r_t = 0.0037 + 0.19I_t^{(493)} + a_t + 0.16a_{t-9} + 0.13a_{t-10}, \quad \sigma^2 = 0.0011. \quad \tag{2}$$

3. Quarterly earnings per share of Caterpillar from the second quarter of 1983 to the second quarter of 2014.

(a) Build a time series model for the earnings series. Perform model checking and write down the fitted model.

**Answer:** Based on the ACF of the first differenced data, I specified an ARIMA(0,1,1)(1,0,1)$_4$ model for the earnings. The fitted model is

$$(1 - 0.99B^4)(1 - B)x_t = (1 - 0.13B)(1 - 0.98B^4)a_t, \quad \sigma^2 = 0.043. \quad \tag{3}$$

This model is adequate with possible some minor outliers. See Figure 5.

(b) Fit the following model to the earnings series:

```r
m5=arima(xt,order=c(0,1,0),seasonal=list(order=c(0,1,1),period=4))
```

where `xt` denotes the earnings series. Write down the fitted model.

**Answer:** The fitted model is

$$(1 - B)(1 - B^4)x_t = (1 - 0.95B^4)a_t, \quad \sigma^2 = 0.044. \quad \tag{4}$$

(c) Compare the two time series models. Which model is preferred in terms of in-sample fitting? Why?

**Answer:** The AIC selects the model in part (a).
(d) Use the backtest procedure to compare the two models via 1-step ahead forecasts. You may use \( t = 101 \) as the starting forecast origin. Which model is preferred? Why?

**Answer:** The model in part (b), i.e. \( m5 \), is selected as it has smaller RMSE and MAE in out-of-sample forecasts.


(a) Fit the linear regression model

\[
y_t = \alpha + \beta x_t + e_t.
\]

Write down the fitted model. What is the \( R^2 \)? Is the model adequate? Why?

**Answer:** The model is

\[
y_t = 0.59 + 0.0295 x_t + e_t, \quad \sigma = 0.17.
\]

The \( R^2 \) of the model is 96.84%. But the model is not adequate. The residuals have strong serial correlations with \( Q(10) = 6546.07 \).

(b) Let \( d_t = (1 - B)x_t \) and \( c_t = (1 - B)y_t \), where \( B \) is the back-shift operator. Here \( c_t \) and \( d_t \) denote the change in weekly gas and oil price, respectively. Consider the linear regression

\[
c_t = \beta d_t + e_t.
\]

Write down the fitted model. What is the \( R^2 \)?

**Answer:** The model is

\[
c_t = 0.0122 d_t + e_t, \quad \sigma = 0.039.
\]

The \( R^2 \) of the linear regression is 30.95%. The residuals of the model in part (a) has serial correlations close to 1 and are decaying slowly.

(c) Is the model in part (b) adequate? If not, refine the model and write down the refined model.

**Answer:** The model is not adequate either as its residuals have significant serial correlations. In addition, the model contains some outliers. After several iterations, we obtain a model with three outliers. The model is

\[
(1 - 0.63B)(c_t - 0.0072d_t - 0.49I_t^{(763)} - 0.25I_t^{(921)} - 0.17I_t^{(925)}) = a_t, \quad \sigma^2 - 8.22 \times 10^{-4}.
\]

This model is adequate. See Figure 6.
Based on the refined model, describe the linear dependence between gasoline price and oil price.

**Answer**: Based on the fitted model, the change in weekly regular gasoline price depends on the change in weekly oil price. The linear dependence has a coefficient 0.0072 after adjusting the effects of lag-1 serial dependence and three outliers.

5. Consider again the gasoline and oil prices of Problem 4. Suppose that one is concerned with taking the first difference. To mitigate the concern, one can perform the first two analyses below:

(a) Perform a proper unit-root test on the weekly gasoline price. Draw the conclusion.

**Answer**: The ADF test with lag = 9 and drift has a test statistic $-1.83$ with $p$-value 0.38. The unit root is not rejected.

(b) Perform a proper unit-root test on the weekly crude oil price. Draw the conclusion.

**Answer**: The ADF test with lag = 8 and drift has a test statistic $-2.11$ with $p$-value 0.27. The unit-root hypothesis cannot be rejected at the 5% level.

(c) Build a time series model for the gasoline price series. Check the adequacy of the model. Does the model imply any business cycles in the gas price? Why?

**Answer**: AIC selects an AR(9) model for the differenced series. Thus, we have an ARIMA(9,1,0) model for the gasoline price. Model checking shows that the model has several outliers. After some iterations, the model is

\[
(1 - 0.82B - 0.06B^3 - 0.05B^8 + 0.08B^9)(1 - B)(r_t - 0.27I_t^{(764)} + 0.09I_t^{(766)})
-0.14I_t^{(922)} + 0.06I_t^{(926)}) = a_t, \quad \sigma^2 = 0.0011.
\]

This model is adequate and its AR polynomial contains complex roots. Therefore, the model implies that there exists business cycles in the gasoline price.
Figure 1: Model checking of an MA(5) model for the daily S&P composite returns.

Figure 2: Model checking of an AR(5) model for the daily S&P composite returns.
Figure 3: Model checking of an AR(10) model for the VWF series.

Figure 4: Model checking of an AR(10) model for the VWF series. with an outlier.
Figure 5: Model checking for the model in Part (a) for the quarterly earnings per share of Caterpillar stock.

Figure 6: Model checking for the final model of weekly changes in gasoline prices with three outliers.