Solutions to Exercises of Chapter 2

Problem 1. From the model, $R_{101} = a_{101} + 0.2a_{100}$. Taking conditional expectation at $t = 100$, we have

$$R_{100}(1) = 0.2a_{1000} = 0.2 \times 0.01 = 0.002$$

The associated forecast error is $e_{100}(1) = R_{101} - R_{100}(1) = a_{101}$. Therefore, the standard deviation of the forecast error is $\sqrt{\text{Var}[e_{100}(1)]} = \text{std}(a_{101}) = 0.025$. For 2-step ahead forecast, we have

$$R_{102} = a_{102} + 0.2a_{101},$$

and $R_{100}(2) = E(R_{102}|F_{100}) = 0$. The forecast error is

$$e_{100}(2) = R_{102} - R_{100}(2) = a_{102} + 0.2a_{101}$$

with standard deviation $\sqrt{\text{Var}[a_{102} + 0.2a_{101}]} = \sqrt{1.04} \sigma_a^2 = 0.0255$. To compute ACF of $R_t$, use the model to obtain

- $\text{Var}(R_t) = (1 + 0.2^2) \sigma_a^2 = 0.00065$.
- $\text{Cov}(R_t, R_{t-1}) = \text{Cov}(a_t + 0.2a_{t-1}, a_{t-1} + 0.2a_{t-2}) = \text{Cov}(0.2a_{t-1}, a_{t-1}) = 0.2 \sigma_a^2 = 0.000125$.
- $\text{Cov}(R_t, R_{t-\ell}) = 0$ for $\ell \geq 2$.

Therefore, ACF of $R_t$ is $\rho_1 = 0.192$, $\rho_\ell = 0$ for $\ell \geq 2$.

Problem 2. Taking expectation of the model, we have

$$E(r_t) = 0.01 + 0.2E(r_{t-2}).$$

Therefore, $E(r_t) = \frac{0.01}{1-0.2} = 0.0125$. Taking variance of the model, we obtain

$$\text{Var}(r_t) = 0.04 \text{Var}(r_{t-2}) + \sigma_a^2.$$ 

Therefore, $\text{Var}(r_t) = \frac{0.02}{1-0.06} = 0.0208$. For autocorrelation function, we may drop the constant term and multiply the equation by $r_{t-\ell}$ to obtain

$$r_tr_{t-\ell} = 0.2r_{t-2}r_{t-\ell} + a_tr_{t-\ell}.$$ 

Taking the expectation for $\ell > 0$, we have $\gamma_\ell = 0.2\gamma_{\ell-2}$, where $\gamma_\ell$ is the lag-$\ell$ autocovariance of $r_t$. Therefore, $\rho_\ell = 0.2\rho_{\ell-2}$ for $\ell > 0$. Since $\rho_0 = 1$, we have $\rho_2 = 0.2, \rho_4 = 0.2^2, \cdots$ for even number $\ell$. For $\ell = 1$, we use $\rho_1 = \rho_{-1}$ to obtain $\rho_1 = 0.2\rho_{-1} = 0.2 \rho_1$. This implies $0.8 \rho_1 = 0$. Therefore, $\rho_\ell = 0$ if $\ell$ is an odd number.
For 1-step ahead forecast at $t = 100$, we have $r_{101} = 0.01 + 0.2r_{99} + a_{101}$. Taking conditional expectation,

$$r_{100}(1) = 0.01 + 0.2r_{99} = 0.01 + 0.2 \times (0.02) = 0.014.$$  

The forecast error is $e_{100}(1) = a_{101}$ with standard deviation $\sqrt{0.02} = 0.141$. For 2-step ahead forecast, using $r_{102} = 0.01 + 0.2r_{101} + a_{102}$, we obtain $r_{100}(2) = 0.01 + 0.2(-0.01) = 0.008$. The associated forecast error is $a_{102}$ with standard deviation $\sqrt{0.02} = 0.141$.

**Problem 3.** Model specification:

- t-ratio of the sample mean = 8.71, indicating the need of a constant term.
- ACF shows significant correlation at lags 1 and 11. For simplicity, focus on $\hat{\rho}_1 = 0.14$ with standard error 0.04. Thus, ACF suggests an MA(1) model.
- PACF also shows significant result at lags 1 and 11. Start with an AR(1) model.

Denote the return by $r_t$. For AR(1) model, estimation gives

$$r_t = 0.004 + 0.143r_{t-1} + a_t, \quad \sigma_a = 0.014$$

where standard errors of the coefficient estimates are 0.001 and 0.038, respectively. Residual ACF shows $Q(10) = 11.3$ with p-value 0.256. Thus, the model is adequate. For MA(1) model, we have

$$r_t = 0.005 + a_t + 0.155a_{t-1}, \quad \sigma_a = 0.014,$$

where standard errors of the coefficients are 0.001 and 0.038, respectively. For residuals, $Q(10) = 9.8$ with p-value 0.367. Again, the MA(1) model is also adequate.

**Problem 4.** Based on the Q-statistics, $Q(5) = 14.5$ (0.013) and $Q(10) = 22.2$ (0.014), where the number in parentheses is the p-value. Thus, at the 5% level, the return is statistically predictable.

**Problem 5.** The test results are

<table>
<thead>
<tr>
<th></th>
<th>HWP</th>
<th>VW</th>
<th>EW</th>
<th>SP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q(10)$</td>
<td>11.7</td>
<td>74.3</td>
<td>829</td>
<td>29.3</td>
</tr>
<tr>
<td>p-value</td>
<td>0.306</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Thus, the individual stock does not have significant serial dependence, but the market indexes are serially correlated.

**Problem 6.** Denote the return by $r_t$. Model specification:

- t-ratio of sample mean = 4.04, indicating the need of a constant term.
• PACF suggests an AR(1) model
• ACF suggests an MA(1) model

Estimation of AR(1) model shows
\[ r_t = 0.826 + 0.227r_{t-1} + a_t, \quad \sigma_a = 5.45 \]
where standard errors of the coefficients are 0.260 and 0.046. Residual Q-statistic gives \( Q(10) = 8.2 \), which is insignificant at the 5\% level. For MA(1) model, we have
\[ r_t = 1.061 + a_t + 0.239a_{t-1}, \quad \sigma_a = 5.44 \]
where standard errors of the coefficients are 0.316 and 0.045. Residual Q-statistic gives \( Q(10) = 6.4 \), which is also insignificant at the 5\% level. The two models are adequate.

Forecasts (standard errors) at the time point \( t = 456 \) are

<table>
<thead>
<tr>
<th>time</th>
<th>AR(1)</th>
<th>MA(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>457</td>
<td>2.610 (5.454)</td>
<td>2.252 (5.439)</td>
</tr>
<tr>
<td>458</td>
<td>1.419 (5.592)</td>
<td>1.061 (5.592)</td>
</tr>
</tbody>
</table>

Model comparison: Express the AR(1) model as
\[ (1 - 0.227B)r_t = 0.826 + a_t. \]
Using long-division, we have
\[ r_t = 1.069 + a_t + 0.227a_{t-1} + 0.052a_{t-2} + 0.012a_{t-3} + 0.003a_{t-4} + \cdots \]
which is very close to the MA(1) model. Thus, the two models are essentially equivalent.

**Problem 7.** Denote the equal-weighted return by \( r_t \). To use AR model, we apply PACF or AIC criterion to select an AR(6) model. The fitted model is
\[
\begin{align*}
r_t &= 0.049 + 0.282r_{t-1} + 0.001r_{t-2} + 0.038r_{t-3} + 0.050r_{t-4} \\
&= +0.065t - 5 + 0.041r_{t-6} + a_t, \quad \sigma_a = 0.621
\end{align*}
\]
where the standard error of the constant term is 0.009 and those of the AR coefficients are about 0.015. Thus, except for the lag-2, all AR coefficients are significant. The model is adequate with \( Q(10) = 7.8 \) for the residuals.

To use ARMA model, the automatic procedure of SCA identifies an ARMA(1,1) model. But the residuals show some serial correlations at lags 5 and 6. I modified the model to take careful of the possibility of weekly effect (5 trading days). The model becomes
\[ (1 - \phi_1 B)r_t = \phi_0 + (1 - \theta B)(1 - \Theta B^5)a_t \]
The fitted model is

\[(1 - 0.404B)r_t = 0.057 + (1 - 0.122B)(1 + 0.075B^5)a_t, \quad \sigma_a = 0.624.\]

All the estimates are significant at the 1% level. This model is adequate even though some minor serial correlations remain in the residuals.

Forecasts (standard errors) of the two models are as follows:

<table>
<thead>
<tr>
<th>time</th>
<th>AR(6) model</th>
<th>ARMA model</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>5053</td>
<td>0.028 (0.621)</td>
<td>0.014 (0.624)</td>
<td>0.434</td>
</tr>
<tr>
<td>5054</td>
<td>0.135 (0.645)</td>
<td>0.107 (0.649)</td>
<td>1.413</td>
</tr>
<tr>
<td>5055</td>
<td>0.165 (0.647)</td>
<td>0.097 (0.653)</td>
<td>0.514</td>
</tr>
<tr>
<td>5056</td>
<td>0.160 (0.648)</td>
<td>0.155 (0.653)</td>
<td>1.645</td>
</tr>
<tr>
<td>5057</td>
<td>0.127 (0.650)</td>
<td>0.069 (0.653)</td>
<td></td>
</tr>
<tr>
<td>5058</td>
<td>0.091 (0.653)</td>
<td>0.090 (0.655)</td>
<td></td>
</tr>
<tr>
<td>5059</td>
<td>0.100 (0.656)</td>
<td>0.093 (0.655)</td>
<td></td>
</tr>
</tbody>
</table>

**Problem 8.** Create indicator (or dummy) variables as follows:

- \( M_t = 1 \) if it is a Monday and \( = 0 \) otherwise.
- \( T_t = 1 \) if it is a Tuesday and \( = 0 \) otherwise.
- \( W_t = 1 \) if it is a Wednesday and \( = 0 \) otherwise.
- \( R_t = 1 \) if it is a Thursday and \( = 0 \) otherwise.

Consider the regression

\[r_t = c_0 + c_1 M_t + c_2 T_t + c_3 W_t + c_4 R_t + e_t.\]

The fitted model is

\[r_t = 0.253 - 0.354M_t - 0.244T_t - 0.093W_t - 0.110R_t + e_t\]

where all estimates are significant at the 1% level. But the residuals show strong serial correlations. The EACF of the residuals suggests an ARMA(1,2) model. I refined the model as

\[r_t = c_0 + c_1 M_t + c_2 T_t + c_3 W_t + c_4 R_t + \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi B} a_t.\]

This is a regression model with time series errors. Parameter estimates and their standard errors are

<table>
<thead>
<tr>
<th>( c_0 )</th>
<th>( c_1 )</th>
<th>( c_2 )</th>
<th>( c_3 )</th>
<th>( c_4 )</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \phi )</th>
<th>( \sigma_a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.256</td>
<td>-0.358</td>
<td>-0.248</td>
<td>-0.096</td>
<td>-0.110</td>
<td>0.597</td>
<td>0.170</td>
<td>0.893</td>
<td>0.608</td>
</tr>
<tr>
<td>0.024</td>
<td>0.024</td>
<td>0.026</td>
<td>0.026</td>
<td>0.024</td>
<td>0.029</td>
<td>0.018</td>
<td>0.024</td>
<td></td>
</tr>
</tbody>
</table>
All estimates are highly significant. In particular, the results indicate the expected return on Friday is different from (or better than) the other days. The Ljung-Box statistics of the residuals give $Q(10) = 18.2$, suggesting no significant serial correlations left in the residuals.

**Problem 9.** Due the small magnitude of the data, I use percentage returns, i.e. $y_t = 100(f_t - f_{t-1})$ and $x_t = 100(s_t - s_{t-1})$. The correlation between $y_t$ and $x_t$ is 0.389. Thus, I start with the simple linear regression model

$$y_t = \beta x_t + e_t.$$ 

The fitted model is $y_t = 0.621x_t + e_t$. The residuals, however, show strong serial dependence, e.g. $Q(10) = 115.0$. I then use EACF to identify an ARMA model for the residual and it suggests clearly an ARMA(1,1) model. Consequently, the model becomes

$$y_t = 0.723x_t + \frac{1 - 0.936B}{1 - 0.820B}a_t$$

where all estimates are highly significant and $\sigma_a = 0.029$. The residual series has no significant serial correlations because $Q(10) = 5.5$, which is low. Note that if one employs autoregressive models only, then he needs a high order model; see Chapter 8 where an AR(8) model is used. This appears to be an excellent example for demonstration the usefulness of using ARMA models. Also, the residuals appear to be heteroscedastic. One can further refine the model by using the models discussed in Chapter 3.

**Problem 10.** Using PACF and ignoring higher order correlations, I start with an AR(3) model. The first four lags of PACF are 0.63, −0.25, −0.16 and −0.13, respectively, with standard error 0.08. The fitted model is

$$y_t = 0.009 + 0.750y_{t-1} - 0.117y_{t-2} - 0.1637y_{t-3} + a_t, \sigma_a = 0.325,$$

where standard errors of the estimates are 0.025, 0.076, 0.095, and 0.076, respectively. The residuals show a significant serial correlation at lag 8, which might be due to the effect of the seasonal adjustment procedure used to process the data. Consider the polynomial $1 - 0.750x + 0.117x^2 + 163x^3$, which can be factored as $(1 + 0.337B)(1 - 1.087B + 0.483B^2)$. For the 2nd factor, the discriminant function is $1.087^2 - 4 \times 0.483 = -0.75 < 0$. Thus, the fitted model supports the existence of business cycles in the U.S. quarterly unemployment rate. The average period of the cycles is $k = 2\pi/\cos^{-1}[1.087/(2 \times \sqrt{0.483})] \approx 9.3$ quarters. [Note the same conclusion is reached if an AR(4) model is used.]

**Problem 11.** Denote the time series by $x_t$. Examining the sample ACF of $x_t$ and $(1 - B)x_t$ shows that second-order difference is needed, i.e. the ACF of $(1 - B)x_t$ remains high, for instance, $\hat{\rho}_{24} = 0.50$. Thus, we employ $y_t = (1 - B)^2x_t$ in the analysis. The
sample ACF of $y_t$ and the automatic procedure of SCA all identify an MA(1) model for $y_t$. Consequently, the model for $x_t$ is

$$(1 - B)^2 x_t = (1 - 0.448B)a_t, \quad \sigma_a = 0.147,$$

where the standard error of the coefficient is 0.061. The residuals show no significant serial correlations as $Q(12) = 16.3$, which corresponds to a p-value 0.13.