Solutions to Homework Assignment 4

**Note:** Sincere there are no true answers to empirical time series modeling, the solutions are for your reference only. You may obtain models that are different from mine.

1. `7cityMonth.txt` data: The data have strong and stable seasonal patterns. Seasonal VARMA models are employed.
   
   (a) Nanjing and Shanghai. Let $z_t$ be the bivariate time series. The final model used is
   \[
   \left( I - \begin{bmatrix} 0.182 & 0 \\ 0 & 0.242 \end{bmatrix} \right) (1 - B^{12}) z_t = \left( I - \begin{bmatrix} 0.872 & 0 \\ 0 & 0.882 \end{bmatrix} \right) a_t
   \]
   where the covariance matrix of $a_t$ is
   \[
   \begin{bmatrix} 1.51 & 1.34 \\ 1.34 & 1.41 \end{bmatrix}.
   \]
   Model checking indicates the model fits the series well. From the model, except for contemporaneous correlation, the two temperature series have no cross dynamic dependence.

   (b) Anqing and Hangzhou. Again, let $z_t$ be the bivariate time series. The final model used is
   \[
   \left( I - \begin{bmatrix} 0.199 & -0.130 \\ 0 & 0.101 \end{bmatrix} \right) (1 - B^{12}) z_t = \left( I - \begin{bmatrix} 0.895 & 0 \\ 0 & 0.919 \end{bmatrix} \right) a_t
   \]
   where the covariance matrix of $a_t$ is
   \[
   \begin{bmatrix} 1.71 & 1.48 \\ 1.48 & 1.55 \end{bmatrix}.
   \]
   Model checking fails to reject the model. In this particular case, the temperature of Anqing depends on the lagged-1 value of Hangzhou, but Hangzhou’s temperature is not affected by that of Anqing. Perhaps the wind from East China Sea plays a role.

2. De-seasoned temperature series of Shanghai, Anqing, and Hangzhou. Let $x_t$ be the 3-dimensional series. A VAR(3) model is employed and the final model is given below:

   \[
   \begin{array}{c}
   \text{AR(1)-matrix} \\
   \\
   \begin{bmatrix}
   [1,] & 0.3787 & 0.2046 & -0.325 \\
   [2,] & 0.1606 & 0.2279 & -0.261
   \end{bmatrix}
   \end{array}
   \]
3. clothing.csv data. Used the first 1805 observations in the analysis. The data contain some large outliers, but I did not handle them specifically. Let $z_t$ be the bivariate time series, the final model fitted is

$$(I - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^7)z_t = (I - \theta_1 B - \theta_2 B^2)(I - \theta_7 B^7)a_t.$$  

The model still encounter some cross-correlations, especially at the seasonal lags. For parameter estimates, see the associated R output.

4. Let $z_t$ be the first three PCA series. In this particular case, it turns out that the third component only needs the seasonal difference, but the first two components require regular
and seasonal difference. Consequently, we need to treat the components differently. That is, 
\[ y_{1t} = (1 - B)(1 - B^7)z_{1t}, \quad y_{2t} = (1 - B)(1 - B^7)z_{2t}, \quad \text{and} \quad y_{3t} = (1 - B^7)z_{3t}. \]
For the \( y_t \) series, I employed the model

\[
(I - \phi_1B - \phi_2B^2 - \phi_3B^3 - \phi_4B^4)(I - \phi_7B^7)y_t = (I - \theta_7B^7)a_t.
\]

The parameter estimates are given below:

Estimates in matrix form:

Regular AR coefficient matrix

\begin{align*}
\text{AR(1)-matrix} & \\
[1,] & 0.00000 & 0.242 & 0.0532 \\
[2,] & -0.01892 & -0.429 & -0.0692 \\
[3,] & -0.00847 & -0.117 & 0.2602 \\
\end{align*}

\begin{align*}
\text{AR(2)-matrix} & \\
[1,] & -0.2232 & -0.1042 & 0.0000 \\
[2,] & 0.0167 & -0.2098 & 0.0489 \\
[3,] & -0.0165 & -0.0831 & 0.1962 \\
\end{align*}

\begin{align*}
\text{AR(3)-matrix} & \\
[1,] & -0.173 & 0.0000 & 0.0000 \\
[2,] & 0.000 & -0.1636 & 0.0416 \\
[3,] & -0.020 & -0.0776 & 0.1278 \\
\end{align*}

\begin{align*}
\text{AR(4)-matrix} & \\
[1,] & -0.2103 & -0.2069 & 0.1252 \\
[2,] & 0.0170 & -0.0905 & 0.0000 \\
[3,] & 0.0146 & 0.0000 & 0.0996 \\
\end{align*}

Seasonal AR coefficient matrix

\begin{align*}
\text{AR(7)-matrix} & \\
[1,] & 0.00000 & 0.000 & 0 \\
[2,] & 0.00000 & -0.031 & 0 \\
[3,] & 0.00101 & 0.000 & 0 \\
\end{align*}

Seasonal MA coefficient matrix

\begin{align*}
\text{MA(7)-matrix} & \\
[1,] & 0.988306631 & -0.05182412 & -0.0406722 \\
[2,] & 0.000000000 & 0.97176130 & 0.0000000 \\
[3,] & 0.006310236 & 0.01715135 & 0.9223127 \\
\end{align*}

Residuals cov-matrix:

\begin{align*}
\text{resi} & \quad \text{resi} & \quad \text{resi} \\
\text{resi} & 0.87245970 & -0.038733827 & -0.008972650 \\
\end{align*}
resi -0.03873383 0.086447334 0.002035884
resi -0.00897265 0.002035884 0.122067467

It still needs some further improvements, but the difficulties might be caused by outliers.

5. The last two PCAs. Let \( z_t \) be the last two principal components. A VAR(5) model fits the series well. The parameter estimates are

\[
\begin{align*}
\text{AR coefficient matrix} \\
\text{AR(1)-matrix} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0.2861 & 0.000 \\
0.0211 & 0.301
\end{bmatrix} \\
\text{standard error} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0.0233 & 0.0000 \\
0.0201 & 0.0234
\end{bmatrix} \\
\text{AR(2)-matrix} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0.161 & -0.0437 \\
0.000 & 0.0890
\end{bmatrix} \\
\text{standard error} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0.0238 & 0.0241 \\
0.0000 & 0.0246
\end{bmatrix} \\
\text{AR(3)-matrix} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0 & 0.116 \\
0 & 0.0000
\end{bmatrix} \\
\text{standard error} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0 & 0.0000 \\
0 & 0.0238
\end{bmatrix} \\
\text{AR(4)-matrix} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0.079 & 0 \\
0.000 & 0
\end{bmatrix} \\
\text{standard error} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0.0238 & 0 \\
0.0000 & 0
\end{bmatrix} \\
\text{AR(5)-matrix} \\
&\begin{bmatrix}
[1,1] & [1,2] \\
[2,1] & [2,2]
\end{bmatrix} \\
&\begin{bmatrix}
0.0434 & -0.0356 \\
0.0212 & 0.0454
\end{bmatrix} \\
\text{standard error}
\end{align*}
\]
Based on the fitted model, $z_t$ does not have seasonal pattern with seasonality 7. Consequently, we found at least two linear combinations of the original seasonal data that have no seasonality, implying common seasonalities exist in the original data.