Lecture 9: Diagnostic Checking

Bus 41910, Time Series Analysis, Mr. R. Tsay

We now discuss the last stage of Box-Jenkins iterative modeling procedure, namely, model checking or diagnostic checking. The objective of this stage is two-fold. First, it checks for possible discrepancy of an entertained model. Secondly, it corrects any model discrepancy if found.

Model checking in time series analysis is very much the same as that of the traditional regression analysis. It emphasizes on residual analysis. Basically, one considers the following statistics:

- Residual plots
- Residual serial correlations
- Outlier detection.

For an entertained ARIMA(p, q) model (or ARIMA(p, d, q) model), the residual series $\hat{a}_t$ is defined by

$$\hat{a}_t = Z_t - \hat{\phi}_1 Z_{t-1} - \cdots - \hat{\phi}_p Z_{t-p} + \hat{\theta}_1 \hat{a}_{t-1} + \cdots + \hat{\theta}_q \hat{a}_{t-q}$$

where the mean of Z_t is assumed to be zero, $\hat{\phi}$'s and $\hat{\theta}$'s are MLE of ϕ's and θ's, respectively, and the starting residuals $\hat{a}_t$ for $t \leq 0$ are defined by the method of estimation used. For instance, $\hat{a}_t = 0$ for $t \leq 0$ under conditional MLE.

There are various ways to plot the residuals. For instance, the time plot is useful in spotting possible serial correlation, non-constant variance, and outliers and normal-score plot (or histogram) is used to check the normality assumption. Residual plots are integral parts of time series analysis.

To check for residual serial correlation, one can use residual ACF and PACF. For an adequate model, the residuals should behave as white noise. The residual ACF at lag-ℓ is defined by

$$\hat{r}_\ell = \frac{\sum_{t=\ell+1} (\hat{a}_t - \bar{a})(\hat{a}_{t-\ell} - \bar{a})}{\sum_{t=1}^{n} (\hat{a}_t - \bar{a})^2}$$

where $\bar{a}$ is the average of the residuals $\hat{a}_t$'s. For a white noise series, the asymptotic variance of the sample ACF is $\frac{1}{n}$. Therefore, one can use $\frac{1}{\sqrt{n}}$ as an approximate standard error to check the significance of individual $\hat{r}_\ell$. More efficiently, one may wish to check simultaneously that several residual ACFs are not significantly different from zero, that is to say, in hypothesis testing, to test the null hypothesis

$$H_0 : r_1 = r_2 = \cdots = r_m = 0$$

where r_i is the theoretical ACF of a_t at lag i. To this end, Ljung and Box (1978, BKA) propose a modified Box-Pierce statistic:

$$Q(m) = n(n+2)\sum_{v=1}^{m} \frac{\hat{r}_v^2}{n-v}$$

which, under the null hypothesis that the entertained ARMA(p,q) model is the "true" model, is asymptotically χ^2_{m-p-q}. This is the Q statistics of ACF computed by SCA.
Remarks: The Ljung-Box $Q(m)$ statistic is asymptotically equivalent to testing for an ARMA(p,q) model against an ARMA$(p,q+m)$ or ARMA$(p+m,q)$ model. This Q statistic is commonly used. However, it is not a powerful statistic in detecting residual series correlation. Of course, its power depends on the choice of m. In the literature, simulation results by Newbold and his associates suggest that $m = O(\ln n)$ is close to an optimal choice.

Next, we consider outlier detection. This problem has gained much attention in the 1980s and various methods are available. In time series analysis, outliers can cause biases in parameter estimation as well as model misspecification, resulting in misleading conclusion. For this reason, several outlier detection and robust estimation procedures have been proposed in the literature for time series analysis. Here we shall consider a simple regression approach which is closely related to the "Intervention Analysis" of Box and Tiao (1976, JASA; also the packet).
Following Fox (1974), there are basically two types of outliers in a time series. The first type of outlier is called the "additive outlier (AO)" which represents that a disturbance is committed to a particular observation. Mathematically, the observed time series is

$$Y_t = Z_t + \omega_a I_t^{(d)}$$

where Z_t is an outlier-free time series, ω_a denotes the magnitude of the disturbance and $I_t^{(d)}$ is an indicator variable defined by

$$I_t^{(d)} = \begin{cases} 1 & \text{if } t = d \\ 0 & \text{if } t \neq d \end{cases}$$

In other words, for an AO model

$$Y_t = Z_t \quad \text{if } t \neq d \quad \text{and} \quad Y_d = Z_d + \omega_a.$$

A typical example of an AO is a typo or a recording error.

Another type of outlier is called an "innovational outlier (IO)", which is a disturbance in the innovational series $\{a_t\}$ and may affect every subsequent observation of the series. Mathematically, an IO model is

$$Y_t = \frac{\theta(B)}{\phi(B)}(a_t + \omega_v I_t^{(d)})$$

where $I_t^{(d)}$ is defined as before and ω_v denotes the magnitude of the disturbance. Rewriting the model as

$$Y_t = Z_t + \frac{\theta(B)}{\phi(B)}\omega_v I_t^{(d)}$$

we see that an IO affects the series through its own dynamic $\frac{\theta(B)}{\phi(B)}$ and, in effect, becomes part of the system thereafter. In practice, an IO often indicates an onset of certain changes in the system. For instance, in a manufacturing process, changing an operator or a measurement instrument may result in an IO.

Of course, many other types of disturbance can happen to a time series. The AO and IO models only two of many possibilities. In Chen and Tiao (1990) and Tsay (1988), two types of disturbances were introduced. They are the level shift and temporary change in level. Mathematically, a level shift (LS) can be described by

$$Y_t = Z_t + \frac{\omega_s}{(1-B)} I_t^{(d)}$$

where ω_s is the amount of shift in the level of Z_t. Writing

$$\frac{1}{(1-B)} = 1 + B + B^2 + \cdots$$

we see that for the above model

$$Y_t = \begin{cases} Z_t & \text{for } t < d \\ Z_t + \omega_s & \text{for } t \geq d. \end{cases}$$

Thus, the fixed constant ω_s is added to evey observation one or after d. Such a level shift is permanent.

In some cases, the effect of a level shift is only temporary. A mathematical model which is capable of describing such a shift is

$$Y_t = Z_t + \frac{\omega_c}{(1-\delta B)} I_t^{(d)}, \quad 0 < \delta < 1.$$

Since

$$\frac{1}{1-\delta B} = 1 + \delta B + \delta^2 B^2 + \delta^3 B^3 + \cdots$$

the magnitudes of level shift at times $d, d+1, d+2, \cdots$ are $\omega_c, \delta\omega_s, \delta^2\omega_s, \cdots$. Thus, the initial shift is ω_s and the subsequent shifts are discounted at the rate δ. With $0 < \delta < 1$, the shift decays exponentially to zero. We refer to such a temporary level shift as a transient change (TC) model. In practice, the value of δ is a prespecified constant. It may assume the value of 0.8 or 0.7.

Outlier Detection. In practice, outliers can occur at any time point is a series. Thus, to detect an outlier, we need to estimate the parameters $\omega_a, \omega_v, \omega_s, \omega_c$ and check the significance of these estimates.

For simplicity, we assume that the time series parameters are known. In practice, the parameters need to be estimated and we employ an iterative procedure to detect outliers. The four outlier models discussed above can be put in the general from

$$Y_t = Z_t + \omega_0 \frac{\omega(B)}{\delta(B)} I_t^{(d)}$$

where

$$\omega_0 = \begin{cases} \omega_a & \text{AO case} \\ \omega_v & \text{IO case} \\ \omega_s & \text{LS case} \\ \omega_c & \text{TC case} \end{cases} \qquad \frac{\omega(B)}{\delta(B)} = \begin{cases} 1 & \text{AO case} \\ \frac{\theta(B)}{\phi(B)} & \text{IO case} \\ \frac{1}{1-B} & \text{LS case} \\ \frac{1}{1-\delta B} & \text{TC case.} \end{cases}$$

Given $\theta(B)$ and $\phi(B)$, define

$$y_t = \frac{\phi(B)}{\theta(B)} Y_t, x_t = \frac{\phi(B)}{\theta(B)} \frac{\omega(B)}{\delta(B)} I_t^{(d)}.$$

Then, we have

$$y_t = \omega_0 x_t + a_t$$

which is precisely a simple linear regression equation. Therefore,

$$\hat{\omega}_0 = \frac{\sum_{t=1}^n y_t x_t}{\sum_{t=1}^n x_t^2} \quad \text{and} \quad \text{Var}(\hat{\omega}_0) = \frac{\sigma_a^2}{\sum_{t=1}^n x_t^2},$$

where n is the sample size. Using this simple technique, we obtain

- IO case: $\hat{\omega}_{v,d} = y_d$ and $\text{Var}(\hat{\omega}_{v,d}) = \sigma_a^2$.
- AO case: $\hat{\omega}_{a,d} = \rho_{a,d}^2 (y_d - \sum_{i=1}^{n-d} \pi_i y_{d+i})$ and $\text{Var}(\hat{\omega}_{a,d}) = \rho_{a,d}^2 \sigma_a^2$ where π's are the π-weights of Z_t and $\rho_{a,d}^2 = (1 + \pi_1^2 + \cdots + \pi_{n-d}^2)^{-1}$.
- LS case: $\hat{\omega}_{s,d} = \rho_{s,d}^2 (y_d - \sum_{i=1}^{n-d} \eta_i y_{d+i})$ and $\text{Var}(\hat{\omega}_{s,d}) = \rho_{s,d}^2 \sigma_a^2$ where η_i's are the coefficient of B^i in the polynomial $\eta(B) = \eta_0 - \eta_1 B - \eta_2 B^2 - \cdots = \frac{\pi(B)}{1-B}$ and $\rho_{s,d}^2 = (1 + \eta_1^2 + \cdots + \eta_{n-d}^2)^{-1}$.
- TC case: $\hat{\omega}_{c,d} = \rho_{c,d}^2 (y_d - \sum_{i=1}^{n-d} \beta_i y_{d+i})$ and $\text{Var}(\hat{\omega}_{c,d}) = \rho_{c,d}^2 \sigma_a^2$ where β_i is the coefficient of B^i in the polynomial $\beta(B) = \beta_0 - \beta_1 B - \cdots = \frac{\pi(B)}{1-\delta B}$ and $\rho_{c,d}^2 = (1 + \beta_1^2 + \cdots + \beta_{n-d}^2)^{-1}$.

Based on the above results, we may employ the test statistics

- Existence of an IO at d: $\lambda_{v,d} = \frac{\hat{\omega}_{v,d}}{\sigma_a}$

- Existence of an AO at d: $\lambda_{a,d} = \frac{\hat{\omega}_{a,d}}{\rho_{a,d}\sigma_a}$
- Existence of a LS at d: $\lambda_{s,d} = \frac{\hat{\omega}_{s,d}}{\rho_{s,d}\sigma_a}$
- Existence of an TC at d: $\lambda_{c,d} = \frac{\hat{\omega}_{c,d}}{\rho_{c,d}\sigma_a}$.

Under the null hypothesis of normality, no disturbance at d and knowing the time series parameters and d, all of the above four statistics are distributed as $N(0,1)$. In practice, the parameters can be replaced by the MLEs. However, since d is unknown, we need to apply the tests to all possible values of d. Consequently, in other words, we need to consider the maximum of test statistics over d. The resulting statistics are no longer normal. However, one can obtain certain percentiles via simulation or using distributions of certain extreme-value statistics. Experience suggests that using a critical value of 3.0 or 3.5 works reasonably well in practice.

We now consider an iterative procedure for time series analysis in the presence of outliers, level-shifts and temporary changes. See Tsay (1988, Journal of Forecasting) and Chang, et al. (1988, Technometrics). The procedure considered here is the very basic one. Some variants can be used to improve the efficacy.

1. Identify an ARMA model for Y_t, estimate the associated parameters. (Here we pretend that there are no outliers in Y_t.)

2. Based on the model of step 1, compute the four test statistics for each time point and identify

   $$\lambda_{v,max} = \max_d\{|\lambda_{v,d}|\}, \lambda_{a,max} = \max_d\{|\lambda_{a,d}|\}, \lambda_{s,max} = \max_d\{|\lambda_{s,d}|\}, \lambda_{c,max} = \max_d\{|\lambda_{c,d}|\},$$

   and denote the time points of these maximum by d_v, d_a, d_s, d_c, respectively.

3. Let $\lambda = \max\{\lambda_{v,max}, \lambda_{a,max}, \lambda_{s,max}, \lambda_{c,max}\}$ and compare λ with the pre-specified critical value C. If $\lambda < C$, there is no outlier and stop. If $\lambda \geq C$, continue to the next step.

4. Compute a modified series Y_t^* by removing the effect of the identified outlier and go to step 1 with Y_t replaced by Y_t^*.

Here we use "identification-detection-removing" cycle to remove the effect of outlier one-by-one.

Illustration: Air-passenger-miles data: Cryer's book.