Lecture 5: Aggregation and Regression Models with Time Series Errors
Bus 41910, Time Series Analysis, Mr. R. Tsay

1 Aggregation

Aggregation is an interesting subject in time series analysis. Two good references for this topic are the textbook by Wei (1990), which contains various results of aggregation, and the paper by Engel (1984, J. Time Series Analysis, pp. 159-171), which considers models of the sums and aggregations of linear ARMA models.
There are two types of aggregation: sum over series and sum over time (or temporal aggregation). Some relevant results are as follows:

a. Sum of two independent time series: Suppose that X_{1t} and X_{2t} are two independent ARMA series of orders (p_1, q_1) and (p_2, q_2), respectively. Let $Z_t = X_{1t} + X_{2t}$. Then, Z_t is an ARMA(p, q) process with $p \leq p_1 + p_2$ and $q \leq \max\{p_1 + q_2, p_2 + q_1\}$.
The reason that "$\leq$" is used is because of the possibility of common factors in the polynomials involved.
Proof: Write the model for X_{it} as

$$\phi_i(B)X_{it} = \theta_i(B)a_{it}.$$

Apply $\phi_1(B)\phi_2(B)$ to Z_t, we have

$$\phi_1(B)\phi_2(B)Z_t = \phi_1(B)\phi_2(B)(X_{1t} + X_{2t}) = \phi_2(B)\theta_1(B)a_{1t} + \phi_1(B)\theta_2(B)a_{2t}.$$

The result follows.
Notice that the "independent" assumption is important, otherwise, X_{1t} may be a lagged version of X_{2t}, e.g. $X_{1t} = X_{2,t-100}$, which can easily violate the stated result.

b. Temporal aggregation: For a time series $\{Y_t\}$, let $\{Z_\ell\}$ be the series consisting of sums of m non-overlapping points of Y_t. For example, for $m = 3$, we can aggregate monthly series to obtain a quarterly series. Mathematically, we can define Z_ℓ as

$$Z_\ell = \sum_{t=m(\ell-1)+1}^{m\ell} Y_t = (1 + B + \cdots + B^{m-1})Y_{m\ell}.$$

If Y_t is an ARMA(p, q) process, then Z_ℓ is also an ARMA process. To determine the model for Z_t, one needs to consider first the implication of the "backshift operator B" on Z_ℓ. By definition,

$$BZ_\ell = Z_{\ell-1} = \sum_{t=m(\ell-2)+1}^{m(\ell-1)} Y_t = (1 + B + \cdots + B^{m-1})Y_{m(\ell-1)} = B^m(1 + B + \cdots + B^{m-1})Y_{m\ell}.$$

Thus, the backshift operator "B" applying to Z_ℓ is amount to applying B^m to Y_t. In other words, we need to understand that "time scales"are different between the original series Y_t and the aggregated series Z_t. For instance, a backshift operator in a quarterly time series is corresponding to B^3 in a monthly series.

Example: The best way to derive a model for the aggregated series Z_t is to consider a simple example. Suppose that Y_t is an ARMA(1,1) process, say

$$(1-\phi B)Y_t = (1-\theta B)a_t,$$

and $m = 2$. Applying the polynomial operator $(1-\phi^2 B)$ to Z_ℓ, we have

$$\begin{aligned}
(1-\phi^2 B)Z_\ell &= Z_\ell - \phi^2 Z_{\ell-1} \\
&= (Y_{2\ell} + Y_{2\ell-1}) - \phi^2(Y_{2(\ell-1)} + Y_{2(\ell-1)-1}) \\
&= Y_{2\ell} - \phi^2 Y_{2\ell-2} + Y_{2\ell-1} - \phi^2 Y_{2\ell-3} \\
&= a_{2\ell} + (\phi-\theta)a_{2\ell-1} - \phi\theta a_{2\ell-2} + a_{2\ell-1} + (\phi-\theta)a_{2\ell-2} - \phi\theta a_{2\ell-3} \\
&\stackrel{def}{=} a_{2\ell} + \omega_1 a_{2\ell-1} + \omega_2 a_{2\ell-2} + \omega_3 a_{2\ell-3}.
\end{aligned}$$

In the above, we have used the result

$$\begin{aligned}
Y_t &= \phi Y_{t-1} + a_t - \theta a_{t-1} \\
&= \phi^2 Y_{t-2} + a_t + (\phi-\theta)a_{t-1} - \phi\theta a_{t-2}.
\end{aligned}$$

Since $2\ell-2$ and $2\ell-3$ correspond to $Z_{\ell-1}$, thus the model for Z_ℓ can be written as

$$(1-\phi^2 B)Z_\ell = (1-\Theta B)b_\ell,$$

which is again an ARMA(1,1) model in the time scale of ℓ.

Example: Consider now that $m = 2$ and Y_t is an AR(2) model, say

$$(1-.5B)(1+.5B)Y_t = (1-.25B^2)Y_t = a_t.$$

Here the coefficients are chosen purposely so that $(.5)^2 = (-.5)^2$. Applying $(1-.25B)$ to Z_ℓ, we have

$$\begin{aligned}
(1-.25B)Z_\ell &= Z_\ell - .25 Z_{\ell-1} \\
&= (Y_{2\ell} + Y_{2\ell-1}) - .25(Y_{2\ell-2} + Y_{2\ell-3}) \\
&= (Y_{2\ell} - .25Y_{2\ell-2}) + (Y_{2\ell-1} - .25Y_{2\ell-3}) \\
&= a_{2\ell} + a_{2\ell-1} \\
&\stackrel{def}{=} b_t,
\end{aligned}$$

which is an AR(1) model in the ℓ time scale. Thus, the fact that the coefficients satisfy $(.5)^2 = (-.5)^2$ reduces the AR order by 1. This phenomenon holds in general aggregation.

General result: For an ARMA(p, q) process Y_t, say $\phi(B)Y_t = \theta(B)a_t$, its m-point non-overlapping aggregate Z_t is an ARMA(p^*, q^*) where $p^* \leq p$ and $q^* \leq [p+1+\frac{q-p-1}{m}]$ with $[x]$ the integer part of x. The inequality "<" occurs when some zeros of $\phi(B)$ satisfy $\lambda_i^m = \lambda_j^m$ for $i \neq j$. If all the zeros of $\phi(B)$ are distinct and $\lambda_i^m \neq \lambda_j^m$ for $i \neq j$, then the equality holds. For further details, see Wei (1990, Chap. 16).

Systematic sampling: Suppose that a time series Y_t is an ARMA(p, q) process. However, one only observes the m-th subseries Z_t, which is defined as

$$Z_t = Y_{mt} \quad \text{with} \quad m \quad \text{a fixed positive integer.}$$

This situation occurs frequently in business and in process control. For instance, one may inspect every 5-th product of a production line. Here again, a backshift operator of Z_t is corresponding to B^m of the original series Y_t. In general, the model for Z_t is ARMA(p, r) where $r \leq [p + \frac{q-p}{m}]$.

2 Regression Models with Time Series Errors

Regression is one of the most commonly used statistical analyses. It requires that the error terms of the model are independent, homogeneous, and uncorrelated with regressors. For proper use of regression models, residual analysis is performed to check the validity of these assumptions. The well-known Durbin-Watson statistic is just one of many model checking statistics in regression analysis. Proper modifications must be exercised if any violation of the assumptions is found. Regression models with time series errors are models designed to deal with "independence" and "homogeneity" assumptions of the error term.
In general, a regression model with time-series errors can be written as

$$y_t = \boldsymbol{\beta}'\boldsymbol{x}_t + e_t, \tag{1}$$

where $\boldsymbol{x}_t$ is the vector of regressors, which may contain the constant 1 as its first element, and e_t follows a stationary time series model, e.g.

$$\phi(B)e_t = \theta(B)a_t.$$

For simplicity, one often assumes that e_t follows an AR(p) model.
If the serial dependence in e_t is overlooked, then the usual t-ratios of the least squares estimates $\widehat{\boldsymbol{\beta}}$ are not reliable, which may lead to erroneous inference. In this lecture, we shall discuss two approaches to analyze regression models with time-series errors.
If the main focus of the analysis is to make inference on $\boldsymbol{\beta}$, then the form of serial dependence in e_t is of secondary interest. In this case, we can use the so-called "HAC" estimators of the covariance matrix of $\widehat{\boldsymbol{\beta}}$ to adjust for the serial dependence in e_t. Here "HAC" stands for *heteroscedastic and autocorrelation consistent* estimator.

If $\{e_t\}$ of Eq. (1) are *iid*, then the covariance matrix of $\widehat{\boldsymbol{\beta}}$ is

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma_e^2 \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1},$$

where $\sigma_e^2 = \mathrm{Var}(e_t)$ and T is the sample size. If e_t has conditional heteroscedasticity, White (1980) proposes the modifed estimator

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1} \left[\frac{1}{T-p}\sum_{t=1}^{T} \hat{e}_t^2 \boldsymbol{x}_t \boldsymbol{x}_t'\right] \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1},$$

where $\hat{e}_t = y_t - \widehat{\boldsymbol{\beta}}' \boldsymbol{x}_t$ is the residual and p is the dimension of $\boldsymbol{x}_t$. This formula is understandable because, under conditional heteroscedasticity, $\hat{e}_t^2$ and $\boldsymbol{x}_t \boldsymbol{x}_t'$ are correlated.
If e_t is also serially correlated, Newey and West (1987) propose the estimator

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1} \widehat{\boldsymbol{C}}_* \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1},$$

where

$$T\widehat{\boldsymbol{C}}_* = \sum_{t=1}^{T} \hat{e}_t^2 \boldsymbol{x}_t \boldsymbol{x}_t' + \sum_{j=1}^{\ell} w_j \sum_{t=j+1}^{T} (\boldsymbol{x}_t \hat{e}_t \hat{e}_{t-j} \boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j} \hat{e}_{t-j} \hat{e}_t \boldsymbol{x}_t'),$$

where ℓ is a trunaction parameter and $w_j = 1 - \frac{j}{\ell+1}$. Newey and West suggest choosing ℓ to be the integer part of $e(T/100)^{2/9}$. This estimator essentially uses a nonparametric method to estimate the covariance matrix of $\sum_{t=1}^{T} \hat{e}_t \boldsymbol{x}_t$. In practice, the result might be sensitive to the choice of ℓ.

Another approach is to specify a time series model for e_t and perform a joint estimation of all parameters simultaneously. This is the approach we take in this class. In practice, we can apply model specification methods (to be discussed later) to the residual $\hat{e}_t$ of the ordinary least squares estimate of $\boldsymbol{\beta}$. With the specified model for e_t, we then perform maximum likelihood estimation of all parameters.
In SCA, the regression model with time-series errors is written as

$$y_t = \boldsymbol{\beta}' \boldsymbol{x}_t + \frac{\theta(B)}{\phi(B)} a_t,$$

where $\theta(B)$ and $\phi(B)$ are finite order polynomials as defined in an ARMA model. In R, the model is written as

$$\phi(B)(y_t - \boldsymbol{\beta}' \boldsymbol{x}_t) = \theta(B) a_t.$$

Of course, the two models are equivalent.

Remark: In fact, SCA also allows for some dynamic dependence of y_t on $\boldsymbol{x}_t$. For instance, if $\boldsymbol{x}_t$ is a scalar, say $\boldsymbol{x}_t = x_t$, then the model in SCA can assume the form

$$y_t = \frac{\omega(B)}{\delta(B)} x_t + \frac{\theta(B)}{\phi(B)} a_t,$$

where $\omega(B) = \omega_0 + \omega_1 B + \cdots + \omega_s B^s$ and $\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$ are polynomials of degree s and r, respectively. This type of models is referred to as a distributed-lag model or transfer function model.