Communications for Statistical Applications and Methods
2015, Vol. 22, No. 3, 209–222
DOI: http://dx.doi.org/10.5351/CSAM.2015.22.3.209
Print ISSN 2287-7843 / Online ISSN 2383-4757

Issues Related to the Use of Time Series in Model Building and Analysis: Review Article

William W. S. Wei[1,a]

[a]Department of Statistics, Temple University, USA

Abstract

Time series are used in many studies for model building and analysis. We must be very careful to understand the kind of time series data used in the analysis. In this review article, we will begin with some issues related to the use of aggregate and systematic sampling time series. Since several time series are often used in a study of the relationship of variables, we will also consider vector time series modeling and analysis. Although the basic procedures of model building between univariate time series and vector time series are the same, there are some important phenomena which are unique to vector time series. Therefore, we will also discuss some issues related to vector time models. Understanding these issues is important when we use time series data in modeling and analysis, regardless of whether it is a univariate or multivariate time series.

Keywords: temporal aggregation, systematic sampling, unit root test, causal relationship, vector time series, contemporal aggregation

1. Introduction

Let z_t be a time series process. For a stationary process, its mean, $E(z_t) = \mu$, and variance, $\gamma_z(0) = E(z_t - \mu)^2 = \sigma_z^2$, are constant. Also, in this case, its autocovariance function (ACF) between z_t and z_{t+k}, $\gamma_z(k) = E(z_t - \mu)(z_{t+k} - \mu) = E(\dot{z}_t\dot{z}_{t+k})$, and autocorrelation function, $\rho_z(k) = \gamma_z(k)/\gamma_z(0)$, are functions of only the time difference. The partial autocorrelation function (PACF) is defined as $\phi_{kk} = \mathrm{Corr}(z_t, z_{t+k}|z_{t+1}, \dots, z_{t+k-1})$. Some commonly used time series processes or models are:

(1) Autoregressive process of order p (AR(p) model)

$$\begin{aligned}\dot{z}_t - \phi_1\dot{z}_{t-1} - \cdots - \phi_p\dot{z}_{t-p} &= a_t,\\ \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)\dot{z}_t &= a_t,\\ \phi_p(B)\dot{z}_t &= a_t.\end{aligned} \tag{1.1}$$

(2) Moving average process of order q (MA(q) model)

$$\begin{aligned}\dot{z}_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-p}\\ &= \left(1 - \theta_1 B - \theta_1 B^2 - \cdots - \theta_q B^q\right) a_t\\ &= \theta_q(B) a_t.\end{aligned} \tag{1.2}$$

[1] Department of Statistics, Temple University, 1810 North 13th Street, 330 Speakman Hall (006-12), Philadelphia, PA 19122, USA. E-mail: wwei@temple.edu

Published 31 May 2015 / journal homepage: http://csam.or.kr
© 2015 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

(3) Autoregressive moving average process (ARMA(p, q) model)

$$\phi_p(B)\dot{z}_t = \theta_q(B)a_t. \tag{1.3}$$

(4) Autoregressive integrated moving average process (ARIMA(p, d, q) model)

$$\phi_p(B)(1-B)^d z_t = \theta_0 + \theta_q(B)a_t. \tag{1.4}$$

The model is stationary if the roots of its associated AR polynomial are all outside the unit circle, and the important characteristics of stationary models can be summarized in the following table:

Model	ACF	PACF
AR(p)	Decreases exponentially	Cuts off at lag p
MA(q)	Cuts off at lag q	Decreases exponentially
ARMA(p, q)	Decreases exponentially	Decreases exponentially

2. Temporal Aggregation Effect on Model Form

Time series are used in many studies either for model building or inference. We must be careful when choosing what kind of time series data is used in the analysis. Since many time series variables like rainfall, industrial production, and sales exist only in some aggregated forms, we will begin with the issue related to the temporal aggregation effect on the model form. Given a time series z_t, let $Z_T = (1 + B + \cdots + B^{m-1})z_{mT}$. For example, with $m = 3$, $Z_1 = (1 + B + B^2)z_3 = z_1 + z_2 + z_3$, $Z_2 = (1 + B + B^2)z_6 = z_4 + z_5 + z_6$, etc. We will call z_t as non-aggregate series and Z_T as aggregate series. For $m = 3$ if z_t is a monthly series, then Z_T will be a quarterly series. To make inference, should we use non-aggregate series, z_t or aggregate series, Z_T? Do they make any difference? Are the time series models for z_t and for Z_T the same?

The first published papers on aggregation effects on ARIMA models were by Tiao (1972) and Amemiya and Wu (1972), and they led to many other studies on the topic including my Ph.D. dissertation and life time research in the area.

To answer the above questions, we need to study the relationship of autocovariances between the non-aggregate and aggregate series z_t and Z_T, or more generally between $w_t = (1-B)^d z_t$ and $U_T = (1-\mathcal{B})^d Z_T$. Define the m-period overlapping sum,

$$\zeta_t = \sum_{j=0}^{m-1} z_{t-j} = \left(1 + B + \cdots + B^{m-1}\right) z_t,$$

and note that $Z_T = \zeta_{mT}$, $(1-\mathcal{B})Z_T = Z_T - Z_{T-1} = \zeta_{mT} - \zeta_{m(T-1)} = (1-B^m)\zeta_{mT}$,

$$\begin{aligned} U_T &= (1-\mathcal{B})^d Z_T = (1-B^m)^d \zeta_{mT} \\ &= \left[\left(1 + B + \cdots + B^{m-1}\right)(1-B)\right]^d \left(1 + B + \cdots + B^{m-1}\right) z_{mT} \\ &= \left(1 + B + \cdots + B^{m-1}\right)^{d+1} (1-B)^d z_{mT} \\ &= \left(1 + B + \cdots + B^{m-1}\right)^{d+1} w_{mT}. \end{aligned}$$

Hence, we have

$$\gamma_U(k) = \left(1 + B + \cdots + B^{m-1}\right)^{2(d+1)} \gamma_w(mk + (d+1)(m-1)). \tag{2.1}$$

Now let us consider an MA(2) model for z_t, $z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$. If $m = 3$, what is the model for Z_T? For this MA(2) model, we have

$$\begin{aligned}
\gamma_z(0) &= \left(1 + \theta_1^2 + \theta_2^2\right)\sigma_a^2, \\
\gamma_z(1) &= (-\theta_1 + \theta_1\theta_2)\,\sigma_a^2, \\
\gamma_z(2) &= -\theta_2\sigma_a^2,
\end{aligned}$$

and

$$\gamma_z(j) = 0, \quad |j| > 2.$$

Note that for $d = 0$ and $m = 3$, from (2.1), we have

$$\gamma_Z = \left(1 + B + B^2\right)^2 \gamma_z(3k+2) = \left(1 + 2B + 3B^2 + 2B^3 + B^4\right)\gamma_z(3k+2),$$

where

$$\begin{aligned}
\begin{bmatrix} \gamma_Z(0) \\ \gamma_Z(1) \\ \gamma_Z(2) \\ \gamma_Z(3) \end{bmatrix}
&= \begin{bmatrix}
1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1
\end{bmatrix}
\begin{bmatrix} \gamma_z(-2) \\ \gamma_z(-1) \\ \gamma_z(0) \\ \gamma_z(1) \\ \vdots \\ \gamma_z(11) \end{bmatrix} \\
&= \begin{bmatrix}
3 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1
\end{bmatrix}
\begin{bmatrix} \gamma_z(0) \\ \gamma_z(1) \\ \gamma_z(2) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 3\gamma_z(0) + 4\gamma_z(1) + 2\gamma_z(2) \\ \gamma_z(1) + 2\gamma_z(2) \\ 0 \\ 0 \end{bmatrix}.
\end{aligned}$$

Hence, Z_T is a MA(1) process, $Z_T = (1 - \Theta\mathcal{B})A_T$ where

$$\begin{aligned}
\gamma_Z(0) &= \left(1 + \Theta^2\right)\sigma_A^2 = 3\left[1 + \theta_1^2 + \theta_2^2\right]\sigma_a^2 + 4[-\theta_1 + \theta_1\theta_2]\sigma_a^2 + 2[-\theta_2]\sigma_a^2, \\
\gamma_Z(1) &= -\Theta\sigma_A^2 = [-\theta_1 + \theta_1\theta_2]\sigma_a^2 + 2\,[-\theta_2]\,\sigma_a^2, \\
\frac{1 + \Theta^2}{-\Theta} &= \frac{3\left[1 + \theta_1^2 + \theta_2^2\right] + 4[-\theta_1 + \theta_1\theta_2] + 2[-\theta_2]}{[-\theta_1 + \theta_1\theta_2] + 2\,[-\theta_2]},
\end{aligned}$$

and

$$\sigma_A^2 = \frac{[-\theta_1 + \theta_1\theta_2]\sigma_a^2 + 2[-\theta_2]\sigma_a^2}{-\Theta}.$$

More generally, we have the following results from Stram and Wei (1986):

(1) Temporal aggregation of the AR(p) process

Suppose that the non-aggregate series z_t follows a stationary AR(p) process,

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right) z_t = a_t.$$

Let $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)z_t$ and δ_i^{-1} for $i = 1, \ldots, p^*$ be the distinct roots of $\phi_p(B)$, each with multiplicity s_i such that $\sum_{i=1}^{p^*} s_i = p$. For any given value of m, let b equal the number of distinct values δ_i^m for $i = 1, \ldots, p^*$. Furthermore, partition the numbers s_i for $i = 1, \ldots, p^*$ into b distinct sets A_i such that s_k and $s_j \in A_i$ if and only if $\delta_k^m = \delta_j^m$. Then the mth order aggregate series, Z_T follows an ARMA(M, N_1) model,

$$\left(1 - \alpha_1 \mathcal{B} - \cdots - \alpha_M \mathcal{B}^M\right) Z_T = \left(1 - \beta_1 \mathcal{B} - \cdots - \beta_{N_1} \mathcal{B}^{N_1}\right) E_T,$$

where $M = \sum_{i=1}^{b} \max A_i$, $\max A_i$ = the largest element in A_i, $N_1 = [p + 1 - (p + 1)/m] - (p - M) = [M + 1 - (p + 1)/m]$, the E_T are white noise with mean 0 and variance σ_E^2, and α_i, β_j, and σ_E^2 are functions of ϕ_k's and σ_a^2.

(2) If $z_t \sim$ ARMA(p, q) model, then $Z_T \sim$ ARMA(M, N_2), where $N_2 = [p+1+(q-p-1)/m]-(p-M)$.

(3) If $z_t \sim$ ARIMA(p, d, q) model, then $Z_T \sim$ ARIMA(M, d, N_3), where $N_3 = [p + d + 1 + (q - p - d - 1)/m] - (p - M)$.

The limiting behavior of aggregates was studied by Tiao (1972) and he showed that given $z_t \sim$ ARIMA(p, d, q) model, the limiting model for the aggregates, Z_T exists, and as $m \to \infty$, $Z_T \to$ IMA(d, d).

When a variable is a stock variable and we observe only every mth value of the variable, i.e., given $z_1, z_2, z_3, \ldots$ but we observe only $Z_T = z_{mT}$. For example, for $m = 3$, $Z_1 = z_3, Z_2 = z_6, \ldots$. We have the following interesting result from Wei (1981).

Given $z_t \sim$ IMA(d, q): $(1 - B)^d z_t = (1 - \theta_1 B - \theta_1 B^2 - \cdots - \theta_q B^q)a_t$. Let $Z_T = z_{mT}$. Then, as $m \to \infty$, $Z_T \to$ IMA($d, d - 1$). Theoretically, every ARIMA(p, d, q) process can be approximated by an IMA(d, q) process. Thus, if z_t is an ARIMA($p, 1, q$), then as $m \to \infty$, Z_T approaches IMA(1, 0), which could very likely explain why most daily stock prices follow a random walk model. In building an underlying time series model, we need to be aware of the effect of the use of aggregate series. In addition to the above cited references, we refer readers to some other useful references including Brewer (1973), Wei (1982), Ansley and Kohn (1983), Weiss (1984), Wei and Stram (1988), Marcellino (1999), Shellman (2004), and Sbrana and Silvestrini (2013).

3. Aggregation Effect on Testing for a Unit Root

Given the AR(1) model

$$z_t = \phi z_{t-1} + a_t, \tag{3.1}$$

where a_t is $N(0, \sigma_a^2)$ white noise process and $t = 1, 2, \ldots, n$. To test a unit root, H_0: $\phi = 1$ vs H_1: $\phi < 1$, since

$$\hat{\phi} = \frac{\sum_{t=2}^{n} z_{t-1} z_t}{\sum_{t=2}^{n} z_{t-1}^2} = 1 + \frac{\sum_{t=2}^{n} z_{t-1} a_t}{\sum_{t=2}^{n} z_{t-1}^2}.$$

Dickey and Fuller (1979) suggested using the following test statistic and showed that under H_0: $\phi = 1$,

$$n(\hat{\phi} - 1) = \frac{n^{-1} \sum_{t=2}^{n} z_{t-1} a_t}{n^{-2} \sum_{t=2}^{n} z_{t-1}^2} \xrightarrow{D} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\}}{\int_0^1 [W(x)]^2 dx}, \tag{3.2}$$

where $W(t)$ is a Wiener process (also known as Brownian motion process). We reject H_0 if the value of the test statistic is too small (negative).

In practice, aggregate data, $Z_T = (1 + B + \cdots + B^{m-1}) z_{mT}$ are often used. It has been shown by Teles *et al.* (2008) that under H_0: $\phi = 1$, $(1 - B) z_t = a_t$, the corresponding model for the aggregate series is

$$Z_T = Z_{T-1} + E_T - \Theta E_{T-1}, \tag{3.3}$$

where the E_{T-1} are independent and identically distributed variables with zero mean and variance σ_E^2 and the parameters Θ and σ_E^2 are determined as follows:

(1) If $m = 1$, then $\Theta = 0$; $\sigma_E^2 = \sigma_a^2$;

(2) If $m \geq 2$, then

$$\Theta = -\frac{2m^2 + 1}{m^2 - 1} + \left\{\left(\frac{2m^2 + 1}{m^2 - 1}\right)^2 - 1\right\}^{\frac{1}{2}},$$

$$\sigma_E^2 = \sigma_a^2 \frac{m\left(2m^2 + 1\right)}{3\left(1 + \Theta^2\right)}.$$

As a result, the corresponding test statistic becomes

$$\begin{aligned} N\left(\hat{\phi} - 1\right) &= \frac{N^{-1} \sum_{T=2}^{N} Z_{T-1} E_T}{N^{-2} \sum_{T=2}^{N} Z_{T-1}^2} - \Theta \frac{N^{-1} \sum_{T=2}^{N} Z_{T-1} E_{T-1}}{N^{-2} \sum_{T=2}^{N} Z_{T-1}^2} \\ &\xrightarrow{D} \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\}}{\int_0^1 [W(x)]^2 dx} + \frac{\frac{m^2-1}{6m^2}}{\int_0^1 [W(x)]^2 dx}. \end{aligned} \tag{3.4}$$

Comparing with (3.2), we see that the limiting distribution of the test statistic for the aggregate time series depends on the order of aggregation m. Since $m \geq 2$, the distribution of the test statistic is shifted to the right, and the shift increases with the order of aggregation. Aggregation leads to empirical significance levels lower than the nominal level and significantly reduces the power of the test.

Example 1. In this example, we simulated a time series of 240 observations from the model $z_t = 0.95 z_{t-1} + a_t$ where the a_t are i.i.d. $N(0, 1)$. This series was then aggregated with $m = 3$.

(1) Test a unit root based on non-aggregate series, z_t (240 observations). Based on the sample autocorrelation function and partial autocorrelation function, we have an AR(1) model. The least squares estimation leads to the following result

$$\hat{z}_t = \underset{(.0182)}{0.9603} z_{t-1}.$$

To test the hypothesis of a unit root, the value of the test statistic is

$$n\left(\hat{\phi}-1\right) = 240(0.9603-1) = -9.53.$$

At $\alpha = 5\%$, the critical point from Dickey and Fuller (1979) is between -8.0 and -7.9 for $n = 240$. Thus, the hypothesis of a unit root is rejected, and we conclude that the underlying model is stationary. This is consistent with the underlying simulated model.

(2) Test a unit root with aggregate series, Z_T, with $m = 3$ (80 observations). The sample autocorrelation function and partial autocorrelation function suggest an AR(1) model (an ARMA(1,1) model was also considered but its MA parameter was not significant). The least squares estimation leads to the following result

$$\hat{Z}_t = \underset{(.0448)}{0.9205} z_{t-1}.$$

To test the hypothesis of a unit root, the value of the test statistic is

$$n\left(\hat{\phi}-1\right) = 80(0.9205-1) = -6.36.$$

Again, at $\alpha = 5\%$, the critical point from Dickey and Fuller (1979) is between -7.9 and -7.7 for $n = 80$. Thus, the hypothesis of a unit root is not rejected, and we conclude that the underlying model is nonstationary. This leads to a wrong conclusion. However, if we use the adjusted critical value given in Teles *et al.* (2008) based on the adjusted test statistic of (3.4), which is between -5.45 and -5.40 for $n = 80$, we will reject the null hypothesis of a unit root and leads to a consistent conclusion.

When aggregate time series are used in modeling and testing, we need to make sure to use a proper adjusted table for the test of its significance.

4. Aggregation Effect on a Dynamic Relation

Time series are often used in regression analysis, which is possible the most commonly used statistical method. So we will also consider the consequence of the use of aggregate series in a regression model. Let us consider the simple regression model,

$$y_t = \alpha x_{t-1} + e_{t-1}, \tag{4.1}$$

which is a one-sided causal relationship. If x_{t-1} is also stochastic, for example, if it follows a MA(1) process, we can also write the joint system as

$$\begin{aligned} x_t &= (1-\theta B)a_t, \\ y_t &= \alpha(1-\theta B)a_{t-1} + e_r, \end{aligned} \tag{4.2}$$

or

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} (1-\theta B) & 0 \\ \alpha(1-\theta B)B & 1 \end{bmatrix} \begin{bmatrix} a_t \\ e_t \end{bmatrix}, \tag{4.3}$$

where a_t and e_t are independent $N(0,\sigma_a^2)$ and $N(0,\sigma_e^2)$, respectively. Let $Y_T = (\sum_{j=0}^{m-1} B^j)y_{mT}$, $X_T = (\sum_{j=0}^{m-1} B^j)x_{mT}$, $W_T = (\sum_{j=0}^{m-1} B^j)x_{mT-1}$, and $E_T = (\sum_{j=0}^{m-1} B^j)e_{mT}$. Equation (4.1) implies that

$$Y_T = \alpha W_T + E_T. \tag{4.4}$$

For $m = 3$, $W_1 = x_0 + x_1 + x_2$, $W_2 = x_3 + x_4 + x_5$, *etc.*, which are not available, and the available data are $X_1 = x_1 + x_2 + x_3$, $X_2 = x_4 + x_5 + x_6$, *etc.* A natural way to estimate W_T is to consider its projection on X_T. Specifically, we let $\mathbf{Z} = [W_T, X_T]'$ and compute its covariance matrix generating function

$$G_Z(\mathbf{B}) = \sum_{k=-\infty}^{\infty} \Gamma_k \mathbf{B}^k = \Gamma_0 + \sum_{k=1}^{\infty} \Gamma_k \left(\mathbf{B}^k + \mathbf{F}^k\right) = \begin{bmatrix} G_{11}(\mathbf{B}) & G_{12}(\mathbf{B}) \\ G_{21}(\mathbf{B}) & G_{22}(\mathbf{B}) \end{bmatrix}, \tag{4.5}$$

with

$$\begin{aligned}
\Gamma_0 &= \mathrm{E}\begin{bmatrix} W_T \\ X_T \end{bmatrix}\begin{bmatrix} W_T & X_T \end{bmatrix} = \mathrm{E}\begin{bmatrix} W_T W_T & W_T X_T \\ X_T W_T & X_T X_T \end{bmatrix} \\
&= \sigma_a^2 \begin{bmatrix} m(1-\theta)^2 + 2\theta & (m-1)(1-\theta)^2 \\ (m-1)(1-\theta)^2 & m(1-\theta)^2 + 2\theta \end{bmatrix}, \\
\Gamma_1 &= \mathrm{E}\begin{bmatrix} W_{T-1} \\ X_{T-1} \end{bmatrix}\begin{bmatrix} W_T & X_T \end{bmatrix} = \mathrm{E}\begin{bmatrix} W_{T-1} W_T & W_{T-1} X_T \\ X_{T-1} W_T & X_{T-1} X_T \end{bmatrix} \\
&= \sigma_a^2 \begin{bmatrix} -\theta & 0 \\ (1-\theta)^2 & -\theta \end{bmatrix}, \\
\Gamma_k &= \mathrm{E}\begin{bmatrix} W_{T-k} \\ X_{T-k} \end{bmatrix}\begin{bmatrix} W_T & X_T \end{bmatrix} = \mathrm{E}\begin{bmatrix} W_{T-k} W_T & W_{T-k} X_T \\ X_{T-k} W_T & X_{T-k} X_T \end{bmatrix} \\
&= 0, \quad k \geq 2,
\end{aligned}$$

and hence

$$\begin{aligned}
G_{11}(\mathbf{B}) &= \sigma_a^2 \left\{\left[m(1-\theta)^2 + 2\theta\right] - \theta(\mathbf{B} + \mathbf{F})\right\} = G_{22}(\mathbf{B}), \\
G_{21}(\mathbf{B}) &= \sigma_a^2 \left\{(m-1)(1-\theta)^2 + (1-\theta)^2 \mathbf{B}\right\},
\end{aligned}$$

and

$$G_{12}(\mathbf{B}) = G_{21}(\mathbf{F}).$$

It follows that

$$\begin{aligned}
\hat{W}_T &= [G_{22}(\mathbf{B})]^{-1} G_{21}(\mathbf{B}) X_T \\
&= \frac{(1-\theta)^2[(m-1) + \mathbf{B}]}{[m(1-\theta)^2 + 2\theta] - \theta(\mathbf{B} + \mathbf{F})} X_T.
\end{aligned} \tag{4.6}$$

It is interesting to note that Equation (4.6) can be rewritten as

$$\begin{aligned}
\hat{W}_T &= \frac{(1-\theta)^2[(m-1) + \mathbf{B}]}{[m(1-\theta)^2 + 2\theta] - \theta(\mathbf{B} + \mathbf{F})} X_T \\
&= \frac{\left[\frac{m-1}{m} + \frac{1}{m}\mathbf{B}\right] X_T}{\left[1 + \frac{\theta}{m(1-\theta)^2}(1 - \mathbf{B})(1 - \mathbf{F})\right]},
\end{aligned} \tag{4.7}$$

which implies that the estimate of $\hat{W}_T$ is the weighted average of X_T and X_{T-1} with weights $(m-1)/m$ and $1/m$, respectively. This is clearly reasonable. The aggregate model then becomes

$$\begin{aligned} Y_T &= \alpha \hat{W}_T + U_T \\ &= \frac{\alpha(1-\theta)^2[(m-1)+\mathbf{B}]}{[m(1-\theta)^2+2\theta]-\theta(\mathbf{B}+\mathbf{F})} X_T + U_T, \end{aligned} \tag{4.8}$$

where $U_T = \alpha(W_T - \hat{W}_T) + E_T = \alpha V_T + E_T$, $G_U(\mathbf{B}) = \alpha^2 G_V(\mathbf{B}) + m\sigma_e^2$, $G_V(\mathbf{B}) = G_{11}(\mathbf{B}) - G_{12}(\mathbf{B})[G_{22}(\mathbf{B})]^{-1}G_{21}(\mathbf{B}) = \sigma_a^2\{[m(1-\theta)^2] + 2\theta - \theta(\mathbf{B}+\mathbf{F}) - [(1-\theta)^4[(m-1)+\mathbf{B}][(m-1)+\mathbf{F}]]/[m(1-\theta)^2 + 2\theta - \theta(\mathbf{B}+\mathbf{F})]\}$, and $\mathbf{F} = \mathbf{B}^{-1}$. Thus, temporal aggregation turns a one-sided causal relationship into a two-sided feedback system. It is important to note that after proposing an underlying model for a study, one should use the same time unit in the hypothesis and data collection for modeling and testing. An improper use of time unit could lead to a very misleading conclusion. For a more detailed description, we refer readers to Tiao and Wei (1976). Other useful references include Wei (1978), and Lütkepohl (1987).

5. Issues Related to Vector Time Series Modeling

5.1. Representation of vector time series models

In studying the relationship of variables, other than the regression model, we often consider vector time series models. Although the basic procedures of model building between univariate time series and vector time series are the same, there are some important phenomena which are unique to vector time series models. We now discuss some special issues of vector time models.

First, let us review some results from univariate time series models. It is well known that we can always write a stationary process as a MA representation

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \psi_j a_{j-1},$$

or

$$\dot{Z}_t = \sum_{j=0}^{\infty} \psi_j a_{j-1} = \sum_{j=0}^{\infty} \psi_j B^j a_t = \left(\sum_{j=0}^{\infty} \psi_j B^j\right) a_t = \psi(B) a_t$$

such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Similarly, we can write an invertible process as an AR representation

$$\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + \cdots + a_t,$$

or

$$\begin{aligned} \dot{Z}_t - \pi_1 \dot{Z}_{t-1} - \pi_2 \dot{Z}_{t-2} + \cdots &= a_t, \\ \pi(B)\dot{Z}_t &= a_t, \end{aligned}$$

such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$. From these two representations, we have the well-known dual relationship between AR(p) and MA(q) models in the univariate time series processes. That is, a finite order AR process corresponds to an infinite order MA process, and a finite order MA process corresponds to an infinite order AR process. For example, an AR(1) model, $(1-\phi B)\dot{Z}_t = a_t$, corresponds to an infinite order MA process, $\dot{Z}_t = (1-\phi B)^{-1}a_t = (1+\phi B+\phi^2 B^2+\cdots)a_t$, and a MA(1) model, $\dot{Z}_t = (1-\theta B)a_t$, corresponds to an infinite order AR process, $(1+\theta B+\theta^2 B^2+\cdots)\dot{Z}_t = a_t$.

Let $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \ldots, Z_{m,t}]'$ be the m-dimensional vector time series. Some commonly used vector time series models are VAR(p), VMA(q), and VARMA(p, q) processes. Again, we can write these vector processes in a moving average representation

$$\dot{\mathbf{Z}}_t = \mathbf{a}_t + \boldsymbol{\psi}_1\mathbf{a}_{t-1} + \boldsymbol{\psi}_2\mathbf{a}_{t-2} + \cdots, \tag{5.1}$$

where $\dot{\mathbf{Z}} = \mathbf{Z} - \boldsymbol{\mu}$ and $\mathbf{a}_t$ is a m-dimensional vector white noise process $N(\mathbf{0}, \boldsymbol{\Sigma})$. We can also write it in an autoregressive representation

$$\dot{\mathbf{Z}}_t - \boldsymbol{\Pi}_1\dot{\mathbf{Z}}_{t-1} - \boldsymbol{\Pi}_2\dot{\mathbf{Z}}_{t-2} + \cdots = \mathbf{a}_t. \tag{5.2}$$

It follows that we can express a VAR(1) process,

$$\boldsymbol{\Phi}(B)\mathbf{Z}_t = (\mathbf{I} - \boldsymbol{\Phi}B)\mathbf{Z}_t = \mathbf{a}_t, \tag{5.3}$$

in the following MA representation

$$\mathbf{Z}_t = [\boldsymbol{\Phi}(B)]^{-1}\mathbf{a}_t = (\mathbf{I} - \boldsymbol{\Phi}B)^{-1}\mathbf{a}_t = \sum_{j=0}^{\infty}\boldsymbol{\Phi}^j\mathbf{a}_{t-j}, \tag{5.4}$$

where $\boldsymbol{\Phi}_0 = \mathbf{I}$. A natural question to ask: is the VMA representation always an infinite order? People often think that the univariate model is a special case of the vector model with dimension equal 1, and so the answer to the question is obviously yes. However, let us consider the following 3-dimensional vector VAR(1) model

$$\begin{cases} Z_{1,t} = .8Z_{2,t} + a_{1,t}, \\ Z_{2,t} = .5Z_{3,t} + a_{2,t}, \\ Z_{3,t} = a_{3,t} \end{cases} \tag{5.5a}$$

or equivalently

$$(\mathbf{I} - \boldsymbol{\Phi}B)\mathbf{Z}_t = \mathbf{a}_t, \tag{5.5b}$$

where $\boldsymbol{\Phi} = \begin{bmatrix} 0 & .8 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & 0 \end{bmatrix}$. Since $\boldsymbol{\Phi}^2 \neq \mathbf{0}$ and $\boldsymbol{\Phi}^j = \mathbf{0}$ for $j > 2$, Equation (5.4) actually represents a VMA(2) model. In fact,

$$[\boldsymbol{\Phi}(B)]^{-1} = \frac{1}{|\boldsymbol{\Phi}(B)|}\mathrm{adj}[\boldsymbol{\Phi}(B)]. \tag{5.6}$$

Thus, the inverse of a non-degenerate VAR(1) matrix polynomial (i.e., $\boldsymbol{\Phi}(B) \neq \mathbf{I}$) will be of a finite order if the determinant $|\boldsymbol{\Phi}(B)|$ is independent of B. For more detailed discussion, we refer readers to Tiao and Tsay (1989), and Shen and Wei (1995).

5.2. Representation of multiplicative seasonal vector autoregressive moving average models

Given a univariate seasonal ARMA model

$$\Phi_P(B^s)\phi_p(B)x_t = \theta_q(B)\Theta_Q(B^s)a_t, \tag{5.7}$$

where

$$\begin{aligned}\Phi_P(B^s) &= 1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps},\\ \phi_p(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p,\\ \theta_q(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q,\\ \Theta_Q(B^s) &= 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Qs},\end{aligned}$$

and a_t is a Gaussian white noise process with mean 0 and a constant variance σ_a^2. The model is often denoted as ARMA$(p, q)\times(P, Q)_s$. To facilitate our discussion, we will use the order of the polynomials appearing in the equation and denote it as ARMA$(P)_s(p)(q)(Q)_s$. When $\mathbf{x}_t = [x_{1,t}, \ldots, x_{k,t}]'$ is a k-dimensional vector, the natural extension is the following multiplicative vector autoregressive moving average VARMA$(P)_s(p)(q)(Q)_s$ model,

$$\boldsymbol{\Phi}_P(B^s)\boldsymbol{\phi}_p(B)\mathbf{x}_t = \boldsymbol{\theta}_q(B)\boldsymbol{\Theta}_Q(B^s)\mathbf{a}_t, \tag{5.8}$$

where

$$\begin{aligned}\boldsymbol{\Phi}_P(B^s) &= \mathbf{I} - \boldsymbol{\Phi}_1 B^s - \cdots - \boldsymbol{\Phi}_P B^{Ps},\\ \boldsymbol{\phi}_p(B) &= \mathbf{I} - \boldsymbol{\phi}_1 B - \cdots - \boldsymbol{\phi}_p B^p,\\ \boldsymbol{\theta}_q(B) &= \mathbf{I} - \boldsymbol{\theta}_1 B - \cdots - \boldsymbol{\theta}_q B^q,\\ \boldsymbol{\Theta}_Q(B^s) &= \mathbf{I} - \boldsymbol{\Theta}_1 B^s - \cdots - \boldsymbol{\Theta}_Q B^{Qs},\end{aligned}$$

are matrix polynomials. The matrix $\mathbf{I}$ is the k-dimensional identity matrix, the $\boldsymbol{\Phi}$s, $\boldsymbol{\phi}$s, $\boldsymbol{\theta}$s, and $\boldsymbol{\Theta}$s are $k \times k$ parameter matrices, and $\mathbf{a}_t$ is a vector Gaussian white noise process with mean vector $\mathbf{0}$ and $\mathrm{E}(\mathbf{a}_t\mathbf{a}_t') = \boldsymbol{\Omega}$.

Note that as expected, the vector model reduces to the univariate model when $k = 1$. Moreover, in such a case, the ARMA$(P)_s(p)(q)(Q)_s$ model can also be written as the following ARMA$(p)(P)_s(Q)_s(q)$ model

$$\phi_p(B)\Phi_P(B^s)x_t = \Theta_Q(B^s)\theta_q(B)a_t. \tag{5.9}$$

As a result, a multiplicative seasonal ARMA model is also traditionally written in the form of ARMA $(P)_s(p)(q)(Q)_s$. This traditional representation has been adopted by many researchers for both univariate and vector time series.

When $k > 1$ and $\mathbf{x}_t$ is a vector process, can we really extend the above operation and write the VARMA$(P)_s(p)(q)(Q)_s$ model, $\boldsymbol{\Phi}_P(B^s)\boldsymbol{\phi}_p(B)\mathbf{x}_t = \boldsymbol{\theta}_q(B)\boldsymbol{\Theta}_Q(B^s)\mathbf{a}_t$, as the following VARMA$(p)(P)_s(Q)_s$ (q) model?

$$\boldsymbol{\phi}_p(B)\boldsymbol{\Phi}_P(B^s)\mathbf{x}_t = \boldsymbol{\Theta}_Q(B^s)\boldsymbol{\theta}_q(B)\mathbf{a}_t. \tag{5.10}$$

Let us consider two simple seemingly equivalent bivariate VAR$(1)_4(1)$ and VAR$(1)(1)_4$ representations with the following parameters

$$\mathbf{\Phi} = \begin{bmatrix} .5 & .5 \\ -.25 & .25 \end{bmatrix},$$
$$\boldsymbol{\phi} = \begin{bmatrix} .9 & .5 \\ .7 & -.5 \end{bmatrix},$$

and the associated noise $\mathbf{a}_t$ is a vector Gaussian white noise process with mean zero and covariance matrix $\mathbf{\Omega}$. For the VAR$(1)_4(1)$ representation,

$$\left(\mathbf{I} - \mathbf{\Phi} B^4\right)(\mathbf{I} - \boldsymbol{\phi} B)\mathbf{x}_t = \mathbf{a}_t,$$
$$\mathbf{x}_{n+1} = \boldsymbol{\phi}\mathbf{x}_n + \mathbf{\Phi}\mathbf{x}_{n-3} - \mathbf{\Phi}\boldsymbol{\phi}\mathbf{x}_{n-4} + \mathbf{a}_{n+1},$$
$$\begin{bmatrix} \hat{x}_{1,n}(1) \\ \hat{x}_{2,n}(1) \end{bmatrix} = \begin{bmatrix} .9 & .5 \\ .7 & -.5 \end{bmatrix}\begin{bmatrix} x_{1,n} \\ x_{2,n} \end{bmatrix} + \begin{bmatrix} .5 & .5 \\ -.25 & .25 \end{bmatrix}\begin{bmatrix} x_{1,(n-3)} \\ x_{2,(n-3)} \end{bmatrix} - \begin{bmatrix} .5 & .5 \\ -.25 & .25 \end{bmatrix}\begin{bmatrix} .9 & .5 \\ .7 & -.5 \end{bmatrix}\begin{bmatrix} x_{1,(n-4)} \\ x_{2,(n-4)} \end{bmatrix}$$
$$= \begin{bmatrix} .9 & .5 \\ .7 & -.5 \end{bmatrix}\begin{bmatrix} x_{1,n} \\ x_{2,n} \end{bmatrix} + \begin{bmatrix} .5 & .5 \\ -.25 & .25 \end{bmatrix}\begin{bmatrix} x_{1,(n-3)} \\ x_{2,(n-3)} \end{bmatrix} - \begin{bmatrix} .8 & 0 \\ -.05 & -.25 \end{bmatrix}\begin{bmatrix} x_{1,(n-4)} \\ x_{2,(n-4)} \end{bmatrix}.$$

For the VAR$(1)(1)_4$ representation,

$$(\mathbf{I} - \boldsymbol{\phi} B)(\mathbf{I} - \mathbf{\Phi} B^4)\mathbf{x}_t = \mathbf{a}_t,$$
$$\mathbf{x}_{n+1} = \boldsymbol{\phi}\mathbf{x}_n + \mathbf{\Phi}\mathbf{x}_{n-3} - \boldsymbol{\phi}\mathbf{\Phi}\mathbf{x}_{n-4} + \mathbf{a}_{n+1},$$
$$\begin{bmatrix} \hat{x}_{1,n}(1) \\ \hat{x}_{2,n}(1) \end{bmatrix} = \begin{bmatrix} .9 & .5 \\ .7 & -.5 \end{bmatrix}\begin{bmatrix} x_{1,n} \\ x_{2,n} \end{bmatrix} + \begin{bmatrix} .5 & .5 \\ -.25 & .25 \end{bmatrix}\begin{bmatrix} x_{1,(n-3)} \\ x_{2,(n-3)} \end{bmatrix} - \begin{bmatrix} .9 & .5 \\ .7 & -.5 \end{bmatrix}\begin{bmatrix} .5 & .5 \\ -.25 & .25 \end{bmatrix}\begin{bmatrix} x_{1,(n-4)} \\ x_{2,(n-4)} \end{bmatrix}$$
$$= \begin{bmatrix} .9 & .5 \\ .7 & -.5 \end{bmatrix}\begin{bmatrix} x_{1,n} \\ x_{2,n} \end{bmatrix} + \begin{bmatrix} .5 & .5 \\ -.25 & .25 \end{bmatrix}\begin{bmatrix} x_{1,(n-3)} \\ x_{2,(n-3)} \end{bmatrix} - \begin{bmatrix} .325 & .575 \\ .475 & .225 \end{bmatrix}\begin{bmatrix} x_{1,(n-4)} \\ x_{2,(n-4)} \end{bmatrix}.$$

The different implications between the two seemingly equivalent representations are clear and cannot be ignored especially when a policy related decision is to be made. Please see Yozgatligil and Wei (2009) for details.

6. Contemporal Aggregation

In addition to temporal aggregation discussed in earlier sections, there is another commonly used aggregation. For example, the total money supply is the aggregate of demand deposits and currency in circulation. The total housing start is the aggregate of housing starts in the north east, north central, south, and west regions, which again are the subaggregates of housing starts in different states. The total sales of a company is the aggregate of the sales achieved by all of its branches throughout the country or countries.

Let $z_{1,t}, z_{2,t}, \ldots, z_{m,t}$ be the m component time series and $Y_t = \sum_{i=1}^{m} z_{i,t}$ be the corresponding series of aggregates. Suppose we are interested in forecasting the future aggregate Y_{t+l} for some l, based on the knowledge of the available time series up to the time t. Clearly, such forecasts can be obtained through the following three methods.

Method 1: based on a model using the aggregate series Y_t and its l-step ahead forecast, $\hat{Y}_t(l)$.

Method 2: based on individual component models using the non-aggregate series $z_{i,t}$ and the sum of the forecasts from all component models, i.e., $\hat{\hat{Y}}_t(l) = \sum_{i=1}^{m} \hat{z}_{i,t}(l)$.

Method 3: based on a joint multiple time series model and the forecast from the joint multiple model, $\hat{\hat{\hat{Y}}}_t(l)$.

Question: what are the relative efficiencies among the three methods in terms of the minimum mean square error forecast? The answers are given below.

(1) $\mathrm{E}[Y_{t+l} - \hat{\hat{\hat{Y}}}_t(l)]^2 \leq \mathrm{E}[Y_{t+l} - \hat{Y}_t(l)]^2$; $\mathrm{E}[Y_{t+l} - \hat{\hat{\hat{Y}}}_t(l)]^2 \leq \mathrm{E}[Y_{t+l} - \hat{\hat{Y}}_t(l)]^2$; and equality holds when $z_{1,t}, z_{2,t}, \ldots, z_{m,t}$ are orthogonal to each other.

(2) The comparison between methods 1 and 2 depends on the model structure; there is no definite winner between methods 1 and 2.

The answer (2) above could be surprising to some people because aggregation normally will cause information loss. For the proof, we refer readers to Wei and Abraham (1981).

Next, let us consider the m-dimensional models related to both time and space where we write it as

$$\mathbf{Z}_t = \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{m,t} \end{bmatrix} = \begin{bmatrix} \text{Time series for space 1} \\ \text{Time series for space 2} \\ \vdots \\ \text{Time series for space m} \end{bmatrix}. \tag{6.1}$$

The commonly used model to describe (6.1) is the following space-time autoregressive moving average (STARMA(p, q)) model,

$$\mathbf{Z}_t = \sum_{k=1}^{p} \sum_{l=0}^{r_k} \phi_{k,l} \mathbf{W}^{(l)} \mathbf{Z}_{t-k} + \mathbf{a}_t - \sum_{k=1}^{q} \sum_{l=0}^{\tau_k} \theta_{k,l} \mathbf{W}^{(l)} \mathbf{a}_{t-k}, \tag{6.2}$$

where $\mathbf{W}^{(l)} = \left\lfloor w_{i,j}^{(l)} \right\rfloor$ are $m \times m$ the spatial weight matrices, $\sum_{j=1}^{m} w_{i,j}^{(l)} = 1$,

$$w_{i,j}^{(l)} = \begin{cases} (0, 1], & \text{if location } j \text{ is the } l^{th} \text{ order neighbor of } i, \\ 0, & \text{otherwise,} \end{cases}$$

and $\phi_{k,l}$ and $\theta_{k,l}$ are autoregressive and moving average parameters at time lag k and space lag l, respectively. p is the autoregressive order, q is the moving average order, r_k is the spatial order for the kth autoregressive term, and τ_k is the spatial order for the kth moving average term. The STARMA(p, q) model becomes a space-time autoregressive (STAR(p)) model when $q = 0$. It becomes a space-time moving average (STMA(q)) model when $p = 0$.

For these STARMA(p, q) models, it will be interesting to study the effect of temporal aggregation, the effect of contemporal aggregation, and more generally, the combining effects of both temporal and contemporal aggregation. Because of the time limitation of this presentation, we refer readers to Arbia *et al.* (2010), Giacomini and Granger (2004), and Hendry and Hubrich (2011) among others on some of these issues.

References

Amemiya, T. and Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model, *Journal of the American Statistical Association*, **67**, 628–632.

Ansley, C. F. and Kohn, R. (1983). Exact likelihood of vector autoregressive-moving average process with missing or aggregated data, *Biometrika*, **70**, 275–278.

Arbia, G., Bee, M. and Espa, G. (2010). Aggregation of regional economic time series with different spatial correlation structures, *Geographical Analysis*, **43**, 78–103.

Brewer, K. W. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models, *Journal of Econometrics*, **1**, 133–154.

Dickey, D. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.

Giacomini, R. and Granger, C. W. J. (2004). Aggregation of space-time processes, *Journal of Econometrics*, **118**, 7–26.

Hendry, D. F. and Hubrich, K. (2011). Comment: The disaggregate forecasts or combing disaggregate information to forecast an aggregate, *Journal of Business & Economic Statistics*, **29**, 216–227.

Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*. Springer-Verlag, Berlin.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis, *Journal of Business and Economic Statistics*, **17**, 129–136.

Sbrana, G. and Silvestrini, A. (2013). Aggregation of exponential smoothing process with an application to portfolio risk evaluation, *Journal of Banking & Finance*, **37**, 1437–1450.

Shellman, S. M. (2004). Time series intervals and statistical inference: The effects of temporal aggregation on event data analysis, *Political Analysis*, **12**, 97–104.

Shen, S. Y. and Wei, W. W. S. (1995). A note on the representation of a vector ARMA model, *Journal of Applied Statistical Science*, **2**, 311–318.

Stram, D. and Wei, W. W. S. (1986). Temporal aggregation in the ARIMA process, *Journal of Time Series Analysis*, **7**, 279–292.

Teles, P., Wei, W. W. S. and Hodgess, E. M. (2008). Testing a unit root based on aggregate time series, *Communications in Statistics - Theory and Methods*, **37**, 565–590.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series, *Biometrika*, **59**, 525–531.

Tiao, G. C. and Box, G. E. P. (1981). Modelling multiple time series with applications, *Journal of the American Statistical Association*, **76**, 802–816.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series, *Journal of the Royal Statistical Society Series B*, **51**, 157–213.

Tiao, G. C. and Wei, W. W. S. (1976). Effect of temporal aggregation on the dynamic relationship of two time series variables, *Biometrika*, **63**, 513–523.

Wei, W. W. S. (1978). The effect of temporal aggregation on parameter estimation in distributed lag model, *Journal of Econometrics*, **8**, 237–246.

Wei, W. W. S. (1981). Effect of systematic sampling on ARIMA models, *Communications in Statistics - Theory and Methods*, **10**, 2389–2398.

Wei, W. W. S and Abraham, B. (1981). Forecasting contemporal time series aggregates, *Communications in Statistics - Theory and Methods*, **10**, 1335–1344.

Wei, W. W. S. (1982). Comment: The effect of systematic sampling and temporal aggregation on causality – A cautionary note, *Journal of the American Statistical Association*, **77**, 316–319.

Wei, W. W. S. and Stram, D. (1988). An eigenvalue approach to the limiting behavior of time series aggregates, *Annals of the Institute of Statistical Mathematics*, **40**, 101–110.

Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models, *Journal of Econometrics*, **26**, 271–281.

Yozgatligil, C. and Wei, W. W. S. (2009). Representation of multiplicative seasonal vector autoregressive moving average models, *The American Statistician*, **63**, 328–334.

Received December 14, 2014; Revised January 28, 2015; Accepted January 28, 2015