THE UNIVERSITY OF CHICAGO
Graduate School of Business
Business 41914, Spring Quarter 2007, Mr. Ruey S. Tsay

Lecture 4: Vector ARMA Models (continued.)

0.1 Asymptotic Distribution for Conditional MLE of VARMA Models

For a well-defined stationary and invertible VARMA(p,q) model, we assume that the innovations $\{\boldsymbol{a}_t\}$ satisfies (a) $E(\boldsymbol{a}_t|F_{t-1}) = \boldsymbol{0}$, (b) $E(\boldsymbol{a}_t\boldsymbol{a}_t'|F_{t-1}) = \boldsymbol{\Sigma} > 0$, and (c) $\boldsymbol{a}_t$ has finite fourth moments, where F_{t-1} denotes the σ-field generated by $\{\boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-2}, \cdots\}$. It has been proven (Dunsmuir and Hannan, 1976, and Hannan and Deistler, 1988, and Reinsel, 1993, p. 117) that the MLE of the model parameters are stongly consistent and asymptotically normally distributed. Here MLE denotes estimates obtained by maximizing the multivariate normal density function.
Next, we discuss two more topics of VARMA models. They are the *impulse response functions* and *forecast error variance decomposition.*

0.2 Impulse Response Functions

For a stationary VARMA(p,q) model, $\boldsymbol{\phi}(B)(\boldsymbol{z}_t - \boldsymbol{\mu}) = \boldsymbol{\theta}(B)\boldsymbol{a}_t$, the MA model representation is

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{a}_{t-i} \tag{1}$$

where $\mathrm{Var}(\boldsymbol{a}_t) = \boldsymbol{\Sigma}$ and $\boldsymbol{\psi}_0 = \boldsymbol{I}$. The matrices $\boldsymbol{\psi}_i$ can be obtained by equating the coefficient matrices of B^i in the equation

$$\sum_{i=0}^{\infty} \boldsymbol{\psi}_i B^i = [\boldsymbol{\phi}(B)]^{-1}\boldsymbol{\theta}(B).$$

Since $\boldsymbol{\Sigma}$ is typically not a diagonal matrix, the components a_{it}s are often correlated, making the interpretation of elements $\psi_{\ell,ij}$ of $\boldsymbol{\psi}_\ell$ complicated. To overcome the difficulty, one can make use of the Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}$. That is,

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{L}',$$

where $\boldsymbol{G}$ is a diagonal matrix and $\boldsymbol{L}$ is a lower triangular matrix with unit diagonal elements. Let $\boldsymbol{b}_t = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. Then, $\mathrm{Var}(\boldsymbol{b}_t) = \boldsymbol{G}$ so that elements of $\boldsymbol{b}_t$ are uncorrelated. The MA representation in Eq. (1) can be rewritten as

$$\begin{aligned}
\boldsymbol{z}_t &= \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_t + \sum_{i=1}^{\infty} \boldsymbol{\psi}_j \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_{t-i} \\
&= \boldsymbol{\mu} + \boldsymbol{\Psi}_0\boldsymbol{b}_t + \sum_{i=1}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{b}_{t-i}, \qquad (2)
\end{aligned}$$

where $\boldsymbol{\Psi}_0 = \boldsymbol{L}$ and $\boldsymbol{\Psi}_i = \boldsymbol{\psi}_i \boldsymbol{L}$ for $i > 0$. Note that $\boldsymbol{\Psi}_0$ is a lower triangular matrix. Thus, by construction, z_{jt} can depend contemporaneously on z_{it} with $i = 1, \cdots, (j-1)$, where $j = 2, \cdots, k$. The matrices $\boldsymbol{\Psi}_i$ thus depend on the ordering of elements in $\boldsymbol{z}_t$. Different orderings give rise to different matrices $\boldsymbol{\Psi}_i$, and there are $k!$ possible orderings of $\boldsymbol{z}_t$.
Since elements of $\boldsymbol{b}_t$ in Eq. (2) are uncorrelated, we have

$$\frac{\partial z_{i,t+\ell}}{\partial b_{jt}} = \frac{\partial z_{i,t}}{\partial b_{j,t-\ell}} = \Psi_{\ell,ij}, \quad i, j = 1, \cdots k,$$

where $\Psi_{\ell,ij}$ is the (i,j)th element of $\boldsymbol{\Psi}_\ell$. Consequently, $\boldsymbol{\Psi}_i$ are the impulse response functions of $\boldsymbol{z}_t$ with respect to $\boldsymbol{b}_t$. As metnioned earlier, these impulse response functions depend on the ordering of elements of $\boldsymbol{z}_t$.

0.3 Forecast Error Variance Decomposition (FEVD)

For the MA representation in Eq. (2), the ℓ-step ahead forecast error at the forecast origin n is

$$\boldsymbol{e}_n(\ell) = \sum_{m=0}^{\ell-1} \boldsymbol{\Psi}_m \boldsymbol{b}_{n+\ell-m}.$$

For the ith element of $\boldsymbol{z}_t$, the forecast error is

$$\begin{aligned} e_{i,n}(\ell) &= \sum_{m=0}^{\ell-1} \sum_{j=1}^{k} \Psi_{m,ij} b_{j,t+\ell-m} \\ &= \sum_{m=0}^{\ell-1} \Psi_{m,i1} b_{1,t+\ell-m} + \cdots + \sum_{m=0}^{\ell-1} \Psi_{m,ik} b_{k,t+\ell-m}. \end{aligned}$$

Since b_{it} $(i = 1, \cdots, k)$ are uncorrelated and have no serial correlations, we have

$$\mathrm{Var}[e_{i,n}(\ell)] = \sigma_{b1}^2 \left(\sum_{m=0}^{\ell-1} \Psi_{m,i1}^2 \right) + \cdots + \sigma_{bk}^2 \left(\sum_{m=0}^{\ell-1} \Psi_{m,ik}^2 \right),$$

where $\sigma_{bj}^2 = \mathrm{Var}(b_{jt})$. Consequently, the portion of the variance of $e_{i,n}(\ell)$ due to the shock $\{b_{jt}\}$ is

$$\mathrm{FEVD}_{i,j}(\ell) = \frac{\sigma_{bj}^2 \left(\sum_{m=0}^{\ell-1} \Psi_{m,ij}^2 \right)}{\sigma_{b1}^2 \left(\sum_{m=0}^{\ell-1} \Psi_{m,i1}^2 \right) + \cdots + \sigma_{bk}^2 \left(\sum_{m=0}^{\ell-1} \Psi_{m,ik}^2 \right)}, \quad i, j = 1, \cdots, k.$$

Again, the FEVD depends on the ordering of the elements of $\boldsymbol{z}_t$.

Remark: Impulse respose weights and FEVD are available in S-Plus. See the commands VAR and impRes in Finmetrics of S-Plus. In SCA, one can obtain the $\boldsymbol{\psi}_i$ of the matrix polynomial $[\boldsymbol{\phi}(B)]^{-1}\boldsymbol{\theta}(B) = \boldsymbol{\psi}(B)$ from the command nopsiweight in MFORE. The Cholesky decomposition can be obtained in SCA by the command chol, e.g., $\boldsymbol{b} = \mathrm{chol}(\boldsymbol{a})$, where $\boldsymbol{b}$ is an upper triangular matrix such that $\boldsymbol{a} = \boldsymbol{b}'\boldsymbol{b}$.

Chapter 3

Unit-Root Nonstationary VARMA Models

We begin with a simple case.

3.1 Multivariate exponential smoothing

Consider first the invertible vector IMA(1,1) model

$$(1 - B)\boldsymbol{z}_t = (\boldsymbol{I} - \boldsymbol{\theta} B)\boldsymbol{a}_t, \tag{3.1}$$

where, for simplicity, $\boldsymbol{z}_0 = \boldsymbol{0}$. Rewrite the model as

$$\boldsymbol{z}_t = \boldsymbol{a}_t - \boldsymbol{\theta}\boldsymbol{a}_{t-1} + \boldsymbol{z}_{t-1}.$$

Then, by repeatedly subsititions, we obtain

$$\boldsymbol{z}_t = \boldsymbol{a}_t + (\boldsymbol{I} - \boldsymbol{\theta})[\boldsymbol{a}_{t-1} + \boldsymbol{a}_{t-2} + \cdots],$$

where it is understood that $\boldsymbol{a}_t = \boldsymbol{0}$ for $t \leq 0$. Thus, the model has strong memory because the cofficient matrix $\boldsymbol{\psi}_i = \boldsymbol{I} - \boldsymbol{\theta}$ does not converge to zero as i increases. Next, from the model, we have

$$(\boldsymbol{I} + \boldsymbol{\theta} B + \boldsymbol{\theta}^2 B^2 + \cdots)(\boldsymbol{I} - \boldsymbol{I} B)\boldsymbol{z}_t = \boldsymbol{a}_t.$$

That is,

$$\boldsymbol{z}_t = (\boldsymbol{I} - \boldsymbol{\theta})\boldsymbol{z}_{t-1} + (\boldsymbol{I} - \boldsymbol{\theta})\boldsymbol{\theta}\boldsymbol{z}_{t-2} + (\boldsymbol{I} - \boldsymbol{\theta})\boldsymbol{\theta}^2\boldsymbol{z}_{t-3} + \cdots + \boldsymbol{a}_t. \tag{3.2}$$

From Eq. (3.2), it is easy to see that the 1-step ahead forecast of $\boldsymbol{z}_t$ at the forecast origin $t-1$ is

$$\begin{aligned}
\hat{\boldsymbol{z}}_{t-1}(1) &= (\boldsymbol{I} - \boldsymbol{\theta})\boldsymbol{z}_{t-1} + (\boldsymbol{I} - \boldsymbol{\theta})\boldsymbol{\theta}\boldsymbol{z}_{t-2} + (\boldsymbol{I} - \boldsymbol{\theta})\boldsymbol{\theta}^2\boldsymbol{z}_{t-3} + \cdots \\
&= (\boldsymbol{I} - \boldsymbol{\theta})[\boldsymbol{z}_{t-1} + \boldsymbol{\theta}\boldsymbol{z}_{t-2} + \boldsymbol{\theta}^2\boldsymbol{z}_{t-3} + \cdots].
\end{aligned}$$

This is a ganeralization of the univariate exponential smoothing model except that the weights decay in matrix form, not in terms of elements of $\boldsymbol{\theta}$.

Finally, the marginal model of z_{it} is a uivariate IMA(1,1) model. Specifically,

$$(1-B)z_{it} = a_{it} - \sum_{j=1}^{k} \theta_{1,ij} a_{j,t-1} \equiv b_{it} - \Theta b_{i,t-1},$$

where Θ and $\mathrm{Var}(b_{it})$ can be obtained from $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$.

3.2 Unit Roots

The multivariate IMA(1,1) model in Eq. (3.1) is unit-root nonstationary. It is a special unti-root process because every element z_{it} has a unit root. Indeed, all zeros of the determinant $|\boldsymbol{I} - \boldsymbol{I}B|$ are *on* the unit circle. In general, a VARMA(p, q) model is said to be unit-root nonstationary if $|\boldsymbol{\psi}(x)| \neq 0$ for $|x| < 1$, but $|\boldsymbol{\psi}(1)| = 0$.
Unit roots and co-integration have generated mush research interest in the 1980s and 1990s. Some important references are given below.
For limiting properties of unit-root processes:

1. Dickey and Fuller (1979, JASA, Vol. 74, p. 427-431). Distribution of the estimates for autoregressive time series.

2. P.C.B. Phillips (1987, Econometrica, Vol. 55, p.277-301.) Ties series regression with a unit root.

   This paper is informative and easy to understand, but it focuses on a single unit root.

3. Chan and Wei (1988, Annals of Statistics, Vol. 16, p. 367-401.) Limiting distribution of least squares estimates of unstable autoregressive processes.

   This paper gives results for multiple unit roots and other roots on the unit circle. The innovations are not serially correlated, however.

For co-integration,

1. Box and Tiao (1977, Biometrika, Vol. 64, p. 355-365.) A canonical analysis of multiple time series.

2. Engle and Granger (1987, Vol. 55, p. 251-276.) Cointegration and error correction: representation, estimation and testing.

For co-integration tests,

1. Johansen, S. (1988, J. Economics Dynamics and Control, Vol. 12, p.231-254.) Statistical analysis of cointegration vectors.

2. Ahn and Reinsel (1990, JASA, Vol. 85, p. 813-823.) Estimation for partially nonstationary multivariate autoregressive models,

and others.

3.2.1 Review of Univariate Results

We begin with a brief review of estimation and testing of unit roots in a univariate ARIMA model.
Basic results
Assume that $\{a_t\}$ is a martigale difference sequence such that (a) $E(a_t|F_{t-1}) = 0$, (b) $E(a_t^2|F_{t-1}) = \sigma_a^2 < \infty$, and $E(|a_t|^{2+\delta}|F_{t-1}) < \infty$ for some $\delta > 0$, where $F_{t-1} = \sigma$-field generated by $\{a_{t-1}, a_{t-2}, \cdots\}$. [See Chan and Wei (1988). Phillips (1987) allows for a_t to have some weak serial correlations (strong mixing condition). Suppose that the sample is $\{a_1, a_2, \cdots, a_T\}$. Let $S_t = \sum_{i=1}^{t} a_i$ be the partial sum of $\{a_t\}$. For $u \in [0, 1]$, define the function

$$X_T(u) = \frac{1}{\sqrt{T}\sigma_a} S_{[Tu]},$$

where $[Tu]$ denotes the integer part of Tu. Then, the functional central limit theorem states that

$$X_T(u) \longrightarrow_d W(u), \quad \text{as} \quad T \to \infty,$$

where $W(u)$ is a standatd Brownian motion on [0,1] and $\to_d$ denotes weak convergence, i.e. convergence in distribution. Furthermore, if h is a continuous functional on D, which is defined as the space of all real-valued functions on [0,1] that are right continuous at each point in [0,1] and have finite left limits, then $h(X_T) \longrightarrow_d h(W)$ as $T \to \infty$. This is the continuous mapping theorem.

AR(1) case:
Consider the simple AR(1) model $z_t = \phi z_{t-1} + a_t$, $t = 1, \cdots, T$ and $z_0 = 0$. Suppose that the null hypothesis is that z_t is a random walk, i.e. $H_o : \phi = 1$ vs $H_a : \phi < 1$. The least squares estimator of ϕ is

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

Under the null hypothesis, $\phi = 1$ so that $z_t = \sum_{i=0}^{t-1} a_{t-i} + z_0$. It can be shown that

$$T(\hat{\phi} - 1) = \frac{T^{-1}\sum_{t=1}^{T} z_{t-1} a_t}{T^{-2}\sum_{t=1}^{T} z_{t-1}^2} = O_p(1),$$

where $O_p(1)$ denotes bounded in probability as $T \to \infty$. In addition, the numerator and the denominator of $T(\hat{\phi}-1)$ possess non-degenerate and non-normal limiting distributions. Dickey and Fuller (1979) show that

$$T(\hat{\phi} - 1) \longrightarrow_d \frac{1}{2}(\Lambda^2 - 1)/\Gamma,$$

where $(\Gamma, \Lambda) = (\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2, \sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i)$, the Z_i are *iid* N(0,1) random variables and $\gamma_i = 2(-1)^{i+1}/[(2i-1)\pi]$. Chan and Wei (1988) give an equivalent representation for the distribution as

$$T(\hat{\phi} - 1) \longrightarrow_d \frac{\int_0^1 W(u)dW(u)}{\int_0^1 W^2(u)du} = \frac{(W^2(1) - 1)/2}{\int_0^1 W^2(u)du},$$

where $W(u)$ is a standard Brownian motion process on [0,1]. Furthermore, the limiting disitribution of the t-ratio statistic

$$\hat{t} = \frac{\hat{\phi} - 1}{s_a(\sum_{t=1}^{T} z_{t-1}^2)^{-1/2}}, \tag{3.3}$$

where $s_a^2 = (\sum_{t=1}^T z_t^2 - \hat{\phi}\sum_{t=1}^T z_{t-1}z_t)/(T-2)$ is the residual mean square, is given by

$$\hat{t} \longrightarrow_d \frac{\int_0^1 W(u)dW(u)}{[\int_0^1 W^2(u)du]^{1/2}}.$$

Fuller (1976, book) gives percentiles of the prior two statistics.

Remark. For the AR(1) model considered with $\phi = 1$, Theorem 3.1 of Phillips (1987) shows that

1. $T^{-2}\sum_{t=1}^T z_{t-1}^2 \to_d \sigma_a^2 \int_0^1 W^2(u)du$,
2. $T^{-1}\sum_{t=1}^T z_{t-1}(z_t - z_{t-1}) \to_d \frac{\sigma_a^2}{2}[W^2(1) - 1]$,
3. $T(\hat{\phi} - 1) \to_d \frac{(1/2)[W^2(1)-1]}{\int_0^1 W^2(u)du}$,
4. $\hat{\phi} \to_p 1$,
5. $\hat{t} \to_d \frac{(1/2)[W^2(1)-1]}{[\int_0^1 W^2(u)du]^{1/2}}$,

where $W(u)$ is the standard Brownian motion process on C, which is the set of all real-valued continuous functions on [0,1]. Note that under the assumption for $\{a_t\}$ stated above, $\sigma^2 = \sigma_u^2$ in Phillips' notation and it is σ_a^2 in our notation.

AR(p) case

For higher-order AR(p) process, $\phi(B)z_t = a_t$, we focus on the case that $\phi(B) = \phi^*(B)(1-B)$ where $p^*(B)$ is a stationary AR polynomial. Let $\phi^*(B) = 1 - \sum_{i=1}^{p-1}\phi_i^* B^i$. the model becomes $\phi(B)z_t = \phi^*(B)(1-B)z_t = (1-B)z_t - \sum_{i=1}^{p-1}\phi_i^*(1-B)z_{t-i} = a_t$. Testing for a unit root in $\phi(B)$ is equivalent to testing $\rho = 1$ in the model

$$z_t = \rho z_{t-1} + \sum_{j=1}^{p-1}\phi_j^*(z_{t-j} - z_{t-j-1}) + a_t.$$

Or equivalently, the same as testing for $\rho - 1 = 0$ in the model

$$\Delta z_t = (\rho - 1)z_{t-1} + \sum_{j=1}^{p-1}\phi_j^*\Delta z_{t-j} + a_t,$$

where $\Delta z_t = z_t - z_{t-1}$. The above model is the univariate version of error-correction form. It is easy to verify that (a) $\rho - 1 = -\phi(1) = \sum_{i=1}^p \phi_i - 1$ and $\phi_j^* = -\sum_{i=j+1}^p \phi_i$. In practice, the least squares estimates of the model

$$\Delta z_t = \beta z_{t-1} + \sum_{j=1}^{p-1}\phi_j^*\Delta z_{t-j} + a_t, \tag{3.4}$$

where $\beta = \rho - 1$, is used. It can be shown that the t-ratio of $\hat{\beta}$ (against 0) has the same limiting distribution as $\hat{t}$ in Eq. (3.3). See, for instance, Fuller (1976, Theorem 8.5.1). This is the well-known augmented Dickey-Fuller unit-root test. Furthermore, the limiting distribution of the LS

estimates $\hat{\phi}_i^*$ in Eq. (3.4) is the same as that of fitting an AR$(p-1)$ model to Δz_t. In other words, limiting properties of the estimates for the stationary part remain unchanged when we treate the unit-root as known a priori.
Remark: In our review, we assume there is no constant term in the model. One can include a constant term in the model and obtain the associated limiting distribution. The limiting distribution is different, but the idea remains the same.

3.2.2 Demonstration

Consider the series of U.S. quarterly unemployment rate and real GDP from 1948 to 2004. The data were from the Federal Reserve Bank at St Louis. The results of unit-root test are given below (using R).

```
> setwd("C:/teaching/mts")
> library(fSeries)
> da=read.table("q-gdpun.txt")
> dim(da)
[1] 228   5
> da[1,]
    V1 V2 V3     V4     V5
1 1948  1  1 7.3878 3.7333
> gdp=da[,4]
> plot(gdp,type='l')
> help(UnitrootTests)

> x=diff(gdp)
> m1=ar(x,method='mle')    <== Find the AR order for the differenced series.
> m1

Call:
ar(x = x, method = "mle")

Coefficients:
      1        2        3        4
 0.2998   0.1355  -0.0798  -0.1084

Order selected 4  sigma^2 estimated as  8.597e-05

> adfTest(gdp,lag=4)

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 4
```

```
  STATISTIC:
    Dickey-Fuller: 6.0361
  P VALUE:
    0.99

Warning message:
p-value greater than printed p-value in: adfTest(gdp, lag = 4)

> unemp=da[,5]
> plot(unemp,type='l')

> m3=ar(unemp,method='mle')
> m3
ar(x = unemp, method = "mle")

Coefficients:
      1        2        3        4        5        6        7        8
 1.6896  -0.7782  -0.0185  -0.0948   0.2125   0.0630  -0.0761  -0.3264
      9       10
 0.4826  -0.1941

Order selected 10  sigma^2 estimated as  0.0769

> m4=adfTest(unemp,lag=9)
> m4

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 9
  STATISTIC:
    Dickey-Fuller: -0.502
  P VALUE:
    0.4559

> adfTest(unemp,lag=9,type=c("c"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 9
```

```
  STATISTIC:
    Dickey-Fuller: -2.6667
  P VALUE:
    0.08465

> adfTest(unemp,lag=9,type=c("ct"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 9
  STATISTIC:
    Dickey-Fuller: -2.785
  P VALUE:
    0.2462
```

3.3 Co-integration

In the literature, a time series z_t is said to be integrated of order 1, i.e. $I(1)$ process, if $(1-B)z_t$ is stationary and invertible. Similarly, z_t is an $I(d)$ process if $(1-B)^d z_t$ is stationary and invertible, where $d > 0$. The order d is referred to as the order of integration or the multiplicity of a unit root. A stationary and invertible time series is said to be $I(0)$ process. .

Consider the multivariate process $\boldsymbol{z}_t$. If z_{it} are $I(1)$ processes, but a non-trivial linear combination $\boldsymbol{\beta}'\boldsymbol{z}_t$ is $I(0)$, then $\boldsymbol{z}_t$ is said to be co-integrated of order 1. Basically, if z_{it} are $I(d)$ nonstationary and $\boldsymbol{\beta}'\boldsymbol{z}_t$ is $I(h)$ with $h < d$, then $\boldsymbol{z}_t$ is co-integrated. In real applications, the case of $d = 1$ and $h = 0$ is of major interest. Thus, co-integration often means that a linear combination of individually unit-root nonstationary time series becomes a stationary and invertible series. The linear combination $\boldsymbol{\beta}$ is called a co-integrating vector.

Suppose that $\boldsymbol{z}_t$ is unit-root nonstationary such that the marginal models for z_{it} have a unit root. If $\boldsymbol{\beta}$ is a $k \times m$ matrix of full rank m, where $m \leq k$, such that $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{z}_t$ is $I(0)$, then $\boldsymbol{z}_t$ is a co-integrated system with m co-integrating vectors, which are the columns of $\boldsymbol{\beta}$. This means that there are $k-m$ unit roots in $\boldsymbol{z}_t$. For the given full-rank $k \times m$ matrix $\boldsymbol{\beta}$ with $m < k$, let $\boldsymbol{\beta}_\perp$ be a $k \times (k-m)$ full-rank matrix such that $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$. Then, $\boldsymbol{n}_t = \boldsymbol{\beta}'_\perp \boldsymbol{z}_t$ is a unit-root nonstationary. The components n_{it} $(i = 1, \cdots, (k-m))$ are referred to as the *common trends* of $\boldsymbol{z}_t$. We shall discuss methods for finding co-integrating vectors and common trends later.

Co-integration implies a long-term stable relationship between variables in forecasting. Since $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{z}_t$ is stationary, it is mean-reverting so that the ℓ-step ahead forecast of $\boldsymbol{w}_{T+\ell}$ at the forecast origin T satisfies

$$\hat{\boldsymbol{w}}_T(\ell) \to_p E(\boldsymbol{w}_t) \equiv \boldsymbol{\mu}_w, \quad \ell \to \infty.$$

This implies that $\boldsymbol{\beta}'\hat{\boldsymbol{z}}_T(\ell) \to \boldsymbol{\mu}_w$ as ℓ increases. Thus, point forecasts of $\boldsymbol{z}_t$ satisfy a long-term stable constraint.

3.3.1 An example of co-integration

To understand co-integration, we consider a simple example. Suppose that the bivariate process $\boldsymbol{z}_t$ follows the model

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} - \begin{bmatrix} 0.5 & -1.0 \\ -0.25 & 0.5 \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix},$$

where the covariance matrix $\boldsymbol{\Sigma}$ of the shock $\boldsymbol{a}_t$ is positive definite. For simplicity, assume that $\boldsymbol{\Sigma} = \boldsymbol{I}$. The prior VARMA(1,1) model, from Tsay (2002, chap. 8), is not a weakly stationary because the two eigenvalues of the AR coefficient matrix are 0 and 1. Rewrite the model as

$$\begin{bmatrix} 1-0.5B & B \\ 0.25B & 1-0.5B \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.2B & 0.4B \\ 0.1B & 1-0.2B \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}. \tag{3.5}$$

Premultiplying the above equation by

$$\begin{bmatrix} 1-0.5B & -B \\ -0.25B & 1-0.5B \end{bmatrix},$$

we obtain

$$\begin{bmatrix} 1-B & 0 \\ 0 & 1-B \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.7B & -0.6B \\ -0.15B & 1-0.7B \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Therefore, each component z_{it} of the model is unit-root nonstationary and follows an ARIMA(0,1,1) model.

However, we can consider a linear transformation by defining

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{z}_t,$$

$$\begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{a}_t.$$

The VARMA model of the transformed series $\boldsymbol{y}_t$ can be obtained as follows:

$$\begin{aligned} \boldsymbol{L}\boldsymbol{z}_t &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{z}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{z}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{z}_{t-1}) + \boldsymbol{b}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{b}_{t-1}. \end{aligned}$$

Thus, the model for $\boldsymbol{y}_t$ is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} - \begin{bmatrix} 1.0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} - \begin{bmatrix} 0.4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_{1,t-1} \\ b_{2,t-1} \end{bmatrix}. \tag{3.6}$$

From the prior model, we see that (a) y_{1t} and y_{2t} are not dynamically related, except for concurrent correlations between b_{1t} and b_{2t}, (b) y_{1t} follows a univariate ARIMA(0,1,1) model, and (c) $_{2t}$ is a stationary series. In fact, y_{2t} is a white noise series. Consequently, there is only one unit root in $\boldsymbol{z}_t$ even though both z_{it} are unit-root nonstationary. In other words, the unit roots in z_{it} are from the same source y_{1t}, which is referred to as the *common trend* of z_{it}. The linear combination $y_{2t} = (0.5, 1)\boldsymbol{z}_t$ is stationary so that $(0.5, 1)'$ is a co-integrating vector for $\boldsymbol{z}_t$. If the co-integration relationship is imposed, the forecasts $\boldsymbol{z}_T(\ell)$ must satisfy the constraint $(0.5, 1)\boldsymbol{z}_T(\ell) = 0$.

3.4 An Error-Correction Form

Consider the model in Eq. (3.5). Since each component z_{it} has a unit root, one is tempting to take the first difference. Let $\Delta \boldsymbol{z}_t = (\boldsymbol{I} - \boldsymbol{I}B)\boldsymbol{z}_t$ be the first differenced series of $\boldsymbol{z}_t$. Using $|\boldsymbol{\phi}(B)| = (1 - B)$, we have

$$\begin{bmatrix} 1-0.5B & B \\ 0.25B & 1-.5B \end{bmatrix}^{-1} = \frac{1}{1-B}\begin{bmatrix} 1-0.5B & -B \\ -0.25B & 1-0.5B \end{bmatrix}.$$

It is then easy to see that the model for $\Delta \boldsymbol{z}_t$ is

$$\begin{aligned} \Delta \boldsymbol{z}_t &= \begin{bmatrix} 1-0.5B & -B \\ -0.25B & 1-0.5B \end{bmatrix}\begin{bmatrix} 1-0.2B & 0.4B \\ 0.1B & 1-0.2B \end{bmatrix}\boldsymbol{a}_t \\ &= \begin{bmatrix} 1-0.7B & -0.6B \\ -0.15B & 1-.7B \end{bmatrix}\boldsymbol{a}_t \\ &= \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.7 & 0.6 \\ 0.15 & 0.7 \end{bmatrix}\begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}. \end{aligned}$$

Thus, $\Delta \boldsymbol{z}_t$ follows a VMA(1) model. Furthermore, it is easy to show that the eigenvalues of the MA(1) matrix is 1.0 and 0.4. Thus, the VMA(1) model is not invertible. This result implies that differencing every components of a co-integrated system leads to a non-invertible model. This phenomenon is called *over-differencing* in the time series literature. As mentioned before, non-invertible model is hard to estimate.

To avoid the non-invertibility, Engle and Granger (1987) proposed the error-correction form of multivariate time series that keeps the MA strucutre of the model. To illustrate, consider the model in Eq. (3.5). Moving the AR(1) part of the model to the right hand side of the equation and subtracting $\boldsymbol{z}_{t-1}$ from the model, we have

$$\begin{aligned} \Delta \boldsymbol{z}_t &= \left(\begin{bmatrix} 0.5 & -1 \\ -0.25 & 0.5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\boldsymbol{z}_{t-1} - \boldsymbol{a}_t - \boldsymbol{\theta}\boldsymbol{a}_{t-1} \\ &= \begin{bmatrix} -0.5 & -1 \\ -0.25 & -0.5 \end{bmatrix}\boldsymbol{z}_{t-1} + \boldsymbol{a}_t - \boldsymbol{\theta}\boldsymbol{a}_{t-1} \\ &= \begin{bmatrix} -1 \\ -0.5 \end{bmatrix}[0.5, 1]\boldsymbol{z}_{t-1} + \boldsymbol{a}_t - \boldsymbol{\theta}\boldsymbol{a}_{t-1}. \end{aligned}$$

This is an *error-correction* form for the model, which has an invertible MA structure, but uses $\boldsymbol{z}_{t-1}$ in the right hand side of the model. The $\boldsymbol{z}_{t-1}$ term is referred to as the error-correction term and its coefficient matrix is of rank 1 representing the number of co-integrating vectors of the system. Given a co-integrated linear VARMA model

$$\boldsymbol{\phi}(B)\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\theta}(B)\boldsymbol{a}_t,$$

where $\boldsymbol{\phi}(B) = \boldsymbol{I} - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i$ and the the model is assumed to be identifiable. The co-integration assumptions implies that $\phi(1) = \boldsymbol{I} - \sum_{i=1}^{p} \boldsymbol{\phi}_i$ is a singular matrix. An *error-correction form* of the model can be obtained by subtracting $\boldsymbol{z}_{t-1}$ from both sides of the model. Some algebra shows that the resulting model is in the form

$$\Delta \boldsymbol{z}_t = \boldsymbol{\Pi} \boldsymbol{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\phi}_i^* \Delta \boldsymbol{z}_{t-i} + \boldsymbol{\phi}_0 + \boldsymbol{\theta}(B)\boldsymbol{a}_t, \tag{3.7}$$

where the coefficient matrics are given by

$$\begin{aligned}
\boldsymbol{\Pi} &= \sum_{i=1}^{p} \boldsymbol{\phi}_i - \boldsymbol{I} = -\boldsymbol{\phi}(1) \\
\boldsymbol{\phi}_{p-1}^* &= -\boldsymbol{\phi}_p \\
\boldsymbol{\phi}_{p-2}^* &= -\boldsymbol{\phi}_{p-1} - \boldsymbol{\phi}_p \\
\vdots &= \vdots \\
\boldsymbol{\phi}_1^* &= -\boldsymbol{\phi}_2 - \cdots - \boldsymbol{\phi}_p.
\end{aligned}$$

Note that the MA part of the model remain unchanged.

Remark: There are many ways to write an error-correction form. For instance, instead of $\boldsymbol{z}_{t-1}$, one can subtracting $\boldsymbol{z}_{t-p}$ from the given VARMA(p,q) model and obtain another error-correction form as

$$\Delta \boldsymbol{z}_t = \boldsymbol{\Pi} \boldsymbol{z}_{t-p} + \sum_{i=1}^{p-1} \boldsymbol{\phi}_i^* \Delta \boldsymbol{z}_{t-i} + \boldsymbol{\phi}_0 + \boldsymbol{\theta}(B)\boldsymbol{a}_t, \tag{3.8}$$

where $\boldsymbol{\Pi} = -\boldsymbol{\phi}(1)$ and the $\boldsymbol{\phi}_i^*$ are given by

$$\begin{aligned}
\boldsymbol{\phi}_1^* &= \boldsymbol{\phi}_1 - \boldsymbol{I} \\
\boldsymbol{\phi}_2^* &= \boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 - \boldsymbol{I} \\
\vdots &= \vdots \\
\boldsymbol{\phi}_{p-1}^* &= \boldsymbol{\phi}_1 + \cdots + \boldsymbol{\phi}_{p-1} - \boldsymbol{I}.
\end{aligned}$$

Since $\boldsymbol{\phi}(1)$ is a singular matrix for a co-integrated system, $\boldsymbol{\Pi}$ is not full rank. Assume that Rank($\boldsymbol{\Pi}$) $= m$. Then, there exist $k \times m$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of rank m such that $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$. This decomposition, however, is not unique. In fact, for any $m \times m$ orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$, it is easy to see that

$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\boldsymbol{P}\boldsymbol{P}'\boldsymbol{\beta}' = (\boldsymbol{\alpha}\boldsymbol{P})(\boldsymbol{\beta}\boldsymbol{P})'.$$

Thus, $\boldsymbol{\alpha}\boldsymbol{P}$ and $\boldsymbol{\beta}\boldsymbol{P}$ are of rank m and may serve as another decomposition of $\boldsymbol{\Pi}$.

The columns of the matrix $\boldsymbol{\beta}$ are co-integrating vectors. Thus, there are m co-integrating vectors for $\boldsymbol{z}_t$, and the system has $k - m$ unit roots.