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 \(\{a_t\}\) satisfies (a) \(E(a_t|F_{t-1}) = 0\), (b) \(E(a_t a_t'|F_{t-1}) = \Sigma > 0\), and (c) \(a_t\) has finite fourth moments, where \(F_{t-1}\) denotes the \(\sigma\)-field generated by \(\{z_{t-1}, z_{t-2}, \cdots\}\). 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 strongly consistent and asymptotically normally distributed. Here MLE denotes estimates obtained by maximizing the multivariate normal density function.

Next, we discuss more on the impulse response functions of VARMA models and introduce the forecast error variance decomposition.

0.2 Impulse Response Functions

For a stationary VARMA\((p,q)\) model, \(\phi(B)(z_t - \mu) = \theta(B)a_t\), the MA model representation is

\[
  z_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}
\]  

(1)

where \(\text{Var}(a_t) = \Sigma\) and \(\psi_0 = I\). As discussed before in the VAR\((p)\) case, the matrices \(\psi_i\) can be obtained by equating the coefficient matrices of \(B^i\) in the equation

\[
  \phi(B) \sum_{i=0}^{\infty} \psi_i B^i = \theta(B).
\]

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

\[\Sigma = LGL' = PP',\]

where \(G\) is a diagonal matrix, \(L\) is a lower triangular matrix with unit diagonal elements, and \(P = LG^{1/2}\). Let \(b_t = P^{-1}a_t\). Then, \(\text{Var}(b_t) = I\) so that elements of \(b_t\) are uncorrelated with unit variance. The MA representation in Eq. (1) can be rewritten as

\[
  z_t = \mu + PP^{-1}a_t + \sum_{i=1}^{\infty} \psi_i PP^{-1}a_{t-i}
  = \mu + \Psi_0 b_t + \sum_{i=1}^{\infty} \Psi_i b_{t-i},
\]

(2)
where $\Psi_0 = P$ and $\Psi_i = \psi_i P$ for $i > 0$. Note that $\Psi_0$ 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 $\Psi_i$ thus depend on the ordering of elements in $z_t$. Different orderings give rise to different matrices $\Psi_i$, and there are $k!$ possible orderings of $z_t$. Since elements of $b_t$ 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 $\Psi_{\ell}$. Consequently, $\Psi_i$ are the impulse response functions of $z_t$ with respect to $b_t$. As mentioned earlier, these impulse response functions depend on the ordering of elements of $z_t$.

### 0.3 Forecast Error Variance Decomposition (FEVD)

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

$$e_n(\ell) = \sum_{m=0}^{\ell-1} \Psi_m b_{n+\ell-m}. $$

For the $i$th element of $z_t$, the forecast error is

$$e_{i,n}(\ell) = \sum_{m=0}^{\ell-1} \sum_{j=1}^{k} \Psi_{m,ij} b_{j,t+\ell-m} = \sum_{m=0}^{\ell-1} \psi_{m,11} b_{1,t+\ell-m} + \cdots + \sum_{m=0}^{\ell-1} \psi_{m,ik} b_{k,t+\ell-m}. $$

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

$$\text{Var}[e_{i,n}(\ell)] = \left( \sum_{m=0}^{\ell-1} \psi_{m,11}^2 \right) + \cdots + \left( \sum_{m=0}^{\ell-1} \psi_{m,ik}^2 \right).$$

Consequently, the portion of the variance of $e_{i,n}(\ell)$ due to the shock $\{b_{jt}\}$ is

$$\text{FEVD}_{i,j}(\ell) = \frac{\left( \sum_{m=0}^{\ell-1} \psi_{m,ij}^2 \right)}{\left( \sum_{m=0}^{\ell-1} \psi_{m,11}^2 \right) + \cdots + \left( \sum_{m=0}^{\ell-1} \psi_{m,ik}^2 \right)}, \quad i, j = 1, \cdots, k.$$

Again, the FEVD depends on the ordering of the elements of $z_t$.

**Remark:** Impulse response 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 $\psi_i$ of the matrix polynomial $[\phi(B)]^{-1} \theta(B) = \psi(B)$ from the command nopsiweight in MFORE. The Cholesky decomposition can be obtained in SCA by the command chol, e.g., $b = \text{chol}(a)$, where $b$ is an upper triangular matrix such that $a = b^t b$. 

2
Chapter 3

Unit-Root Nonstationary VARMA Models

Unit root plays an important role both in theory and applications of time series analysis. Properties of unit-root nonstationary time series and the practical implications of unit roots have attracted much attention in the time series and econometric literature. In this chapter, we shall discuss some of the fundamental properties of unit roots in multivariate time series. We begin with a simple case.

3.1 Multivariate exponential smoothing

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

\[(1 - B)z_t = (I - \theta B)a_t,\]  \hspace{1cm} (3.1)

where, for simplicity, \(z_0 = 0\). Rewrite the model as

\[z_t = a_t - \theta a_{t-1} + z_{t-1}.\]

Then, by repeatedly substitutions, we obtain

\[z_t = a_t + (I - \theta)(a_{t-1} + a_{t-2} + \cdots),\]

where it is understood that \(a_t = 0\) for \(t \leq 0\). Thus, the model has strong memory because the coefficient matrix \(\psi_i = I - \theta\) does not converge to zero as \(i\) increases. Next, from the model, we have

\[(I + \theta B + \theta^2 B^2 + \cdots)(I - IB)z_t = a_t.\]

That is,

\[z_t = (I - \theta)z_{t-1} + (I - \theta)\theta z_{t-2} + (I - \theta)\theta^2 z_{t-3} + \cdots + a_t.\]  \hspace{1cm} (3.2)

From Eq. (3.2), it is easy to see that the 1-step ahead forecast of \(z_t\) at the forecast origin \(t - 1\) is

\[
\hat{z}_{t-1}(1) = (I - \theta)z_{t-1} + (I - \theta)\theta z_{t-2} + (I - \theta)\theta^2 z_{t-3} + \cdots = (I - \theta)[z_{t-1} + \theta z_{t-2} + \theta^2 z_{t-3} + \cdots].
\]
This is a generalization of the univariate exponential smoothing model except that the weights decay in matrix form, not in terms of elements of θ.

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

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

where $\Theta$ and $\text{Var}(b_{it})$ can be obtained from $\theta$ and $\Sigma$. Since the models are invertible, the series $z_{it}$ is commonly referred to as an I(1) process in the econometric literature. That is, it is an integrated processes of order 1.

### 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 $|I - IB|$ are on the unit circle. In general, a VARMA($p,q$) model is said to be unit-root nonstationary if $|\psi(x)| \neq 0$ for $|x| < 1$, but $|\psi(1)| = 0$. Since the $z_{it}$ series is invertible, it is commonly referred to as an I(1) process in the econometric literature. That is, it is an integrated process of order 1.

The unit-root processes and co-integration have attracted lots of research attention in the 1980s and 1990s. In particular, various methods have been proposed for unit-root and co-integration tests. In what follows, we introduce some basic results of unit-root statistics. Our goal is to provide readers some basic knowledge concerning unit roots and cointegration, and to introduce some applications. It is impossible to include all the available results for unit-root theory and cointegration.

### 3.3 Review of Univariate Results

We begin with a brief review of estimation and testing of unit roots in a univariate time series. We adopt the approach of Phillips (1987) with a single unit root. The case of unit roots with multiplicity greater than 1 can be handled via the work of Chan and Wei (1988).

Consider the discrete-time process \{z_t\} generated by

$$z_t = \pi z_{t-1} + y_t, \quad t = 1, 2, \ldots \quad (3.3)$$

where $\pi = 1$, $z_0$ is a fixed real number, and $y_t$ is a stationary time series to be defined shortly. It will become clear later that the starting value $z_0$ has no effect on the limiting distributions discussed in this chapter.

Define the partial sum of \{y_t\}

$$S_t = \sum_{i=1}^{t} y_i. \quad (3.4)$$

For simplicity, we define $S_0 = 0$. Then, $z_t = S_t + z_0$. A fundamental result to unit-root theory is the limiting behavior of $S_t$. To this end, one must properly standardize the partial sum $S_T$ as $T \to \infty$. It is common in the literature to employ the average variance of $S_T$ given by

$$\sigma^2 = \lim_{T \to \infty} E(T^{-1}S_T^2) \quad (3.5)$$
which is assumed to exist and positive. Define
\[ X_T(r) = \frac{1}{\sqrt{T\sigma}} S_{[Tr]}, \quad 0 \leq r \leq 1, \quad (3.6) \]
where \([Tr]\) denotes the integer part of \(Tr\). In particular, \(X_T(1) = \frac{1}{\sqrt{T\sigma}} S_T\). Under certain conditions, \(X_T(r)\) is shown to converge weakly to the well known standard Brownian motion or the Wiener process. This is commonly referred to as the functional central limit theorem.

**Basic assumption A**: Assume that \(\{y_t\}\) is a stationary time series such that (a) \(E(y_t) = 0\) for all \(t\), (b) \(\sup_t E(|y_t|^\beta) < \infty\) for some \(\beta > 2\), (c) the average variance \(\sigma^2\) of Eq. (3.5) exists and is positive, and (d) \(y_t\) is strong mixing with mixing coefficients \(\alpha_m\) that satisfy
\[ \sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty. \]

Strong mixing is a measure of serial dependence of a time series \(\{y_t\}\). Let \(F_{-\infty}^q\) and \(F_{\infty}^r\) be the \(\sigma\)-field generated by \(\{y_q, y_{q-1}, \ldots\}\) and \(\{y_r, y_{r+1}, \ldots\}\), respectively. That is, \(F_{-\infty}^q = F\{y_q, y_{q-1}, \ldots\}\) and \(F_{\infty}^r = F\{y_r, y_{r+1}, \ldots\}\). We say that \(y_t\) satisfies a strong mixing condition if there exists a positive function \(\alpha(.)\) satisfying \(\alpha(n) \to 0\) as \(n \to \infty\) so that
\[ |P(A \cap B) - P(A)P(B)| < \alpha_{r-q}, \quad A \in F_{-\infty}^q, \quad B \in F_{\infty}^r. \]
If \(y_t\) is strong mixing, then the serial dependence between \(y_t\) and \(y_{t-h}\) approaches zero as \(h\) increases.

The following two theorems are widely used in unit root study. See Herrndorf (1983) and Billingsley (1968) for more details.

**Functional Central Limit Theorem (FCLT)**: If \(\{y_t\}\) satisfies the basic assumption A, then \(X_T(r) \Rightarrow W(r)\), where \(W(r)\) is a standard Brownian motion for \(r \in [0,1]\) and \(\Rightarrow\) denotes weak convergence, i.e., convergence in distribution.

**Continuous Mapping Theorem**: If \(X_T(r) \Rightarrow W(r)\) and \(h(.)\) is a continuous functional on \(D[0,1]\), the space of all real valued functions on \([0,1]\) that are right continuous at each point on \([0,1]\) and have finite left limits, then \(h(X_T(r)) \Rightarrow h(W(r))\) as \(T \to \infty\).

The ordinary least squares estimate of \(\pi\) in Eq. (3.3) is
\[ \hat{\pi} = \frac{\sum_{t=1}^{T} z_{t-1} z_t}{\sum_{t=1}^{T} z_{t-1}^2}, \]
and its variance is estimated by
\[ \text{var}(\hat{\pi}) = \frac{s^2}{\sum_{t=1}^{T} z_{t-1}^2}, \]
where \(s^2\) is the residual variance given by
\[ s^2 = \frac{1}{T-1} \sum_{t=1}^{T} (z_t - \hat{\pi} z_{t-1})^2. \quad (3.7) \]
The usual $t$-ratio for testing the null hypothesis $H_o : \pi = 1$ versus $H_a : \pi < 1$ is given by

$$t_\pi = \left( \frac{\sum_{t=1}^{T} z_{t-1}^2}{s} \right) \frac{1/2}{\frac{\hat{\pi} - 1}{s\sqrt{\sum_{t=1}^{T} z_{t-1}^2}}} = \frac{\sum_{t=1}^{T} z_{t-1}y_t}{s\sqrt{\sum_{t=1}^{T} z_{t-1}^2}}. \quad (3.8)$$

We have the following basic results of unit-root process $z_t$.

**Theorem:** Suppose that $\{y_t\}$ satisfies the basic assumption A and $\sup_t E|y_t|^{3+\eta} < \infty$, where $\beta > 2$ and $\eta > 0$, then as $T \to \infty$, we have

(a) $T^{-2} \sum_{t=1}^{T} z_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr$,

(b) $T^{-1} \sum_{t=1}^{T} z_{t-1}y_t \Rightarrow \frac{\sigma^2}{T} (W(1)^2 - \frac{\sigma^2}{T})$,

(c) $T(\hat{\pi} - 1) \Rightarrow \frac{(1/2)(W(1)^2 - (\sigma_y^2/\sigma^2))}{\int_0^1 W(r)^2 dr}$,

(d) $\hat{\pi} \rightarrow_p 1$, where $\rightarrow_p$ denotes convergence in probability,

(c) $t_\pi \Rightarrow \frac{(\sigma/(2\sigma_y))|W(1)^2 - (\sigma_y^2/\sigma^2)|}{\left[\int_0^1 W(r)^2 dr\right]^{1/2}}$,

where $\sigma^2$ and $\sigma_y^2$ are defined as

$$\sigma^2 = \lim_{T \to \infty} E(T^{-1}S_T^2), \quad \sigma_y^2 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E(y_t^2).$$

**Proof.** For part (a), we have

$$T^{-2} \sum_{t=1}^{T} z_{t-1}^2 = T^{-2} \sum_{t=1}^{T} (S_{t-1} + z_0)^2$$

$$= T^{-2} \sum_{t=1}^{T} (S_{t-1}^2 + 2z_0S_{t-1} + z_0^2)$$

$$= \sigma^2 \sum_{t=1}^{T} \left( \frac{1}{\sigma\sqrt{T}} S_{t-1} \right)^2 \frac{1}{T} + 2z_0\sigma T^{-1/2} \sum_{t=1}^{T} \left( \frac{1}{\sigma\sqrt{T}} S_{t-1} \right) \frac{1}{T} + T^{-1}z_0^2$$

$$= \sigma^2 \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \left( \frac{1}{\sigma\sqrt{T}} S_{[T]} \right)^2 dr + 2z_0\sigma T^{-1/2} \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \frac{1}{\sigma\sqrt{T}} S_{[T]} dr + T^{-1}z_0^2$$

$$= \sigma^2 \int_0^1 X_T^2(r) dr + 2z_0\sigma T^{-1/2} \int_0^1 X_T(r) dr + T^{-1}z_0^2$$

$$\Rightarrow \sigma^2 \int_0^1 W(r)^2 dr, \quad T \to \infty.$$
\begin{align*}
T \sum_{t=1}^{T} S_{t-1} y_t + z_0 \bar{y} \\
= T^{-1} \sum_{t=1}^{T} \frac{1}{2} (S_t^2 - S_{t-1}^2 - y_t^2) + z_0 \bar{y} \\
= (2T)^{-1} S_T^2 - (2T)^{-1} \sum_{t=1}^{T} y_t^2 + z_0 \bar{y} \\
= \frac{\sigma^2}{2} X_T(1)^2 - \frac{1}{2} T^{-1} \sum_{t=1}^{T} y_t^2 + z_0 \bar{y} \\
\Rightarrow \frac{\sigma^2}{2} \left( W(1)^2 - \frac{\sigma_y^2}{\sigma^2} \right),
\end{align*}

because \( \bar{y} \to 0 \) and \( T^{-1} \sum_{t=1}^{T} y_t^2 \to \sigma_y^2 \) almost surely as \( T \to \infty \).

Part (c) follows parts (a) and (b) and the continuous mapping theorem. Part (d) follows part (c).

In particular, \( \hat{\phi} \) converges to 1 at the rate of \( T^{-1} \), not the usual rate \( T^{-1/2} \). This is referred to as the super consistency in the unit-root literature.

Using the fast convergence rate of \( \hat{\pi} \), parts (a) and (b), and Eq. (3.7), we have

\begin{align*}
s^2 &= \frac{1}{T - 1} \sum_{t=1}^{T} (z_t - \hat{\pi} z_{t-1})^2 \\
&= \frac{1}{T - 1} \sum_{t=1}^{T} [(z_t - z_{t-1}) + (1 - \hat{\pi}) z_{t-1}]^2 \\
&= \frac{1}{T - 1} \sum_{t=1}^{T} y_t^2 + 2(1 - \hat{\pi})(T - 1)^{-1} \sum_{t=1}^{T} z_{t-1} y_t + (T - 1)^{-1}(1 - \hat{\pi})^2 \sum_{t=1}^{T} \sigma_y^2 \\
&\to_p \sigma_y^2,
\end{align*}

because the last two terms vanish as \( T \to \infty \). Thus, the \( t \)-ratio can be written as

\begin{align*}
t_\pi &= \frac{\sum_{t=1}^{T} z_{t-1} y_t}{\sigma_y (\sum_{t=1}^{T} z_{t-1}^2)^{1/2}} \\
&= \frac{\sigma - 2T^{-1} \sum_{t=1}^{T} z_{t-1} y_t}{\sigma_y [(\sigma - 2T^{-1})^2 \sum_{t=1}^{T} z_{t-1}^2]^{1/2}}.
\end{align*}

Using the results of parts (a) and (b), and continuous mapping theorem, we have

\begin{align*}
t_\pi \to_d \frac{\sigma}{\sigma_y} \left[ W(1)^2 - \frac{\sigma_y^2}{\sigma^2} \right] \\
\int_0^1 W(r)^2 dr)^{1/2}.
\end{align*}

In what follows, we consider the unit-root theory for some special cases. Since \( y_t \) is stationary with zero mean, we have

\begin{equation}
E(S_T^2) = T \gamma_0 + 2 \sum_{i=1}^{T-1} (T - i) \gamma_i \tag{3.9}
\end{equation}

where \( \gamma_i \) is the lag-\( i \) autocovariance of \( y_t \).
3.3.1 AR(1) case

Consider the simple AR(1) model \( z_t = \pi z_{t-1} + a_t \), i.e., \( y_t = a_t \) being an iid sequence of random variables with mean zero and variance \( \sigma_a^2 > 0 \). In this case, it is easy to see that, from Eq. (3.9),

\[
\sigma^2 = \lim_{T \to \infty} E(T^{-1}S^2_T) = \sigma_a^2, \quad \sigma_y^2 = \sigma_a^2.
\]

Consequently, for the random-walk model \( z_t = z_{t-1} + a_t \), we have

1. \( T^{-2} \sum_{t=1}^{T} z_{t-1}^2 \to \sigma_a^2 \int_0^1 W^2(r)dr \),
2. \( T^{-1} \sum_{t=1}^{T} z_{t-1}(z_t - z_{t-1}) \to \frac{\sigma_a^2}{2}[W^2(1) - 1], \)
3. \( T(\hat{\pi} - 1) \to \frac{0.5[W^2(1)-1]}{\int_0^1 W^2(r)dr}, \)
4. \( t_\pi \to_d \frac{0.5[W^2(1)-1]}{[\int_0^1 W^2(r)dr]^{1/2}}. \)

The critical values of \( t_\pi \) has been tabulated by several authors. See, for instance, Fuller (1976, Table 8.5.2).

3.3.2 AR(p) Case

We start with the AR(2) case in which \( (1 - B)(1 - \phi B)z_t = a_t \), where \( |\phi| < 1 \). The model can be written as

\[ z_t = z_{t-1} + y_t, \quad y_t = \phi y_{t-1} + a_t. \]

For the stationary AR(1) process \( y_t \), \( \sigma_y^2 = \sigma_a^2/(1 - \phi^2) \) and \( \gamma_i = \phi^i \gamma_0 \). Thus, by Eq. (3.9), \( \sigma^2 = \sigma_a^2(1 + \phi)/(1 - \phi) \). Consequently, the limiting distributions discussed depend on the AR(1) coefficient \( \phi \). For instance, the \( t \)-ratio of \( \hat{\pi} \) becomes

\[ t_\pi \to \frac{(1+\phi^2)[W(1)^2 - \frac{1}{(1+\phi)^2}]}{[\int_0^1 W(r)^2dr]^{1/2}}. \]

Such a dependence makes it difficult to use \( t_\pi \) in unit-root testing, because the asymptotic critical values depend on the nuisance parameter \( \phi \). This dependence continues to hold for the general AR(p) process \( y_t \). To overcome this difficulty, Said and Dickey (1984) consider the augmented Dickey-Fuller test statistic.

For a higher-order AR(p) process, \( \phi(B)z_t = a_t \), we focus on the case that \( \phi(B) = \phi^*(B)(1 - B) \) where \( \phi^*(B) \) 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 \( \pi = 1 \) in the model

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

Or equivalently, the same as testing for \( \rho = 1 = 0 \) in the model

\[ \Delta z_t = (\pi - 1)z_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta z_{t-j} + a_t, \]
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) $\pi - 1 = -\phi(1) = \sum_{i=1}^{p} \phi_i - 1$ and $\phi^*_j = -\sum_{i=j+1}^{p} \phi_i$. In practice, the linear model

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

(3.10)

where $\beta = \pi - 1$, is used. The least squares estimate of $\beta$ can then be used in unit-root testing. Specifically, testing $H_o : \pi = 1$ versus $H_a : \pi < 1$ is equivalent to testing $H_o : \beta = 0$ versus $H_a : \beta < 0$. It can be shown that the $t$-ratio of $\hat{\beta}$ (against 0) has the same limiting distribution as $t_\pi$ in the random-walk case. In other words, for an AR($p$) model with $p > 1$, by including the lagged variables of $\Delta z_t$ in the linear regression of Eq. (3.10), one can remove the nuisance parameters in unit-root testing. This is the well-known augmented Dickey-Fuller unit-root test. Furthermore, the limiting distribution of the LS estimates $\hat{\phi}^*_j$ in Eq. (3.10) is the same as that of fitting an AR($p-1$) model to $\Delta z_t$. In other words, limiting properties of the estimates for the stationary part remain unchanged when we treat the unit-root as known a priori.

**Remark:** In our discussion, we assume that there is no constant term in the model. One can include a constant term in the model and obtain the associated limiting distribution for unit-root testing. The limiting distribution will be different, but the idea remains the same.

### 3.3.3 MA(1) Case

Next, assume that $z_t = z_{t-1} + y_t$ and $y_t = a_t - \theta a_{t-1}$ with $|\theta| < 1$. In this case, we have $\gamma_0 = (1+\theta^2)\sigma^2_a$, $\gamma_1 = -\theta \sigma^2_a$, and $\gamma_i = 0$ for $i > 1$. Consequently, $\sigma^2_y = (1+\theta^2)\sigma^2_a$ and, by Eq. (3.9), $\sigma^2 = (1-\theta)^2 \sigma^2_a$. The limiting distributions of unit-root statistics become

1. $T^{-2} \sum_{t=1}^{T} z_{t-1}^2 \Rightarrow (1-\theta^2)\sigma^2_a \int_{0}^{1} W^2(r)dr$, 
2. $T^{-1} \sum_{t=1}^{T} (z_{t} - z_{t-1}) \Rightarrow \frac{(1-\theta^2)\sigma^2_a}{2} [W^2(1) - \frac{1+\theta^2}{(1-\theta)^2}]$, 
3. $T(\pi - 1) \Rightarrow \frac{1-\theta}{2\sqrt{1+\theta^2}} [W^2(1) - \frac{1+\theta^2}{(1-\theta)^2}] \int_{0}^{1} W^2(r)dr$, 
4. $t_\pi \rightarrow d \frac{1-\theta}{2\sqrt{1+\theta^2}} [W^2(1) - \frac{1+\theta^2}{(1-\theta)^2}] \int_{0}^{1} W^2(r)dr]^{1/2}$.

From the results, it is clear that when $\theta$ is close to 1 the asymptotic behavior of $t_\pi$ is rather different from that of the case when $y_t$ is a white noise series. This is not surprising because when $\theta$ approaches 1, the $z_t$ process is close to being a white noise series. This might explain the severe size distortions of Phillips-Perron unit-root test statistics seen in Table 1 of Phillips and Perron (1988).

### 3.3.4 Unit-Root Tests

Suppose that the univariate time series $z_t$ follows the AR($p$) model $\phi(B)z_t = a_t$, where $\phi(B) = (1 - B)\phi^*(B)$ such that $\phi^*(1) \neq 0$. That is, $z_t$ has a single unit root. In this subsection, we summarize the framework of unit-root tests commonly employed in the literature. Three models are often employed in the test. They are
1. No constant:
\[ \Delta z_t = \beta z_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta z_{t-i} + a_t. \] (3.11)

2. With constant:
\[ \Delta z_t = \alpha + \beta z_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta z_{t-i} + a_t. \] (3.12)

3. With time trend:
\[ \Delta z_t = \omega_0 + \omega_1 t + \beta z_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta z_{t-i} + a_t. \] (3.13)

The null hypothesis is \( H_0 : \beta = 0 \) and the alternative hypothesis is \( H_a : \beta < 0 \). It turns out that the limiting distributions of the \( t \)-ratios are different for the three models employed, even though they are all functions of the standard Brownian motion under the null hypothesis of a single unit root. Critical values of the \( t \)-ratios have been obtained via simulation in the literature. See, for example, Fuller (1976, Table 8.5.2).

### 3.3.5 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 the package \texttt{fUnitRoots} of R. The three models discussed above are denoted by type being \texttt{nc}, \texttt{c}, \texttt{ct}, respectively.

```r
> setwd("C:/Users/rst/teaching/mts/sp2011")
> library(fUnitRoots)
> 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
```
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

3.4 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 \((1 - B)^dz_t\) 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 \( z_t \). If \( z_{it} \) are \( I(1) \) processes, but a non-trivial linear combination \( \beta'z_t \) is \( I(0) \), then \( z_t \) is said to be co-integrated of order 1. Basically, if \( z_{it} \) are \( I(d) \) nonstationary and \( \beta'z_t \) is \( I(h) \) with \( h < d \), then \( z_t \) 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 \( \beta \) is called a co-integrating vector.

Suppose that \( z_t \) is unit-root nonstationary such that the marginal models for \( z_{it} \) have a unit root. If \( \beta \) is a \( k \times m \) matrix of full rank \( m \), such that \( w_t = \beta'z_t \) is \( I(0) \), then \( z_t \) is a co-integrated system with \( m \) co-integrating vectors, which are the columns of \( \beta \). This means that
there are \( k - m \) unit roots in \( z_t \). For the given full-rank \( k \times m \) matrix \( \beta \) with \( m < k \), let \( \beta_\perp \) be a \( k \times (k - m) \) full-rank matrix such that \( \beta'_\perp \beta_\perp = 0 \). Then, \( n_t = \beta'_\perp z_t \) is a unit-root nonstationary. The components \( n_{it} \ (i = 1, \ldots, (k - m)) \) are referred to as the common trends of \( z_t \). 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 \( w_t = \beta' z_t \) is stationary, it is mean-reverting so that the \( \ell \)-step ahead forecast of \( w_{T+\ell} \) at the forecast origin \( T \) satisfies

\[
\hat{w}_T(\ell) \to_p E(w_t) \equiv \mu_w, \quad \ell \to \infty.
\]

This implies that \( \beta' \hat{z}_T(\ell) \to \mu_w \) as \( \ell \) increases. Thus, point forecasts of \( z_t \) satisfy a long-term stable constraint.

### 3.4.1 An example of co-integration

To understand co-integration, we consider a simple example. Suppose that the bivariate process \( z_t \) follows the model

\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix}
= \begin{bmatrix}
  0.5 & -1.0 \\
  -0.25 & 0.5
\end{bmatrix}
\begin{bmatrix}
  z_{1,t-1} \\
  z_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
  a_{1t} \\
  a_{2t}
\end{bmatrix},
\]

where the covariance matrix \( \Sigma \) of the shock \( a_t \) is positive definite. For simplicity, assume that \( \Sigma = I \). 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

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

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

\[
\begin{bmatrix}
  y_{1t} \\
  y_{2t}
\end{bmatrix}
= \begin{bmatrix}
  1.0 & -2.0 \\
  0.5 & 1.0
\end{bmatrix}
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} \equiv Lz_t,
\]

\[
\begin{bmatrix}
  b_{1t} \\
  b_{2t}
\end{bmatrix}
= \begin{bmatrix}
  1.0 & -2.0 \\
  0.5 & 1.0
\end{bmatrix}
\begin{bmatrix}
  a_{1t} \\
  a_{2t}
\end{bmatrix} \equiv La_t.
\]

The VARMA model of the transformed series \( y_t \) can be obtained as follows:

\[
Lz_t = L\Phi z_{t-1} + La_t - L\Theta a_{t-1}
= L\Phi L^{-1}z_{t-1} + La_t - L\Theta L^{-1}La_{t-1}
= L\Phi L^{-1}(Lz_{t-1}) + b_t - L\Theta L^{-1}b_{t-1}.
\]
Thus, the model for $y_t$ is

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

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) $y_{2t}$ is a stationary series. In fact, $y_{2t}$ is a white noise series. Consequently, there is only one unit root in $z_t$ 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 $y_{2t} = (0.5, 1)z_t$ is stationary so that $(0.5, 1)'$ is a co-integrating vector for $z_t$. If the co-integration relationship is imposed, the forecasts $z_T(\ell)$ must satisfy the constraint $(0.5, 1)z_T(\ell) = 0$.

### 3.5 An Error-Correction Form

Consider the model in Eq. (3.14). Since each component $z_{it}$ has a unit root, one is tempting to take the first difference. Let $\Delta z_t = (I - IB)z_t$ be the first differenced series of $z_t$. Using $|\phi(B)| = (1 - B)$, we have

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

It is then easy to see that the model for $\Delta z_t$ is

$$
\Delta z_t = \begin{bmatrix} 1 - 0.5B & -B \\ -0.25B & 1 - 0.5B \end{bmatrix} \begin{bmatrix} 1 - 0.2B & 0.4B \\ 0.1B & 1 - 0.2B \end{bmatrix} a_t
$$

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

Thus, $\Delta z_t$ 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 structure of the model. To illustrate, consider the model in Eq. (3.14). Moving the AR(1) part of the model to the right hand side of the equation and subtracting $z_{t-1}$ from the model, we have

$$
\Delta z_t = \left( \begin{bmatrix} 0.5 & -1 \\ -0.25 & 0.5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) z_{t-1} - a_t - \theta a_{t-1}
$$

$$
= \begin{bmatrix} -0.5 & -1 \\ -0.25 & -0.5 \end{bmatrix} z_{t-1} + a_t - \theta a_{t-1}
$$

14
This is an error-correction form for the model, which has an invertible MA structure, but uses \( z_{t-1} \) in the right hand side of the model. The \( z_{t-1} \) 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

\[
\phi(B)z_t = \phi_0 + \theta(B)a_t,
\]

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

\[
\Delta z_t = \Pi z_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta z_{t-i} + \phi_0 + \theta(B)a_t,
\]

(3.16)

where the coefficient matrices are given by

\[
\Pi = \sum_{i=1}^{p} \phi_i - I = -\phi(1)
\]

\[
\phi^*_{p-1} = -\phi_{p}
\]

\[
\phi^*_{p-2} = -\phi_{p-1} - \phi_{p}
\]

\[
\vdots \quad \vdots
\]

\[
\phi^*_{1} = -\phi_{2} - \cdots - \phi_{p}.
\]

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 \( z_{t-1} \), one can subtracting \( z_{t-p} \) from the given VARMA\((p,q)\) model and obtain another error-correction form as

\[
\Delta z_t = \Pi z_{t-p} + \sum_{i=1}^{p-1} \phi_i^* \Delta z_{t-i} + \phi_0 + \theta(B)a_t,
\]

(3.17)

where \( \Pi = -\phi(1) \) and the \( \phi_i^* \) are given by

\[
\phi^*_{1} = \phi_{1} - I
\]

\[
\phi^*_{2} = \phi_{1} + \phi_{2} - I
\]

\[
\vdots \quad \vdots
\]

\[
\phi^*_{p-1} = \phi_{1} + \cdots + \phi_{p-1} - I.
\]

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

\[
\alpha \beta' = \alpha PP' \beta' = (\alpha P)(\beta P)'.
\]
Thus, $\alpha P$ and $\beta P$ are of rank $m$ and may serve as another decomposition of $\Pi$. The columns of the matrix $\beta$ are co-integrating vectors. Thus, there are $m$ co-integrating vectors for $z_t$, and the system has $k - m$ unit roots.

References


