Multivariate Time Series Analysis and Its Applications
[Tsay (2005), chapter 8]

Insights:
• Price movements in one market can spread easily and instantly to another market [economic globalization and internet communication]
• Financial markets are more dependent on each other than ever before, and one must consider them jointly to better understand the dynamic structure of global finance
  o One market may lead the other market under some circumstances
• In multiple assets, the dynamic relationships between returns of the assets play an important role in decision making
Weak stationarity and cross-correlation matrices

$k$-dimensional time series: $r_t = (r_{1t}, \ldots, r_{kt})'$

$r_t$ is weakly stationary if its first and second moments are time-invariant. Let its means vector and covariance matrix be,

$$\mu = E(r_t), \quad \Gamma_0 = E[(r_t - \mu)(r_t - \mu)'],$$

where the expectation is taken element by element over the joint distribution of $r_t$. The mean $\mu$ is a $k$-dimensional vector consisting of the unconditional expectations of the components of $r_t$. The covariance matrix $\Gamma_0$ is a $k \times k$ matrix. The $i$th diagonal element of $\Gamma_0$ is the variance of $r_{it}$, whereas the $(i, j)$th element of $\Gamma_0$ is the covariance between $r_{it}$ and $r_{jt}$. We write $\mu = (\mu_1, \ldots, \mu_k)'$ and $\Gamma_0 = [\Gamma_{ij}(0)]$ when the elements are needed.
Cross-correlation matrices:

\[ \rho_0 \equiv [\rho_{ij}(0)] = D^{-1} \Gamma_0 D^{-1} \]

\[ D = \text{diag}\{\sqrt{\Gamma_{11}(0)}, \ldots, \sqrt{\Gamma_{kk}(0)}\}. \]

Correlation coefficient between \( r_{it} \) and \( r_{jt} \) (concurrent or contemporaneous):

\[ \rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(r_{it}, r_{jt})}{\text{std}(r_{it})\text{std}(r_{jt})} \]

It’s easy to see that \( \rho_{ij}(0) = \rho_{ji}(0) \), \(-1 \leq \rho_{ij}(0) \leq 1\), and \( \rho_{ij}(0) = 1 \) for \( 1 \leq i, j \leq k \).

So, \( \rho(0) \) is a symmetric matrix with unit diagonal elements.
Lag-$l$ cross-covariance matrix:

$$\Gamma_{\ell} \equiv [\Gamma_{ij}(\ell)] = E[(r_t - \mu)(r_{t-\ell} - \mu)'],$$

The $(i,j)$th element of $\Gamma_{\ell}$ is the covariance between $r_{it}$ and $r_{j,t-l}$

Lag-$l$ cross-correlation matrix:

$$\rho_{\ell} \equiv [\rho_{ij}(\ell)] = D^{-1}\Gamma_{\ell}D^{-1},$$

Lag-$l$ correlation coefficient between $r_{it}$ and $r_{i,t-\ell}$:

$$\rho_{ij}(\ell) = \frac{\Gamma_{ij}(\ell)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(r_{it}, r_{j,t-\ell})}{\text{std}(r_{it})\text{std}(r_{jt})},$$

$\rho_{ii}(\ell)$ is the lag-$l$ autocorrelation coefficient of $r_{it}$
Linear dependence of a weakly stationary:

1. The diagonal elements \( \{\rho_{ii}(\ell)|\ell = 0, 1, \ldots\} \) are the autocorrelation function of \( r_{it} \).

2. The off-diagonal element \( \rho_{ij}(0) \) measures the concurrent linear relationship between \( r_{it} \) and \( r_{jt} \).

3. For \( \ell > 0 \), the off-diagonal element \( \rho_{ij}(\ell) \) measures the linear dependence of \( r_{it} \) on the past value \( r_{j,t-\ell} \).

In general, we have,

1. \( r_{it} \) and \( r_{jt} \) have no linear relationship if \( \rho_{ij}(\ell) = \rho_{ji}(\ell) = 0 \) for all \( \ell \geq 0 \).

2. \( r_{it} \) and \( r_{jt} \) are concurrently correlated if \( \rho_{ij}(0) \neq 0 \).

3. \( r_{it} \) and \( r_{jt} \) have no lead–lag relationship if \( \rho_{ij}(\ell) = 0 \) and \( \rho_{ji}(\ell) = 0 \) for all \( \ell > 0 \). In this case, we say the two series are uncoupled.

4. There is a unidirectional relationship from \( r_{it} \) to \( r_{jt} \) if \( \rho_{ij}(\ell) = 0 \) for all \( \ell > 0 \), but \( \rho_{ji}(v) \neq 0 \) for some \( v > 0 \). In this case, \( r_{it} \) does not depend on any past value of \( r_{jt} \), but \( r_{jt} \) depends on some past values of \( r_{it} \).

5. There is a feedback relationship between \( r_{it} \) and \( r_{jt} \) if \( \rho_{ij}(\ell) \neq 0 \) for some \( \ell > 0 \) and \( \rho_{ji}(v) \neq 0 \) for some \( v > 0 \).
Sample cross-covariance matrix:

\[ \hat{\Gamma}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} (r_t - \bar{r})(r_{t-\ell} - \bar{r})', \quad \ell \geq 0, \]

where \( \bar{r} = (\sum_{t=1}^{T} r_t) / T \)

Sample cross-correlation matrix:

\[ \hat{\rho}_\ell = \hat{D}^{-1} \hat{\Gamma}_\ell \hat{D}^{-1}, \quad \ell \geq 0, \]

where, \( \hat{D} \) is a \( k \times k \) diagonal matrix of the sample standard deviations of the component series.

Fuller (1976): \( \hat{\rho}_\ell \) is consistent but is biased in a finite sample.
The finite sample distribution of $\hat{\rho}_\ell$ is rather complicated partly because of the presence of conditional heteroscedasticity and high kurtosis.

Tsay (2005) recommends using a proper bootstrap resampling method to obtain an approximate estimate of the distribution.

See examples 8.1 and 8.2 (Tsay, chapter 8, pp. 343-346)

Using simplifying notation of Tiao and Box (1981):

1. “+” means that the corresponding correlation coefficient is greater than or equal to $2/\sqrt{T}$,
2. “−” means that the corresponding correlation coefficient is less than or equal to $-2/\sqrt{T}$, and
3. “.” means that the corresponding correlation coefficient is between $-2/\sqrt{T}$ and $2/\sqrt{T}$,
Multivariate Portmanteau Tests:

Hosking (1980, 1981) and Li and McLeod (1981): The multivariate case of univariate Ljung-Box statistic $Q(m)$.

$H_0 : \rho_1 = \cdots = \rho_m = 0$

$H_a : \rho_i \neq 0$ for some $i \in \{1, \ldots, m\}$

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T - \ell} tr(\hat{\Gamma}' \hat{\Gamma}_0^{-1} \hat{\Gamma} \hat{\Gamma}_0^{-1}),$$

Where $tr(A)$ is the trace of the matrix $A$.

Under null hypothesis and some regularity conditions, $Q_k(m) \sim \chi^2_{k^2 m}$ (asymptotically).
Rewritten:

\[ Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T - \ell} b'_\ell (\hat{\rho}_0^{-1} \otimes \hat{\rho}_0^{-1}) b_\ell, \]

\[ b_\ell = \text{vec}(\hat{\rho}'_\ell) \]

**Note:** if it rejects the null hypothesis, then we build a multivariate model for the series to study the lead-lag relationships between the component series.

See application in pp. 348-349 (Tsay, 2005)
Vector autoregressive models - VAR

VAR(1):

\[ r_t = \phi_0 + \Phi r_{t-1} + a_t \]

where \( \phi_0 \) is a \( k \)-dimensional vector, \( \Phi \) is a \( k \times k \) matrix, and \( \{a_t\} \) is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix \( \Sigma \). In application, the covariance matrix \( \Sigma \) is required to be positive definite; otherwise, the dimension of \( r_t \) can be reduced. In the literature, it is often assumed that \( a_t \) is multivariate normal.

To simplifying, consider the bivariate case \((k = 2)\):

\[ r_{1t} = \phi_{10} + \Phi_{11} r_{1,t-1} + \Phi_{12} r_{2,t-1} + a_{1t}, \]
\[ r_{2t} = \phi_{20} + \Phi_{21} r_{1,t-1} + \Phi_{22} r_{2,t-1} + a_{2t}, \]

\( \Phi_{12} \) denotes the linear dependence of \( r_{1t} \) on \( r_{2,t-1} \)
If \( \Phi_{12} = 0 \), then \( r_{1t} \) does not depend on \( r_{2,t-1} \)
Φ_{12} = 0 and Φ_{21} ≠ 0 then there is a unidirectional relationship from r_{1t} to r_{2t}. Φ_{12} ≠ 0 and Φ_{21} ≠ 0, then there is a feedback relationship between the two series.

But, the reduced-form model doesn’t show explicitly the concurrent dependence between the component series.

Make a Cholesky decomposition: \( \Sigma = LGL' \), \( L \) a lower triangular matrix with unit diagonal elements, \( G \) a diagonal matrix.

\[ L^{-1}\Sigma (L')^{-1} = G \]

Define \( b_t = (b_{1t}, \ldots, b_{kt})' = L^{-1}a_t \). Then

\[ E(b_t) = L^{-1}E(a_t) = 0, \quad \text{Cov}(b_t) = L^{-1}\Sigma (L^{-1})' = L^{-1}\Sigma (L')^{-1} = G. \]

Since \( G \) is a diagonal matrix, the components of \( b_t \) are uncorrelated.
Structural form:

\[ \begin{align*}
L^{-1}r_t &= L^{-1}\phi_0 + L^{-1}\Phi r_{t-1} + L^{-1}a_t = \phi_0^* + \Phi^* r_{t-1} + b_t, \\
\phi_0^* &= L^{-1}\phi_0, \quad \Phi^* = L^{-1}\Phi 
\end{align*} \]

The kth equation of the model:

\[ r_{kt} + \sum_{i=1}^{k-1} w_{ki} r_{it} = \phi_{k,0}^* + \sum_{i=1}^{k} \Phi_{ki}^* r_{i,t-1} + b_{kt}, \]

Because \( b_{kt} \) is uncorrelated with \( b_{it} \), thus this equation shows explicitly the concurrent linear dependence of \( r_{kt} \) on \( r_{it} \).

See example 8.3, pp. 350 (Tsay, 2005)
Stationarity condition and moments of a VAR(1) model:

\[ \mu \equiv E(r_t) = (I - \Phi)^{-1} \phi_0 \]

Rewritten,

\[ \tilde{r}_t = \Phi \tilde{r}_{t-1} + a_t. \]

where \( \tilde{r}_t = r_t - \mu \)

and so,

\[ \tilde{r}_t = a_t + \Phi a_{t-1} + \Phi^2 a_{t-2} + \Phi^3 a_{t-3} + \cdots. \]

conclusions:

1) \( \text{Cov}(a_t, r_{t-1}) = 0 \)
2) \( \text{Cov}(r_t, a_t) = \Sigma \)
3) \( \Phi^j \) must converge to zero as \( j \to \infty \)

This means that the \( k \) eigenvalues of \( \Phi \) must be less than 1 in modulus

\( \Rightarrow \) necessary and sufficient conditions for weak stationarity of \( r_t \).
$|\lambda I - \Phi| = \lambda^k \left| I - \Phi \frac{1}{\lambda} \right|$

the eigenvalues of $\Phi$ are the inverses of the zeros of the determinant $|I - \Phi B|$. Thus, an equivalent sufficient and necessary condition for stationarity of $r_t$ is that all zeros of the determinant $|\Phi(B)|$ are greater than one in modulus; that is, all zeros are outside the unit circle in the complex plane. Fourth, using the expression, we have

$$\text{Cov}(r_t) = \Gamma_0 = \Sigma + \Phi \Sigma \Phi' + \Phi^2 \Sigma (\Phi^2)' + \cdots = \sum_{i=0}^{\infty} \Phi^i \Sigma (\Phi^i)'$$

Furthermore,

$$E(\tilde{r}_t \tilde{r}_t') = \Phi E(\tilde{r}_{t-1} \tilde{r}_{t-1})', \quad \ell > 0.$$ 

And so,

$$\Gamma_\ell = \Phi \Gamma_{\ell-1}, \quad \ell > 0,$$

$$\Gamma_\ell = \Phi^\ell \Gamma_0, \quad \text{for} \quad \ell > 0.$$
And the correlation,

\[ \rho_\ell = D^{-1/2} \Phi \Gamma_{\ell-1} D^{-1/2} = D^{-1/2} \Phi D^{1/2} D^{-1/2} \Gamma_{\ell-1} D^{-1/2} = \Upsilon \rho_{\ell-1}, \]

where \( \Upsilon = D^{-1/2} \Phi D^{1/2} \). Consequently, the CCM of a VAR(1) model satisfies

\[ \rho_\ell = \Upsilon^\ell \rho_0, \quad \text{for} \quad \ell > 0. \]
\textbf{VAR}(p):

\[ r_t = \phi_0 + \Phi_1 r_{t-1} + \cdots + \Phi_p r_{t-p} + a_t, \quad p > 0, \]

\(\Rightarrow\) \hspace{1cm}

\[(I - \Phi_1 B - \cdots - \Phi_p B^p)r_t = \phi_0 + a_t,\]

\(\Rightarrow\)

\[\Phi(B)r_t = \phi_0 + a_t.\]

\text{Mean:} \quad \mu = E(r_t) = (I - \Phi_1 - \cdots - \Phi_p)^{-1}\phi_0 = [\Phi(1)]^{-1}\phi_0

\text{Rewritten:} \quad \tilde{r}_t = \Phi_1 \tilde{r}_{t-1} + \cdots + \Phi_p \tilde{r}_{t-p} + a_t.

Similarly,

- \(\text{Cov}(r_t, a_t) = \Sigma\), the covariance matrix of \(a_t\);
- \(\text{Cov}(r_{t-\ell}, a_t) = 0\) for \(\ell > 0\);
- \(\Gamma_\ell = \Phi_1 \Gamma_{\ell-1} + \cdots + \Phi_p \Gamma_{\ell-p}\) for \(\ell > 0\).
Yule-Walker equation for a CCM (cross-correlation matrix):

$$\rho_\ell = \gamma_1 \rho_{\ell-1} + \cdots + \gamma_p \rho_{\ell-p} \quad \text{for} \quad \ell > 0,$$

where $\gamma_i = D^{-1/2} \Phi_i D^{1/2}$.

It’s possible to rewritten a VAR(p) as a VAR(1) and use the further conclusions.

Let $\tilde{x}_t = (\tilde{r}_{t-p+1}', \tilde{r}_{t-p+2}', \ldots, \tilde{r}_t)'$ and $b_t = (0, \ldots, 0, a_t)'$

$$x_t = \Phi^* x_{t-1} + b_t,$$

where $\Phi^*$ is a $kp \times kp$ matrix given by

$$\Phi^* = \begin{bmatrix}
0 & I & 0 & 0 & \cdots & 0 \\
0 & 0 & I & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \cdots & I \\
\Phi_p & \Phi_{p-1} & \Phi_{p-2} & \Phi_{p-3} & \cdots & \Phi_1
\end{bmatrix},$$

$\Phi^*$ is called the *companion* matrix of the matrix polynomial $\Phi(B)$. 
The results of a VAR(1) model can now be used to derive properties of the VAR(p) model.

example, from the definition, $x_t$ is weakly stationary if and only if $r_t$ is weakly stationary. Therefore, the necessary and sufficient condition of weak stationarity for the VAR($p$) model in Eq. (8.13) is that all eigenvalues of $\Phi^*$ in Eq. (8.15) are less than 1 in modulus. Similar to the VAR(1) case, it can be shown that the condition is equivalent to all zeros of the determinant $|\Phi(B)|$ being outside the unit circle.

structure of the coefficient matrices $\Phi_\ell$ thus provides information on the lead–lag relationship between the components of $r_t$. 
Estimation of a VAR(p) model:

We need to choose the order specification.

Consider the consecutive VAR models estimated by the ordinary least squares: VAR(1), VAR(2), …, VAR(i), …

\[ r_t = \phi_0 + \Phi_1 r_{t-1} + a_t \]

\[ r_t = \phi_0 + \Phi_1 r_{t-1} + \Phi_2 r_{t-2} + a_t \]

\[ \vdots \]

\[ r_t = \phi_0 + \Phi_1 r_{t-1} + \cdots + \Phi_i r_{t-i} + a_t \]

\[ \vdots \]

For a VAR(i), the residuals

\[ \hat{a}_t^{(i)} = r_t - \hat{\phi}_0^{(i)} - \hat{\Phi}_1^{(i)} r_{t-1} - \cdots - \hat{\Phi}_i^{(i)} r_{t-i} \]
Residual covariance matrix

\[ \hat{\Sigma}_i = \frac{1}{T - 2i - 1} \sum_{t=i+1}^{T} \hat{a}_t^{(i)} (\hat{a}_t^{(i)})', \quad i \geq 0 \]

To specify the order \( p \), one can test the null hypothesis \( H_o : \Phi_\ell = 0 \) versus \( H_a : \Phi_\ell \neq 0 \) sequentially for \( l = 1, 2, \ldots \) using the test statistic

\[ M(i) = -(T - k - i - \frac{3}{2}) \ln \left( \frac{|\hat{\Sigma}_i|}{|\hat{\Sigma}_{i-1}|} \right) \]

to testing a VAR(i) model versus a VAR(i-1) model.

is asymptotically a chi-squared distribution with \( k^2 \) degrees of freedom (see Tiao and Box (1981))

Alternatively, one can use an Akaike information criterion (AIC). In this case, we have to use the ML estimator of \( \Sigma \),

\[ \tilde{\Sigma}_i = \frac{1}{T} \sum_{t=i+1}^{T} \hat{a}_t^{(i)} [\hat{a}_t^{(i)}] ' \]
The AIC of a \( \text{VAR}(i) \) model under the normality assumption is defined as

\[
\text{AIC}(i) = \ln(|\tilde{\Sigma}_i|) + \frac{2k^2i}{T}.
\]

The order \( p \) will be such that \( \text{AIC}(p) = \min \text{AIC}(i) \)

Or using another information criterion,

\[
\text{BIC}(i) = \ln(|\tilde{\Sigma}_i|) + \frac{k^2i \ln(T)}{T},
\]

\[
\text{HQ}(i) = \ln(|\tilde{\Sigma}_i|) + \frac{2k^2i \ln(\ln(T))}{T}
\]

See example 8.4, pp. 356-358
Forecasting a VAR(p):

1-step ahead:
- forecast at the time origin \( h \): \( r_h(1) = \phi_0 + \sum_{i=1}^{p} \Phi_i r_{h+1-i} \)
- forecast error: \( e_h(1) = a_{h+1} \)
- covariance matrix of the forecast error: \( \Sigma \)

2-step ahead: substitute \( r_{h+1} \) by its forecast
- forecast at the time origin \( h \):
  \[
  r_h(2) = \phi_0 + \Phi_1 r_h(1) + \sum_{i=2}^{p} \Phi_i r_{h+2-i} 
  \]
- forecast error: \( e_h(2) = a_{h+2} + \Phi_1 [r_t - r_h(1)] = a_{h+2} + \Phi_1 a_{h+1} \)
- covariance matrix of the forecast error: \( \Sigma + \Phi_1 \Sigma \Phi_1' \)

If \( r_t \) is weakly stationary, then the \( l \)-step ahead forecast \( r_h(l) \) converges to its mean \( \mu \) as the horizon \( h \) increases and the covariance matrix of its forecast error converges to the covariance matrix of \( r_t \).
Impulse response function:

MA(∞) representation: \( r_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots \)

where \( \mu = [\Phi(1)]^{-1} \phi_0 \)

the coefficient matrix \( \Psi_i \) being the impact of the past innovation \( a_{t-i} \) on \( r_t \)

\( \Psi_i \) is often referred to as the impulse response function of \( r_t \).

However, the elements of \( a_t \) are often correlated. We need to use a Cholesky decomposition.
\[ r_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots \]
\[ = \mu + LL^{-1} a_t + \Psi_1 LL^{-1} a_{t-1} + \Psi_2 LL^{-1} a_{t-2} + \cdots \]
\[ = \mu + \Psi_0^* b_t + \Psi_1^* b_{t-1} + \Psi_2^* b_{t-2} + \cdots , \]

The coefficient matrices $\Psi_i^*$ are called the *impulse response function* of $r_t$ with the orthogonal innovations $b_t$. $b_t = L^{-1} a_t$

$\Psi_i^*(\ell)$ is the impact of $b_{j,t}$ on the future observation $r_{i,t+\ell}$.

**Problem:** the result depends on the ordering of the components of $r_t$.

**Example:**

\[ IBM_t = 1.16 + 0.02IBM_{t-1} + 0.11SP5_{t-1} + a_{1t} , \]
\[ SP5_t = 0.50 - 0.01IBM_{t-1} + 0.08SP5_{t-1} + a_{2t} . \]
Figure 8.5. Residual plots of fitting a VAR(1) model to the monthly log returns, in percentages, of IBM stock and the S&P 500 index. The sample period is from January 1926 to December 1999.

there exist clusters of outlying observations.
Figure 8.6. Forecasting plots of a fitted VAR(1) model to the monthly log returns, in percentages, of IBM stock and the S&P 500 index. The sample period is from January 1926 to December 1999.
Figure 8.7. Plots of impulse response functions of orthogonal innovations for a fitted VAR(1) model to the monthly log returns, in percentages, of IBM stock and the S&P 500 index. The sample period is from January 1926 to December 1999.
VECTOR MOVING AVERAGE MODELS - VMA(q)

\[ r_t = \theta_0 + a_t - \Theta_1 a_{t-1} - \cdots - \Theta_q a_{t-q} \quad \text{or} \quad r_t = \theta_0 + \Theta(B) a_t, \]

where \( \theta_0 \) is a \( k \)-dimensional vector, \( \Theta_i \) are \( k \times k \) matrices, and \( \Theta(B) = I - \Theta_1 B - \cdots - \Theta_q B^q \) is the MA matrix polynomial in the back-shift operator \( B \).

\[ \mu = E(r_t) = \theta_0. \]

Let \( \tilde{r}_t = r_t - \theta_0 \) be the mean-corrected VAR(q) process.

1. \( \text{Cov}(r_t, a_t) = \Sigma, \)
2. \( \Gamma_0 = \Sigma + \Theta_1 \Sigma \Theta'_1 + \cdots + \Theta_q \Sigma \Theta'_q, \)
3. \( \Gamma_\ell = 0 \) if \( \ell > q \), and
4. \( \Gamma_\ell = \sum_{j=\ell}^{q} \Theta_j \Sigma \Theta'_{j-\ell} \) if \( 1 \leq \ell \leq q \), where \( \Theta_0 = -I. \)

So, the cross-correlation matrix, \( \rho_\ell = 0, \quad \ell > q. \)
Application: VMA(1) $\Rightarrow$ bivariate MA(1) model

$$r_t = \theta_0 + a_t - \Theta a_{t-1} = \mu + a_t - \Theta a_{t-1}$$

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

We have a finite-memory model, because $r_t$ only depends on the current and past shocks.

The off-diagonal elements of $\Theta$ show the dynamic dependence between the component series.

Results:

1. They are uncoupled series if $\Theta_{12} = \Theta_{21} = 0$.
2. There is a unidirectional dynamic relationship from $r_{1t}$ to $r_{2t}$ if $\Theta_{12} = 0$, but $\Theta_{21} \neq 0$. The opposite unidirectional relationship holds if $\Theta_{21} = 0$, but $\Theta_{12} \neq 0$.
3. There is a feedback relationship between $r_{1t}$ and $r_{2t}$ if $\Theta_{12} \neq 0$ and $\Theta_{21} \neq 0$. 
Procedure:
1. use the sample cross-correlation matrices to specify the order $q$ for a VMA(q) model, $\rho_\ell = 0$ for $\ell > q$
2. estimate the specifies model by using either the conditional or exact likelihood method (the exact method is preferred when the sample size is not large)
3. the fitted model should be checked for adequacy applying the $Q_k(m)$ statistics to the residual series.
VECTOR ARMA MODELS - VARMA

Here, one of the issues is the *identifiability* problem. Examples:

VMA(1) = VAR(1):

\[
\begin{bmatrix}
  r_{1t} \\
  r_{2t}
\end{bmatrix}
= \begin{bmatrix}
  a_{1t} \\
  a_{2t}
\end{bmatrix}
- \begin{bmatrix}
  0 & 2 \\
  0 & 0
\end{bmatrix}
\begin{bmatrix}
  a_{1,t-1} \\
  a_{2,t-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  r_{1t} \\
  r_{2t}
\end{bmatrix}
- \begin{bmatrix}
  0 & -2 \\
  0 & 0
\end{bmatrix}
\begin{bmatrix}
  r_{1,t-1} \\
  r_{2,t-1}
\end{bmatrix}
= \begin{bmatrix}
  a_{1t} \\
  a_{2t}
\end{bmatrix}
\]

VARMA(1,1):

\[
\begin{bmatrix}
  r_{1t} \\
  r_{2t}
\end{bmatrix}
- \begin{bmatrix}
  0.8 & -2 \\
  0 & 0
\end{bmatrix}
\begin{bmatrix}
  r_{1,t-1} \\
  r_{2,t-1}
\end{bmatrix}
= \begin{bmatrix}
  a_{1t} \\
  a_{2t}
\end{bmatrix}
- \begin{bmatrix}
  -0.5 & 0 \\
  0 & 0
\end{bmatrix}
\begin{bmatrix}
  a_{1,t-1} \\
  a_{2,t-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  r_{1t} \\
  r_{2t}
\end{bmatrix}
- \begin{bmatrix}
  0.8 & -2 + \eta \\
  0 & \omega
\end{bmatrix}
\begin{bmatrix}
  r_{1,t-1} \\
  r_{2,t-1}
\end{bmatrix}
= \begin{bmatrix}
  a_{1t} \\
  a_{2t}
\end{bmatrix}
- \begin{bmatrix}
  -0.5 & \eta \\
  0 & \omega
\end{bmatrix}
\begin{bmatrix}
  a_{1,t-1} \\
  a_{2,t-1}
\end{bmatrix}
\]

They are identical.
This problem results in a situation similar to the exact multicollinearity in a regression analysis.


Therefore, VAR and VMA models are sufficient in most financial applications and when VARMA models are used, only lower order models are entertained.
UNIT ROOT NONSTATIONARITY AND COINTEGRATION

Consider the bivariate ARMA(1,1) model:

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

This is not a weakly stationary model because the two eigenvalues of the AR coefficient matrix are 0 and 1.

Rewriting,

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

Premultiplying by,

\[
\begin{bmatrix}
  1 - 0.5B & -B \\
  -0.25B & 1 - 0.5B
\end{bmatrix}
\]
We obtain,

\[
\begin{bmatrix}
1 - B & 0 \\
0 & 1 - B \\
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t} \\
\end{bmatrix}
=
\begin{bmatrix}
1 - 0.7B & -0.6B \\
-0.15B & 1 - 0.7B \\
\end{bmatrix}
\begin{bmatrix}
a_{1t} \\
a_{2t} \\
\end{bmatrix}
\]

So, each component is unit-root nonstationary and follows an ARIMA(0,1,1).

Considering the line transformation, L,

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

\[
\begin{bmatrix}
b_{1t} \\
b_{2t} \\
\end{bmatrix}
=
\begin{bmatrix}
1.0 & -2.0 \\
0.5 & 1.0 \\
\end{bmatrix}
\begin{bmatrix}
a_{1t} \\
a_{2t} \\
\end{bmatrix}
\equiv La_t.
\]
\[ Lx_t = L\Phi x_{t-1} + La_t - L\Theta a_{t-1} \]
\[ = L\Phi L^{-1}Lx_{t-1} + La_t - L\Theta L^{-1}La_{t-1} \]
\[ = L\Phi L^{-1}(Lx_{t-1}) + b_t - L\Theta L^{-1}b_{t-1}. \]

Thus, the model for \( y_t \) is

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

Results:

1. \( y_{1t} \) and \( y_{2t} \) are uncoupled series with concurrent correlation equal to that between \( b_{1t} \) and \( b_{2t} \).

2. \( y_{1t} \) follows ARIMA(0,1,1)

3. \( y_{2t} \) is a white noise
So, there is *only* a single unit root in the system, or, in other words, there is a *common trend* of $x_{1t}$ and $x_{2t} = \text{cointegration}$.  

$$y_{2t} = 0.5 x_{1t} + x_{2t}$$ does not have a unit root.

$x_{1t}$ and $x_{2t}$ is cointegrated if:

(a) both of them are unit-root nonstationary;

(b) they have a linear combination that is a unit-root stationary.

For a $k$-dimensional unit-root nonstationary time series, cointegration exists if there are less than $k$ unit roots in the system.

If there are $h$ unit root series, $0 < h < k$ and $k-h$ is the number of cointegration factors, i.e. the number of different linear combinations that are unit-root stationary, called cointegration vectors.

At the example: $y_{2t} = (0.5, 1) x_t$ and $(0.5, 1)'$ is a cointegrating vector.
The idea of cointegration is highly relevant in financial study: the prices must move in unison; otherwise there exists some arbitrage opportunity for investors.

**Error-correction form:**

Engle and Granger (1987) discuss an error-correction representation for a cointegrated system.

At the example, let \( \Delta x_t = x_t - x_{t-1} \)

\[
\begin{bmatrix}
\Delta x_{1t} \\
\Delta x_{2t}
\end{bmatrix} = 
\begin{bmatrix}
-0.5 & -1.0 \\
-0.25 & -0.5
\end{bmatrix}
\begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix} + 
\begin{bmatrix}
 a_{1t} \\
 a_{2t}
\end{bmatrix} - 
\begin{bmatrix}
 0.2 & -0.4 \\
-0.1 & 0.2
\end{bmatrix}
\begin{bmatrix}
a_{1,t-1} \\
a_{2,t-1}
\end{bmatrix}
\]

\[
= 
\begin{bmatrix}
-1 \\
-0.5
\end{bmatrix}
\begin{bmatrix}
0.5 & 1.0
\end{bmatrix}
\begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix} + 
\begin{bmatrix}
 a_{1t} \\
 a_{2t}
\end{bmatrix} - 
\begin{bmatrix}
 0.2 & -0.4 \\
-0.1 & 0.2
\end{bmatrix}
\begin{bmatrix}
a_{1,t-1} \\
a_{2,t-1}
\end{bmatrix}
\]

This is a stationary model because both \( \Delta x_t \) and \([0.5, 1.0]x_t = y_{2t}\) are unit-root stationary.
For a cointegrated VARMA(p,q) model with \( m \) cointegrating factors (\( m<k \)):

\[
\Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + a_t - \sum_{j=1}^{q} \Theta_j a_{t-j},
\]

where \( \alpha \) and \( \beta \) are \( k \times m \) full-rank matrices. The AR coefficient matrices \( \Phi_i^* \) are functions of the original coefficient matrices \( \Phi_j \). Specifically, we have

\[
\Phi_j^* = - \sum_{i=j+1}^{p} \Phi_i, \quad j = 1, \ldots, p - 1,
\]

\[
\alpha \beta' = \Phi_p + \Phi_{p-1} + \cdots + \Phi_1 - I = -\Phi(1).
\]

The time series \( \beta' x_t \) is unit-root stationary
the columns of \( \beta \) are the cointegration vectors of \( x_t \)

The model can be estimated by likelihood methods
COINTEGRATED VAR MODELS

\( k \)-dimensional VAR(p):

\[
x_t = \mu_t + \Phi_1 x_{t-1} + \cdots + \Phi_p x_{t-p} + a_t,
\]

\( a_t \) is assumed to be Gaussian and \( \mu_t = \mu_0 + \mu_1 t \)

Let \( \Phi(B) = I - \Phi_1 B - \cdots - \Phi_p B^p \) the autoregressive polynomial of the VAR.

Consider \( x_t \sim I(1) \), i.e. \( (1-B) x_t \sim I(0) \)

An error-correction model (ECM):

\[
\Delta x_t = \mu_t + \Pi x_{t-1} + \Phi_1^* \Delta x_{t-1} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + a_t,
\]

Where \( \Pi = \alpha \beta' = -\Phi(1) \) and \( \Pi x_{t-1} \) is the error-correction term.

3 cases:
1. Rank($\Pi$) = 0. This implies $\Pi = 0$ and $x_t$ is not cointegrated. The ECM of Eq. (8.36) reduces to

$$\Delta x_t = \mu_t + \Phi_1^* \Delta x_{t-1} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + a_t,$$

so that $\Delta x_t$ follows a VAR($p - 1$) model with deterministic trend $\mu_t$.

2. Rank($\Pi$) = $k$. This implies that $|\Phi(1)| \neq 0$ and $x_t$ contains no unit roots; that is, $x_t$ is $I(0)$. The ECM model is not informative and one studies $x_t$ directly.

3. $0 < \text{Rank}(\Pi) = m < k$. In this case, one can write $\Pi$ as

$$\Pi = \alpha \beta',$$  \hspace{1cm} (8.37)

where $\alpha$ and $\beta$ are $k \times m$ matrices with Rank($\alpha$) = Rank($\beta$) = $m$. The ECM of Eq. (8.36) becomes

$$\Delta x_t = \mu_t + \alpha \beta' x_{t-1} + \Phi_1^* \Delta x_{t-1} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + a_t.$$  \hspace{1cm} (8.38)

This means that $x_t$ is cointegrated with $m$ linearly independent cointegrating vectors, $w_t = \beta' x_t$, and has $k - m$ unit roots that give $k - m$ common stochastic trends of $x_t$. 
If \( x_t \) is cointegrated with \( \text{Rank}(\Pi) = m \), then a simple way to obtain a presentation of the \( k - m \) common trends is to obtain an orthogonal complement matrix \( \alpha_\perp \) of \( \alpha \); that is, \( \alpha_\perp \) is a \( k \times (k - m) \) matrix such that \( \alpha'_\perp \alpha = 0 \), a \( (k - m) \times m \) zero matrix, and use \( y_t = \alpha'_\perp x_t \). To see this, one can premultiply the ECM by \( \alpha'_\perp \) and use \( \Pi = \alpha \beta' \) to see that there would be no error-correction term in the resulting equation. Consequently, the \( (k - m) \)-dimensional series \( y_t \) should have \( k - m \) unit roots.

**Specifications**

1. \( \mu_t = \mathbf{0} \): In this case, all the component series of \( x_t \) are \( I(1) \) without drift and the stationary series \( w_t = \beta' x_t \) has mean zero.

2. \( \mu_t = \mu_0 = \alpha c_0 \), where \( c_0 \) is an \( m \)-dimensional nonzero constant vector. The ECM becomes

\[
\Delta x_t = \alpha (\beta' x_{t-1} + c_0) + \Phi_1^* \Delta x_{t-1} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + a_t,
\]
so that the components of $x_t$ are $I(1)$ without drift, but $w_t$ have a nonzero mean $-c_0$. This is referred to as the case of restricted constant.

3. $\mu_t = \mu_0$, which is nonzero. Here the component series of $x_t$ are $I(1)$ with drift $\mu_0$ and $w_t$ may have a nonzero mean.

4. $\mu_t = \mu_0 + \alpha c_1 t$, where $c_1$ is a nonzero vector. The ECM becomes

$$\Delta x_t = \mu_0 + \alpha (\beta' x_{t-1} + c_1 t) + \Phi_1^* \Delta x_{t-1} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + \alpha_t,$$

so that the components of $x_t$ are $I(1)$ with drift $\mu_0$ and $w_t$ has a linear time trend related to $c_1 t$. This is the case of restricted trend.

5. $\mu_t = \mu_0 + \mu_1 t$, where $\mu_1$ are nonzero. Here both the constant and trend are unrestricted. The components of $x_t$ are $I(1)$ and have a quadratic time trend and $w_t$ have a linear trend.
Cointegration Test:

Let $H(m)$ be the null hypothesis that the rank of $\mathbf{\Pi}$ is $m$.

$H(0) \subset \cdots \subset H(m) \subset \cdots \subset H(k)$.

Mathematically, the rank of $\mathbf{\Pi}$ is the number of nonzero eigenvalues of $\mathbf{\Pi}$.

$H_o : \text{Rank}(\mathbf{\Pi}) = m \quad \text{versus} \quad H_a : \text{Rank}(\mathbf{\Pi}) > m$.

Trace cointegration test:

Johansen (1988) proposes the likelihood ratio (LR) statistic

$$L K_{tr}(m) = -(T - p) \sum_{i=m+1}^{k} \ln(1 - \hat{\lambda}_i)$$

If $\text{Rank}(\mathbf{\Pi}) = m$, then $\hat{\lambda}_i$ should be small for $i > m$ and hence $L K_{tr}(m)$ should be small.
Due the presence of unit roots, the asymptotic distribution of $LK_\text{fr}(m)$ is not chi-squared, but a function of standard Brownian motions. Thus, critical values of $LK_\text{fr}(m)$ must be obtained via simulation.

Maximum eigenvalue test:

Johansen (1988) also considers a sequential procedure to determine the number of cointegrating vectors. Specifically, the hypotheses of interest are

$$H_0 : \text{Rank}(\Pi) = m \quad \text{versus} \quad H_a : \text{Rank}(\Pi) = m + 1.$$ 

The LK ratio test statistic, called the maximum eigenvalue statistic, is

$$LK_{\text{max}}(m) = -(T - p) \ln(1 - \hat{\lambda}_{m+1}).$$

Again, critical values of the test statistics are nonstandard and must be evaluated via simulation.

See example 8.6.5, pp.385.