Lecture 8
Vector Autoregressive Models

CHEN Ying
FE5209 Financial Econometrics

Outline

- Cross correlation matrix
- Vector Autoregressive (VAR) models
- Yule-Walker Equation of stable VAR models
- Forecasting with VAR models
- Model identification

Readings
SDA chapter 10.3
MTS chapter 1
FTS chapter 8

MTS: Multivariate Time Series Analysis with R and Financial Applications by Ruey Tsay

Why consider multiple series jointly?

Examples of multiple time series:

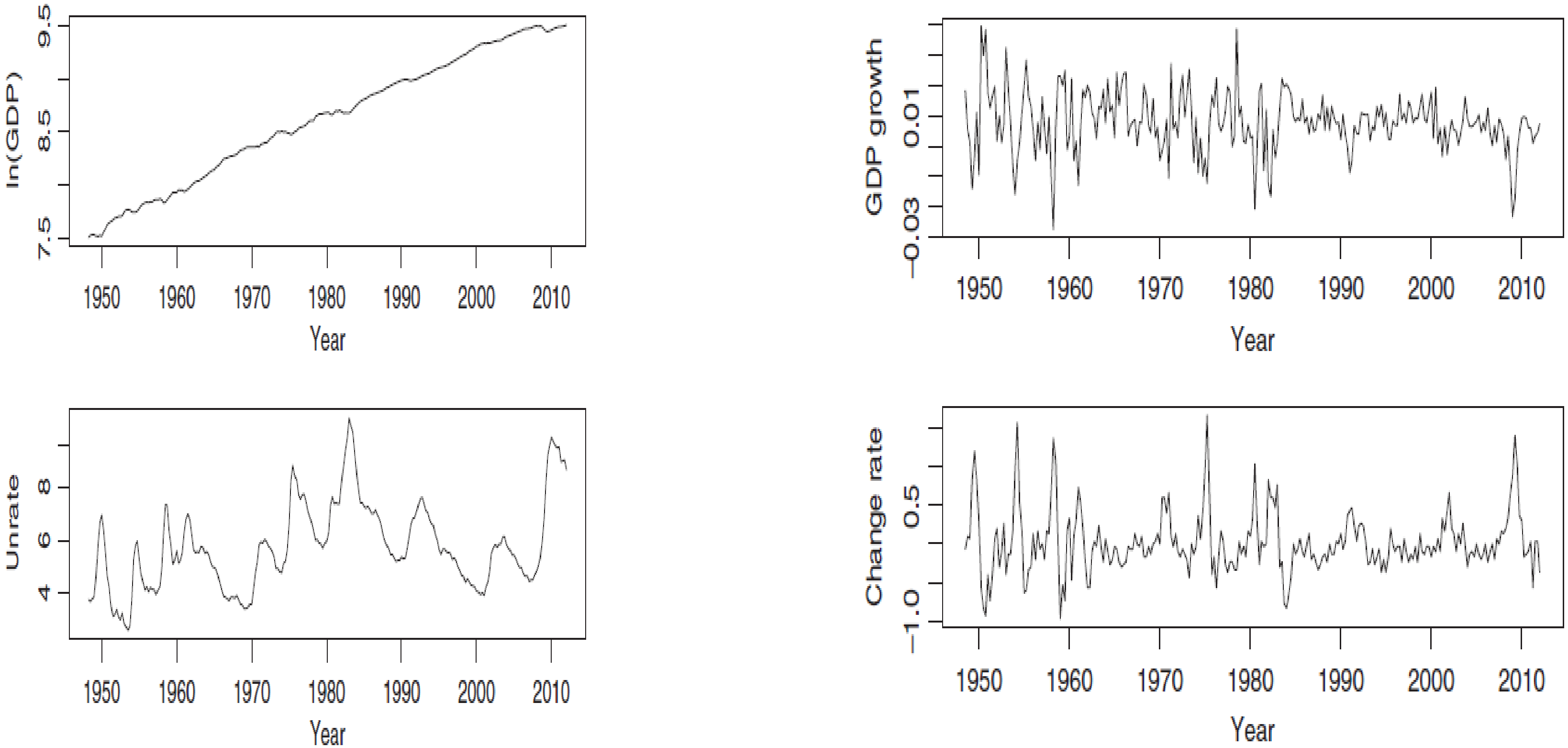
- ❑ U.S. quarterly GDP and unemployment rate series;
- ❑ Monthly log returns of IBM stock and S&P 500 index;
- ❑ The daily closing prices of oil related ETFs, e.g. oil services holdings (OIH) and energy select section SPDR (XLE);
- ❑ Quarterly GDP growth rates of Canada, United Kingdom, and United States.

Why consider multiple series jointly?

(a) To study the dynamic relationship between the series

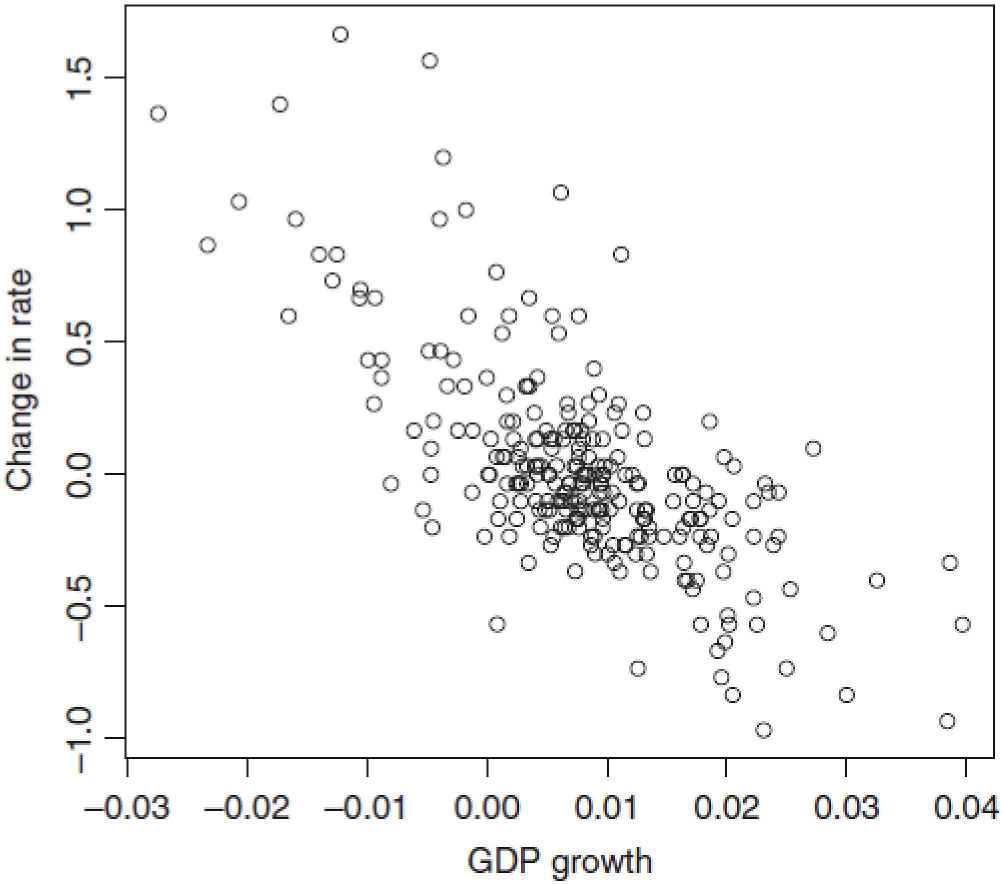
(b) Improve the accuracy of forecasts (use more information).

GDP and unemployment rate



Left: Time plots of U.S. quarterly real GDP (in logarithm) and unemployment rate from 1948 to 2011.
Right: Time plots of the growth rate of U.S. quarterly real GDP (in logarithm) and the change series of unemployment rate from 1948 to 2011. The data are seasonally adjusted. 8_bivariate.R

GDP and unemployment rate



Scatter plot of the changes in quarterly U.S. unemployment rate versus the growth rate of quarterly real GDP (in logarithm) from the second quarter of 1948 to the last quarter of 2011. The data are seasonally adjusted.
8_bivariate.R

Unemployment rates

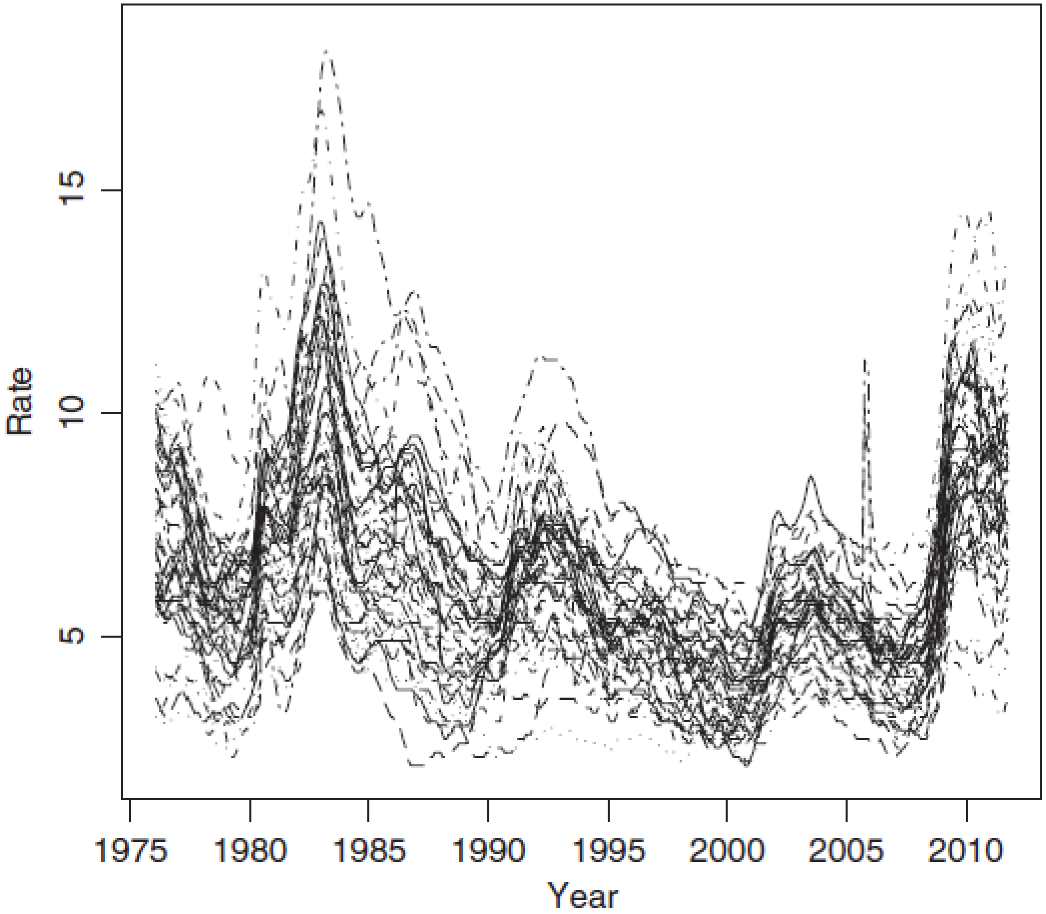


Figure: Time plots of the monthly unemployment rates of the 50 states in the United States from January 1976 to September 2011. The data are seasonally adjusted.
8_bivariate.R

IBM and S&P 500 Index

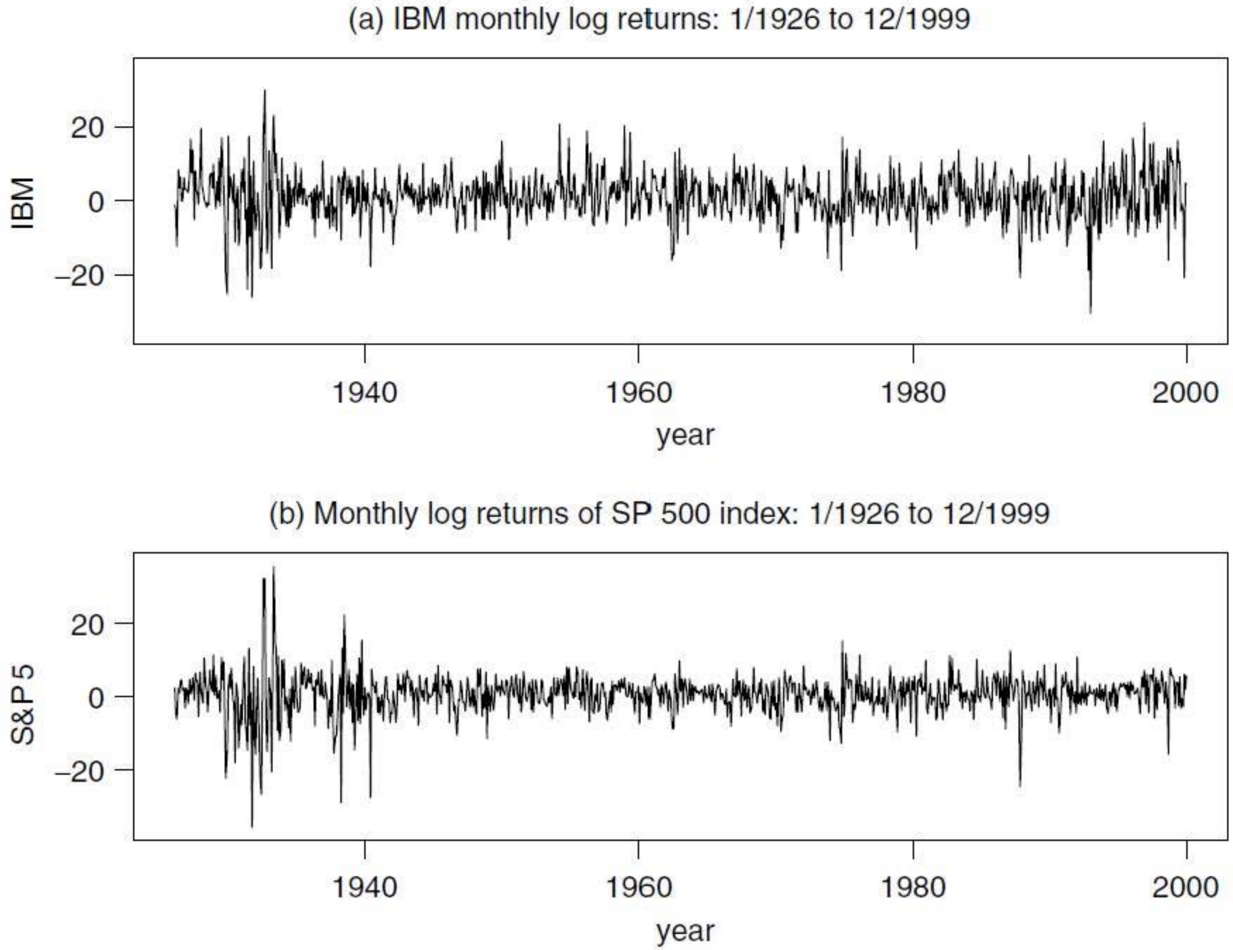


Figure: Time plots of (a) monthly log returns in percentages for IBM stock and (b) the S&P 500 index from January 1926 to December 1999.

IBM and S&P 500 Index

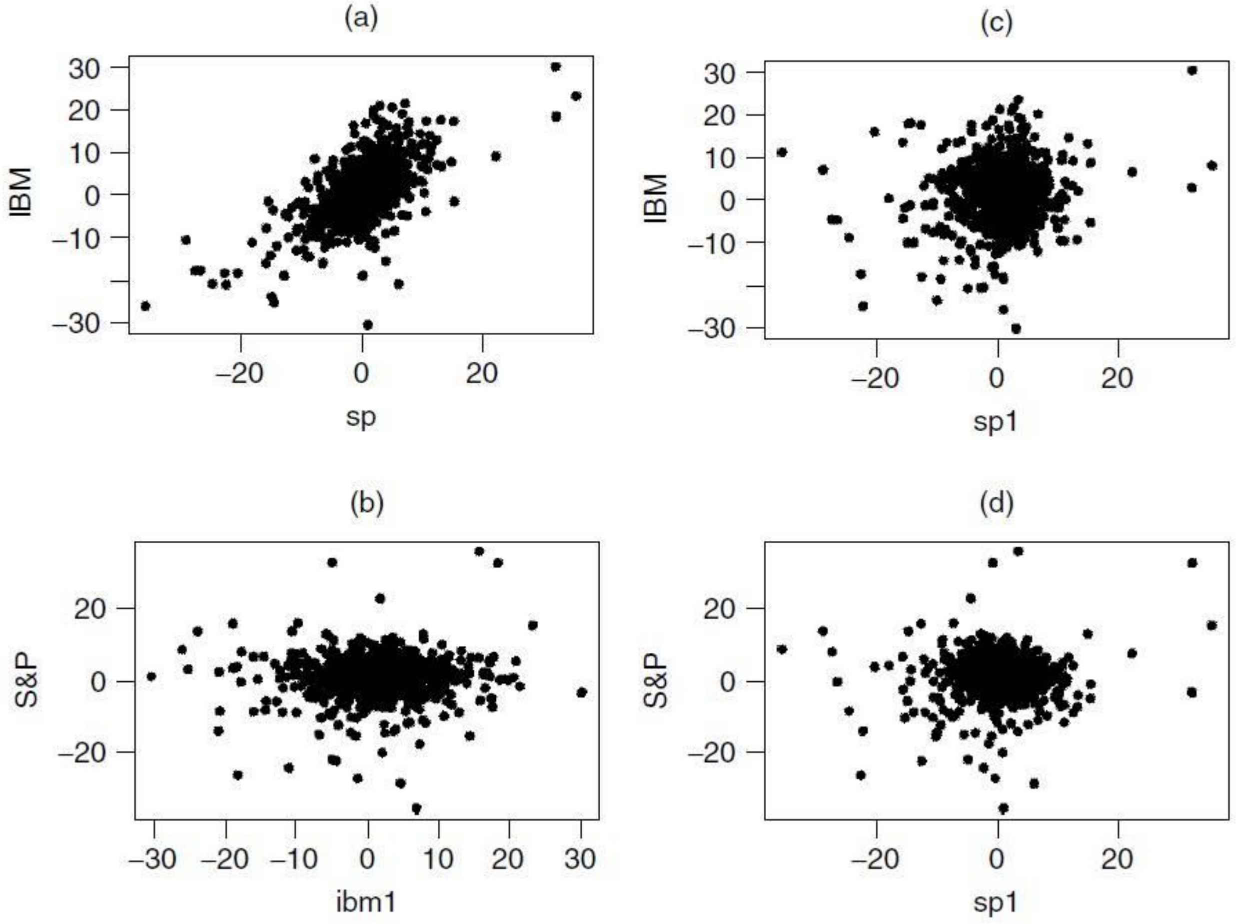


Figure: Scatter plots for monthly log returns of IBM stock and the S&P 500 index: (a) concurrent plot of IBM versus S&P 500 (b) S&P 500 versus lag-1 IBM, (c) IBM versus lag-1 S&P 500, and (d) S&P 500 versus lag-1 S&P 500.

Stationarity

- ❑ Strict stationarity: distributions are time-invariant
- ❑ Weak stationarity: the first two moments are time-invariant

Weak stationarity of $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots \boldsymbol{Y}_T$:

$$\boldsymbol{\mu}_{\boldsymbol{Y}} = E(\boldsymbol{Y}_t)$$

$$\boldsymbol{\Gamma}_k = Cov(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-k}) = E[(\boldsymbol{Y}_t - \boldsymbol{\mu}_{\boldsymbol{Y}})(\boldsymbol{Y}_{t-k} - \boldsymbol{\mu}_{\boldsymbol{Y}})']$$

are time-invariant.

In particular, $\boldsymbol{\Gamma}_0$ is the covariance matrix of $\boldsymbol{Y}_t$.

Cross-covariance matrix

Consider two series $\boldsymbol{Y}_t = \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix}$

$$\boldsymbol{\mu_Y} = E(\boldsymbol{Y}_t) = \begin{bmatrix} E(Y_{1t}) \\ E(Y_{2t}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

For lag k

$$\begin{aligned}\boldsymbol{\Gamma}_k &= E[(\boldsymbol{Y}_t - \boldsymbol{\mu_Y})(\boldsymbol{Y}_{t-k} - \boldsymbol{\mu_Y})'] \\ &= \begin{bmatrix} E(Y_{1t} - \mu_1)(Y_{1,t-k} - \mu_1) & E(Y_{1t} - \mu_1)(Y_{2,t-k} - \mu_2) \\ E(Y_{2t} - \mu_2)(Y_{1,t-k} - \mu_1) & E(Y_{2t} - \mu_2)(Y_{2,t-k} - \mu_2) \end{bmatrix} \\ &= \begin{bmatrix} \Gamma_{11}(k) & \Gamma_{12}(k) \\ \Gamma_{21}(k) & \Gamma_{22}(k) \end{bmatrix} = \begin{bmatrix} \gamma_k \; of \; Y_1 & \Gamma_{12}(k) \\ \Gamma_{21}(k) & \gamma_k \; of \; Y_2 \end{bmatrix}\end{aligned}$$

Asymmetric if $k \neq 0$. Consider $\boldsymbol{\Gamma}_1$:

$$\Gamma_{12}(1){:}\; Cov(Y_{1t}, Y_{2,t-1})$$
$$\Gamma_{21}(1){:}\; Cov(Y_{2t}, Y_{1,t-1})$$

We have: $\boldsymbol{\Gamma}_{-k} \neq \boldsymbol{\Gamma}_k \;\; but \; \boldsymbol{\Gamma}_{-k} = \boldsymbol{\Gamma}'_k$ under stationarity.

$$\Gamma_{12}(-k) = cov(Y_{1t}, Y_{2,t+k}) = cov(Y_{2,t+k}, Y_{1t}) = Cov(Y_{2t}, Y_{1,t-k}) = \Gamma_{21}(k).$$

Cross-Correlation Matrix (CCM)

Let the diagonal matrix $\boldsymbol{D}$ be

$$\boldsymbol{D} = \begin{bmatrix} std(Y_{1t}) & 0 \\ 0 & std(Y_{2t}) \end{bmatrix} = \begin{bmatrix} \sqrt{\Gamma_{11}(0)} & 0 \\ 0 & \sqrt{\Gamma_{22}(0)} \end{bmatrix}$$

For a stationary series $\boldsymbol{Y}_t$, the cross-correlation matrix $\boldsymbol{\rho}_k$ is defined as:

$$\boldsymbol{\rho}_0 = \left[\rho_{ij}(0)\right] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_0\boldsymbol{D}^{-1}$$

$$\rho_{ij}(0) = \frac{\gamma_{ij}(0)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}} = \frac{Cov(Y_{it}, Y_{jt})}{std(Y_{it})std(Y_{jt})}$$

$$\boldsymbol{\rho}_k = \left[\rho_{ij}(k)\right] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_k\boldsymbol{D}^{-1}$$

$$\rho_{ij}(k) = \frac{\gamma_{ij}(k)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}} = \frac{Cov(Y_{it}, Y_{j,t-k})}{std(Y_{it})std(Y_{j,t-k})}$$

Thus, $\rho_{ij}(k)$ is the cross-correlation between Y_{it} and $Y_{j,t-k}$.

Obviously, $\boldsymbol{\rho}_0$ is symmetric, but $\boldsymbol{\rho}_k$ is in general not symmetric for $\mathrm{k} \neq 0$.

From stationarity: $\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}'_{-k}$, $\boldsymbol{\rho}_k = \boldsymbol{\rho}'_{-k}$

For instance: $cor\left(Y_{1t}, Y_{2,t-1}\right) = cor\left(Y_{1,t+1}, Y_{2t}\right) = cor\left(Y_{2t}, Y_{1,t+1}\right)$.

CCM

Cross-correlation matrix $\{\boldsymbol{\rho}_k | k = 0,1,\ldots\}$ of a weakly stationary vector time series:

- The diagonal elements $\{\rho_{ii}(k) | k = 0,1,\ldots\}$ are the autocorrelation function of Y_{it}.
- The off-diagonal element $\rho_{ij}(\mathbf{0})$ measures the concurrent linear relationship between Y_{it} and Y_{jt}.
- For $k > 0$, the off-diagonal element $\rho_{ij}(k)$ measures the linear dependence of Y_{it} on the past value $Y_{j,t-k}$. Therefore, if $\rho_{ij}(k) = 0$ for all $k > 0$, then Y_{it} does not depend linearly on any past value $Y_{j,t-k}$ of the Y_{jt} series.

$$\begin{bmatrix} \rho_k \; of \; Y_1 & \rho_{12}(k) & \cdots & \rho_{1d}(k) \\ \rho_{21}(k) & \rho_k \; of \; Y_2 & & \rho_{2d}(k) \\ \vdots & & \ddots & \vdots \\ \rho_{d1}(k) & \rho_{d2}(k) & \cdots & \rho_k \; of \; Y_d \end{bmatrix}$$

CCM

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

Two series are **uncoupled**: Y_{it} and Y_{jt} have no lead-lag relationship if $\rho_{ij}(k) = 0$ and $\rho_{ji}(k) = 0$ for all $k > 0$.

There is a **unidirectional** relationship from Y_{it} and Y_{jt} if $\rho_{ij}(k) = 0$ for all $k > 0$, but $\rho_{ji}(v) \neq 0$ for some $v > 0$.
$\Rightarrow$ Y_{it} does not depend on any past value of Y_{jt}, but Y_{jt} depends on some past values of Y_{it}.

There is a **feedback** relationship between Y_{it} and Y_{jt} if $\rho_{ij}(k) \neq 0$ for some $k > 0$ and $\rho_{ji}(v) \neq 0$ for some $v > 0$.

Sample cross-correlation matrix

Given the sample $\{\boldsymbol{y}_t\}_{t=1}^{n}$, we construct the sample mean and covariance matrix as

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{y}_t, \qquad \widehat{\boldsymbol{\Gamma}}_0 = \frac{1}{n-1}\sum_{t=1}^{n} (\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}})(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}})'$$

These sample quantities are estimates of $\boldsymbol{\mu}_{\boldsymbol{Y}}$ and $\boldsymbol{\Gamma}_0$, respectively.

The lag-k sample cross-covariance matrix is defined as

$$\widehat{\boldsymbol{\Gamma}}_k = \frac{1}{n-1}\sum_{t=k+1}^{n} (\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}})(\boldsymbol{y}_{t-k} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}})'$$

The lag-k sample cross-correlation matrix (CCM) is then

$$\widehat{\boldsymbol{\rho}}_k = \widehat{\boldsymbol{D}}^{-1}\widehat{\boldsymbol{\Gamma}}_k\widehat{\boldsymbol{D}}^{-1}$$

where $\widehat{\boldsymbol{D}} = \mathrm{diag}\{\widehat{\Gamma}_{0,11}^{\frac{1}{2}}, \cdots, \widehat{\Gamma}_{0,dd}^{\frac{1}{2}}\}$, where $\widehat{\Gamma}_{0,ii}$ is the (i,i)th element of $\widehat{\boldsymbol{\Gamma}}_0$.

Multivariate White Noise

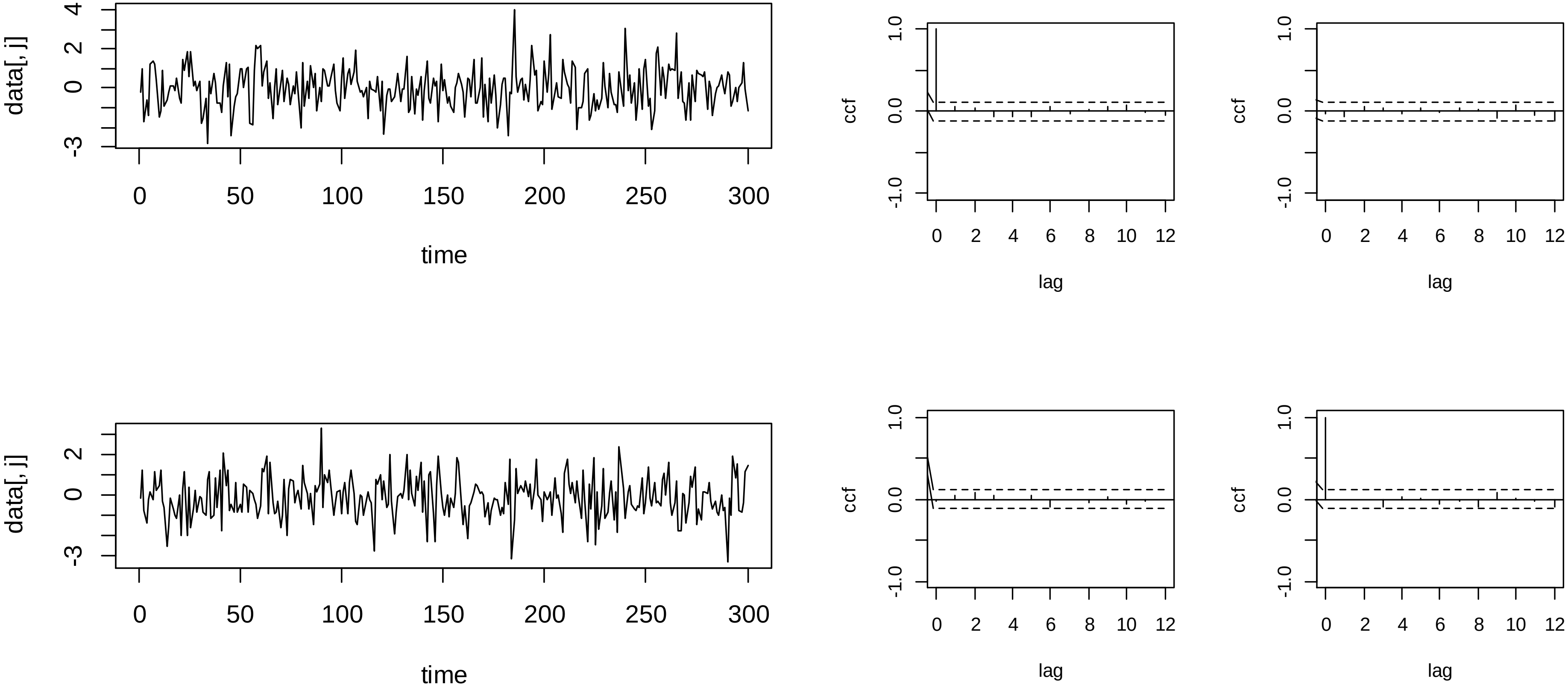
A ***d*-dimensional** multivariate time series $\boldsymbol{\epsilon_1}, \boldsymbol{\epsilon_2}, \ldots$ is a weak $WN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ process if

1. $E(\boldsymbol{\epsilon_t}) = 0$ for all t,
2. $COV(\boldsymbol{\epsilon_t}) = \boldsymbol{\Sigma}$ for all t, and
3. for all $t \neq t'$, all components of $\boldsymbol{\epsilon_t}$ are uncorrelated with all components of $\boldsymbol{\epsilon_t}'$.

Notice that if $\boldsymbol{\Sigma}$ is not diagonal, then there is concurrent cross-correlation between the components of $\boldsymbol{\epsilon_t}$ because $Corr(\epsilon_{j,t}\epsilon_{j',t}) = \Sigma_{j,j'}$; in other words, there may be nonzero *contemporaneous* correlations. However, for all $1 \leq j, j' \leq d$,

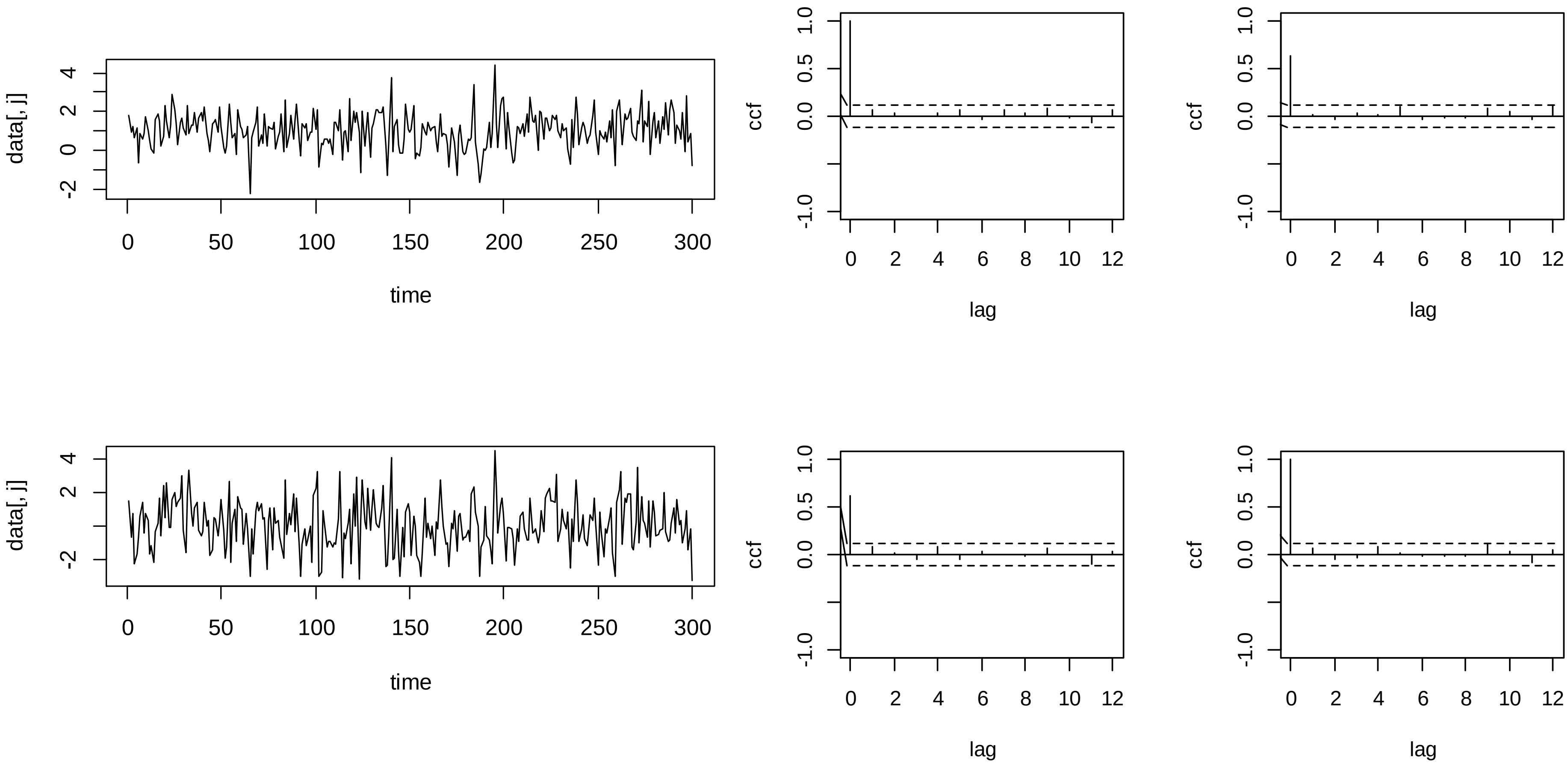
$Corr(\epsilon_{j,t}\epsilon_{j',t'}) = 0$ if $t \neq t'$.

Example: White Noise



Simulated white noise: N(**0,I**$_2$) and the Cross-Correlation Matrix.
8_whitenoise_CCM.R

Example: White Noise



Simulated white noise: $N\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}\mathbf{1} & \mathbf{.8}\\ \mathbf{.8} & \mathbf{2}\end{pmatrix}\right)$ and the Cross-Correlation Matrix (CCM). 8_whitenoise_CCM.R

Linearity

$\boldsymbol{Y}_t$ is a linear function of independent and identically distributed (IID) random vectors, i.e.

$$\begin{pmatrix} Y_{1t} \\ \vdots \\ Y_{dt} \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} + \begin{pmatrix} \sum_{i=0}^{\infty} \varphi_{1i} \varepsilon_{1,t-i} \\ \vdots \\ \sum_{i=0}^{\infty} \varphi_{di} \varepsilon_{d,t-i} \end{pmatrix}$$

Matrix form:

$$\boldsymbol{Y}_t = \boldsymbol{c} + \sum_{i=0}^{\infty} \boldsymbol{\varphi}_i \boldsymbol{\varepsilon}_{t-i}$$

where $\mathbf{c}$ is a constant vector, $\boldsymbol{\varphi}_0 = \boldsymbol{I}$, $\{\boldsymbol{\varepsilon}_j\}$ is a sequence of IID random vectors with mean zero and positive-definite covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$.

Multivariate Portmanteau test

If $\boldsymbol{Y}_t$ is a stationary linear process and $\boldsymbol{\varepsilon}_t$ is normally distributed, then $\widehat{\boldsymbol{\rho}}_k$ is a consistent estimate of $\boldsymbol{\rho}_k$. The asymptotic covariance of elements of $\widehat{\boldsymbol{\rho}}_k$ is complicated in general. See Chapter 10 of Box, Jenkins and Reinsel (1994).

In practice, we are often interested in testing the hypothesis

$$H_0\colon \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \boldsymbol{0} \text{ vs } H_a\colon \boldsymbol{\rho}_i \neq \boldsymbol{0} \text{ for some } 1 \leq i \leq m,$$

where m is a positive integer.

Multivariate version of Ljung-Box $Q(m)$ statistics

$$Q_d(m) = n^2 \sum_{k=1}^{m} \frac{1}{n-k} tr(\widehat{\boldsymbol{\Gamma}}_k' \widehat{\boldsymbol{\Gamma}}_0^{-1} \widehat{\boldsymbol{\Gamma}}_k \widehat{\boldsymbol{\Gamma}}_0^{-1}) \xrightarrow{\mathcal{L}} \chi^2_{md^2}$$

where n is the sample size, d is the dimension of $\boldsymbol{y}_t$, and $tr(\boldsymbol{A})$ is the trace of the matrix $\boldsymbol{A}$. This is also called the *multivariate Portmanteau test*.

Remark: For a fitted VAR(p) model, the $Q_d(m)$ statistic of the residuals is asymptotically $\chi^2_{md^2-g}$ distributed, where g is the no. of estimated parameters in the AR coefficient matrices.

See Hosking (1980, JASA and 1981, JRSSB) and Li and McLeod (1981, JRSSB).

Example: IBM and S&P 500 Index

(a) Summary Statistics

Ticker	Mean	Standard Error	Skewness	Excess Kurtosis	Minimum	Maximum
IBM	1.240	6.729	−0.237	1.917	−30.37	30.10
S&P 500	0.537	5.645	−0.521	8.117	−35.58	35.22

(b) Cross-Correlation Matrices

Lag 1		Lag 2		Lag 3		Lag 4		Lag 5	
0.08	0.10	0.02	−0.06	−0.02	−0.07	−0.02	−0.03	0.00	0.07
0.04	0.08	0.02	−0.02	−0.07	−0.11	0.04	0.02	0.00	0.08

(c) Simplified Notation

$$\begin{bmatrix} + & + \\ \cdot & + \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & - \\ - & - \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & + \end{bmatrix}$$

Table: Summary Statistics and Cross-Correlation Matrices of Monthly Log Returns of IBM Stock and the S&P 500 Index: January 1926 to December 1999. (+: CCF>$2/\sqrt{T}$, −: CCF<$-2/\sqrt{T}$ and $\cdot$: between.)

Example: IBM and S&P 500 Index

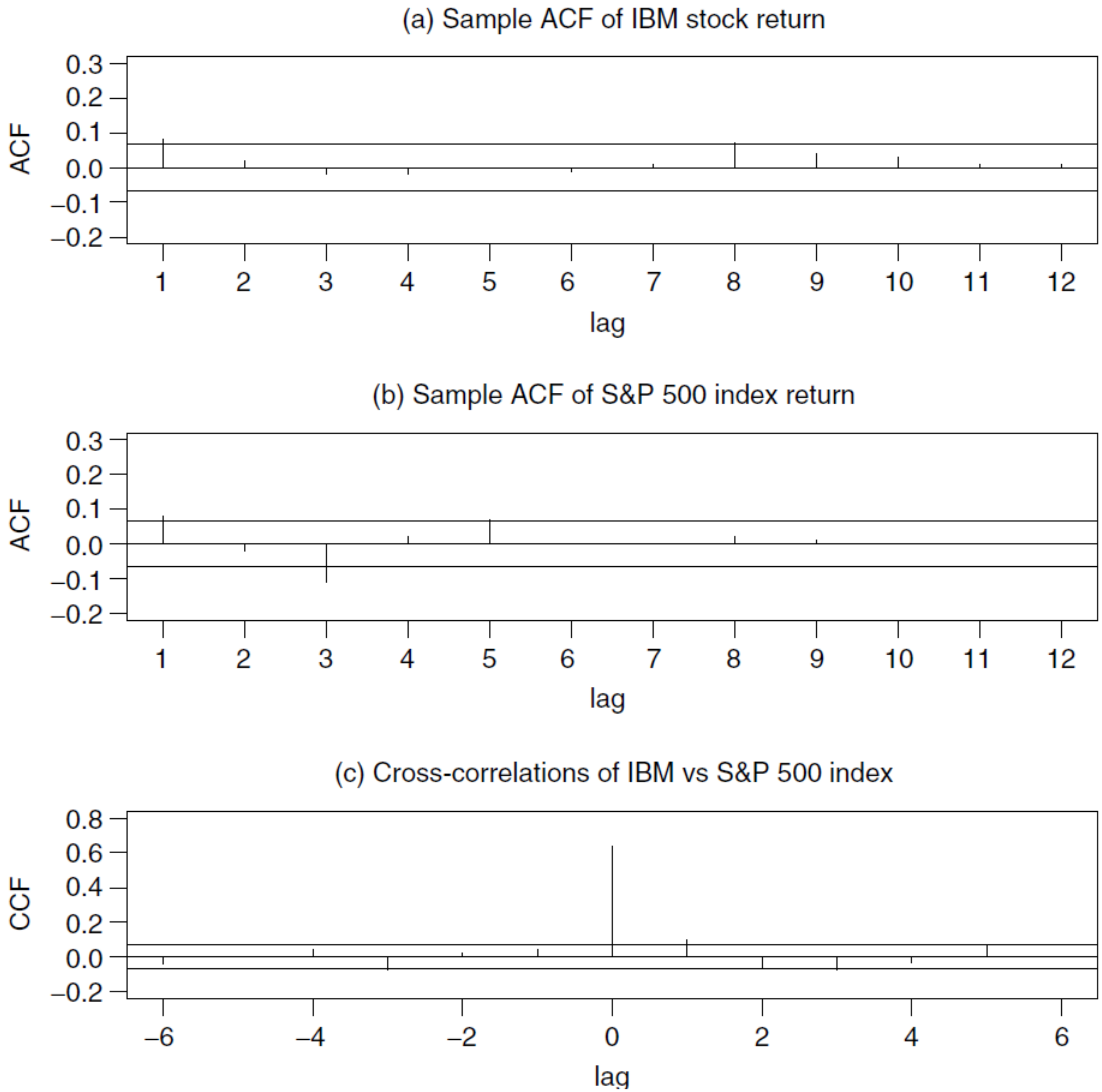
The results show the CCM for the monthly log returns of IBM stock and the S&P 500 index.

Significant cross-correlations at the approximate 5% level appear mainly at lags 1 and 3.

S&P 500 index returns have some marginal autocorrelations at lags 1 and 3.

IBM stock returns depend weakly on the previous returns of the S&P 500 index.

Example: IBM and S&P 500 Index



Sample auto- and cross-correlation functions of two monthly log returns: (a) sample ACF of IBM stock returns, (b) sample ACF of S&P 500 index returns, and (c) cross-correlations between IBM stock return and lagged S&P 500 index returns.

Example: IBM and S&P 500 Index

Applying the $Q_d(m)$ statistics to the bivariate monthly log returns of IBM stock and the S&P 500 index, we have

$$Q_2(1) = 9.81, Q_2(5) = 47.06, Q_2(10) = 71.65.$$

Based on asymptotic chi-squared distributions with degrees of freedom $4(= md^2 = 1 \times 2^2)$, 20, and 40, the p-values of these $Q_2(m)$ statistics are 0.044, 0.001, and 0.002, respectively.

The portmanteau tests thus confirm the existence of serial dependence in the bivariate return series at the 5% significance level.

Vector Autoregressive Models (VAR)

Given $\boldsymbol{Y}_t \in R^d$, the VAR(1) model is:

$$\boldsymbol{Y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t, \qquad \text{or} \qquad (\boldsymbol{I} - \boldsymbol{\phi}_1 \boldsymbol{B})\boldsymbol{Y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{I}$ denotes the $d \times d$ identity matrix.

VAR(1) model for 2-dimension return series:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

matrix form

$$\boldsymbol{Y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\phi}_0$ is a d-dimensional vector, $\boldsymbol{\phi}_1$ is a $d \times d$ matrix, and $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$

$$cov(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

where $\sigma_{12} = \sigma_{21}$.

VAR(1)

The VAR(1) model considers the interdependence between the two time series. Such a model is called a **dynamic simultaneous equations** model.

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

or equivalently,

$$Y_{1t} = \phi_{10} + \phi_{11} Y_{1,t-1} + \phi_{12} Y_{2,t-1} + \varepsilon_{1t}$$
$$Y_{2t} = \phi_{20} + \phi_{21} Y_{1,t-1} + \phi_{22} Y_{2,t-1} + \varepsilon_{2t}$$

where ϕ_{11} and ϕ_{12} denote the dependence of Y_{1t} on the past returns $Y_{1,t-1}$ and $Y_{2,t-1}$, respectively.

Granger causality:

Uncoupled TS: $\phi_{12} = \phi_{21} = 0$

Unidirectional relation: $\phi_{12} = 0, \phi_{21} \neq 0$

Feedback relation: $\phi_{12} \neq 0, \phi_{21} \neq 0$

Stationarity of VAR(1)

The stationarity condition of a VAR(1) model is that all eigenvalues of $\boldsymbol{\phi}_1$ are less than one in modulus.

To see this, we may assume that $\boldsymbol{\phi}_0 = 0$ and the time series starts at $t = 0$ with initial value $\boldsymbol{Y}_0$. Then, it is easy to see that

$$\boldsymbol{Y}_t = \sum_{i=0}^{t-1} \boldsymbol{\phi}_1^i \, \varepsilon_{t-i} + \boldsymbol{\phi}_1^t \boldsymbol{Y}_0$$

If $\boldsymbol{Y}_t$ is stationary, then $\boldsymbol{Y}_t$ does not depend on $\boldsymbol{Y}_0$ as $t \to \infty$. Consequently, $\boldsymbol{\phi}_1^i \to \mathbf{0}$ as $i \to \infty$. Thus, all eigenvalues of $\boldsymbol{\phi}_1$ should be less than one in absolute value.

Equivalently, $\{\boldsymbol{Y}_t\}$ is stationary if roots of the polynomial

$$|\boldsymbol{I} - \boldsymbol{\phi}_1 \boldsymbol{z}|$$

are greater than 1 in modulus. (A generalization of univariate case).

Mean of $\boldsymbol{Y}_t$ satisfies

$$(\boldsymbol{I} - \boldsymbol{\phi}_1)\boldsymbol{\mu} = \boldsymbol{\phi}_0, \text{or } \ \boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{\phi}_1)^{-1}\boldsymbol{\phi}_0$$

if the inverse exists.

Stationarity

$$\begin{bmatrix} 1-\phi_{11}B & -\phi_{12}B \\ -\phi_{21}B & 1-\phi_{22}B \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

This gives the solution

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 1-\phi_{11}B & -\phi_{12}B \\ -\phi_{21}B & 1-\phi_{22}B \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$
$$= \frac{1}{\Delta} \begin{bmatrix} 1-\phi_{22}B & \phi_{21}B \\ \phi_{12}B & 1-\phi_{11}B \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where

$$\Delta = (1-\phi_{11}B)(1-\phi_{22}B) - (\phi_{12}B)(\phi_{21}B)$$
$$= 1 - (\phi_{11}+\phi_{22})B + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21})B^2$$

Let λ_1 and λ_2 are the roots of the equation

$$1 - (\phi_{11}+\phi_{22})\mathrm{z} + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21})z^2 = 0$$

In order that we have a convergent expansion for Y_{1t} and Y_{2t} in terms of ε_{1t} and ε_{2t} we should have $|\lambda_1| > 1$ and $|\lambda_2| > 1$.

Invertibility

A time series $\boldsymbol{Y}_t$ is said to be invertible if

$$\boldsymbol{Y}_t = \varepsilon_t + \sum_{j=1}^{\infty} \boldsymbol{\pi}_j \boldsymbol{Y}_{t-j}$$

That is, $\boldsymbol{Y}_t$ can be written as a linear combination of its past values and ε_t.

Similarly to weak stationarity, for an invertible series $\boldsymbol{Y}_t$, $\boldsymbol{\pi}_j \rightarrow \mathbf{0}$ as $j \rightarrow \infty$.

The VAR model is invertible.

Invertibility condition for the MA(1) model

$Y_t = \beta\epsilon_{t-1} + \epsilon_t,$ $\quad \rho_1 = \frac{\beta}{1+\beta^2}$

$Y_t = \frac{1}{\beta}\epsilon_{t-1} + \epsilon_t,$ $\quad \rho_1 = \frac{\frac{1}{\beta}}{1+\frac{1}{\beta^2}} = \frac{\beta}{1+\beta^2}$

$|\beta| < 1$, so that the most recent observations have higher weight than the observations from the more distant past

YW equation of VAR(1)

Under stationarity.

Let $\boldsymbol{z_t} = \boldsymbol{Y}_t - \boldsymbol{\mu}$, where $\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{Y}_t)$. The VAR(1) model can be written as

$$\boldsymbol{z_t} = \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{\varepsilon}_{t.}$$

Post-multiplying the model by $\boldsymbol{z}'_{t-l}$ and taking expectation, we obtain the moment equation

$$\boldsymbol{\Gamma}_l = \boldsymbol{\phi}_1 \, \boldsymbol{\Gamma}_{l-1}, \qquad l > 0.$$

Consequently, for a stationary VAR(1) model, we have

$$\boldsymbol{\Gamma}_l = \boldsymbol{\phi}_1^l \boldsymbol{\Gamma}_0 \text{ for } l > 0.$$

In particular, for $l = 1$,

$$\boldsymbol{\Gamma}_1 = \boldsymbol{\phi}_1 \, \boldsymbol{\Gamma}_0$$

Since $\boldsymbol{\Gamma}_0$ is nonsingular, we have $\boldsymbol{\phi}_1 = \boldsymbol{\Gamma}_1 \boldsymbol{\Gamma}_0^{-1}$. Pre- and post multiplying the prior moment equation by $\boldsymbol{D}^{-1}$, we further obtain

$$\boldsymbol{\rho}_l = \boldsymbol{\phi}_{\mathbf{1}}^* \boldsymbol{\rho}_{l-1},$$

for $l > 0$, where $\boldsymbol{\phi}_{\mathbf{1}}^* = \boldsymbol{D}^{-1} \, \boldsymbol{\phi}_1 \boldsymbol{D}$ with $\boldsymbol{D}$ being the diagonal matrix of standard deviations of the component series.

VAR(p) Models

Given $\boldsymbol{Y}_t \in R^d$. In this case with d endogenous variables and p lags, we can write the VAR(p) model in matrix notation as

$$\boldsymbol{Y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{Y}_{t-p} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{Y}_t$ and its lagged values, and $\boldsymbol{\varepsilon}_t$ are d$\times$ 1 vectors and $\boldsymbol{\Phi}_1, \cdots \boldsymbol{\Phi}_p$ are $d \times d$ matrices of constants to be estimated.

$\{\boldsymbol{Y}_t\}$ is stationary if roots of the polynomial

$$|\boldsymbol{I} - \boldsymbol{\phi}_1 \boldsymbol{z} - \cdots - \boldsymbol{\phi}_p \boldsymbol{z}^p|$$

are greater than 1 in modulus. (A generalization of univariate case)

Example: IBM and S&P 500 Index

$$\boldsymbol{Y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\phi}_3 \boldsymbol{Y}_{t-3} + \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t \sim (\boldsymbol{0}, \boldsymbol{\Sigma})$$

Table: Estimation results of a VAR(3) model for the monthly log returns, in percentages, of IBM stock and the S&P 500 Index.

	$\boldsymbol{\Phi}_0$	$\boldsymbol{\Phi}_1$		$\boldsymbol{\Phi}_3$		$\boldsymbol{\Sigma}$	
(a) Full Model							
Estimate	1.20	0.011	0.108	0.039	-0.112	44.44	23.51
	0.58	-0.013	0.084	-0.007	-0.105	23.51	31.29
SD	0.23	0.043	0.051	0.044	0.052		
	0.19	0.036	0.043	0.037	0.044		
(b) Simplified Model (Remove insignificant parameters)							
Estimate	1.24	0	0.117	0	-0.083	44.48	23.51
	0.57	0	0.073	0	-0.109	23.51	31.29
SD	0.23		0.040		0.040		
	0.19		0.033		0.033		

Example: IBM and S&P 500 Index

The fitted model:

$$\begin{bmatrix} \mathrm{IBM}_t \\ \mathrm{SP}_t \end{bmatrix} = \begin{bmatrix} 1.24 \\ 0.57 \end{bmatrix} + \begin{bmatrix} 0.117 \\ 0.073 \end{bmatrix} \mathrm{SP}_{t-1} - \begin{bmatrix} 0.083 \\ 0.109 \end{bmatrix} \mathrm{SP}_{t-3} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix},$$

indicating that SP_t is the driving factor of the bivariate series.

$Q_2(4) = 18.17$ (p-value: 0.111) and $Q_2(8) = 41.26$ (0.051). Hence the fitted model is adequate.

Q: Degrees of freedom?

Degrees of freedom are $12 (= md^2 - g = 4 \times 2^2 - 4)$ and 28.

Equivalence of VAR(p) and VAR(1)

In order to compute explicit solutions of VAR(p) models, we make use of the key fact that any VAR(p) model is equivalent to some VAR(1) model after introducing appropriate additional variables.

This is an important simplification as VAR(1) models can be characterized with simple intuitive formulas.

To illustrate this point, first write down a bivariate model of order one in matrix notation, that is,

$$\boldsymbol{x_t} = \boldsymbol{A_1 x_{t-1}} + \boldsymbol{s_t} + \boldsymbol{\epsilon_t}$$

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

And explicitly

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t-1} + s_{1,t} + \epsilon_{1,t}$$
$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t-1} + s_{2,t} + \epsilon_{2,t}$$

Equivalence of VAR(p) and VAR(1)

We observe that any VAR(1) model becomes an arithmetic multivariate random walk if $\boldsymbol{A_1}$ is an identity matrix and $\boldsymbol{s_t}$ is a constant vector. In particular, in the bivariate case, a VAR(1) is a random walk if

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and $\boldsymbol{s_t} = (s_1\ s_2)$.

Consider now a bivariate VAR(2) model of order two:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_{1,t-2} \\ x_{2,t-2} \end{bmatrix} + \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

Equivalence of VAR(p) and VAR(1)

Let us introduce a new vector variable $z_t = x_{t-1}$. The VAR(2) model can then be rewritten as follows:

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t-1} + b_{11}z_{1,t-1} + b_{12}z_{2,t-1} + s_{1,t} + \epsilon_{1,t}$$
$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t-1} + b_{21}z_{1,t-1} + b_{22}z_{2,t-1} + s_{2,t} + \epsilon_{2,t}$$
$$z_{1,t} = x_{1,t-1}$$
$$z_{2,t} = x_{2,t-1}$$

or in matrix form

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} s_{1,t} \\ s_{2,t} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ 0 \\ 0 \end{bmatrix}$$

The above considerations can be generalized. Any AR(p) or VAR(p) model can be transformed into a first-order VAR(1) model by adding appropriate variables. In particular, an n-dimensional VAR(p) model of the form

$$\boldsymbol{x_t} = \left(\boldsymbol{A_1}L + \boldsymbol{A_2}L^2 + \cdots + \boldsymbol{A_p}L^p\right)\boldsymbol{x_t} + \boldsymbol{s_t} + \boldsymbol{\epsilon_t}$$

is transformed into the following np-dimensional VAR(1) model.

Equivalence of VAR(p) and VAR(1)

$$\boldsymbol{X_t = AX_t + S_t + W_t}$$

where

$$\boldsymbol{X_t} = \begin{bmatrix} \boldsymbol{x_t} \\ \boldsymbol{x_{t-1}} \\ \vdots \\ \boldsymbol{x_{t-p+1}} \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A_1} & \boldsymbol{A_2} & \cdots & \boldsymbol{A_{p-1}} & \boldsymbol{A_p} \\ \boldsymbol{I_n} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I_n} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I_n} & \boldsymbol{0} \end{bmatrix}, \boldsymbol{S_t} = \begin{bmatrix} \boldsymbol{s_t} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}, \boldsymbol{W_t} = \begin{bmatrix} \boldsymbol{\epsilon_t} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}$$

$\boldsymbol{X_t}$, $\boldsymbol{S_t}$ and $\boldsymbol{W_t}$ are $np \times 1$ vectors and $\boldsymbol{A}$ is a $np \times np$ square matrix. In order to compute explicit solutions of higher-order VAR processes, we have therefore only to consider VAR(1) models.

It can be demonstrated that the reverse characteristic equation of this VAR(1) system and that of the original VAR(p) system have the same roots.

Forecasting with VAR Models

One of the key objectives of financial modeling is forecasting.

Forecasting entails a criterion for forecasting as we have to concentrate a probability distribution in a point forecast. A widely used criterion is the minimization of the mean square error (MSE).

Suppose that a process y_t is generated by a VAR(p) process. It can be demonstrated that the optimal h-step ahead forecast according to the MSE criterion is the conditional expectation:

$$E_t(\boldsymbol{y}_{t+h}) \equiv E(\boldsymbol{y}_{t+h} | \boldsymbol{y}_s, s \leq t)$$

If the error terms are strict white noise, then the optimal forecast of a VAR model can be computed as follows:

$$E_t(\boldsymbol{y}_{t+h}) = \boldsymbol{v} + \boldsymbol{A}_1 E_t(\boldsymbol{y}_{t+h-1}) + \cdots + \boldsymbol{A}_p E_t(\boldsymbol{y}_{t+h-p})$$

This formula remains valid if the noise term is a martingale difference sequence (see Chapter 6 for a definition). If the error term is white noise, the above forecasting formula will be the best linear predictor.

Forecasting

The minimum mean squared forecast error is the criterion commonly used to produce point forecasts of a vector time series. That is, the forecast origin n, the forecast $Y_n(h)$ is obtained by $Y_n(h) = argmin_g E[(y_{n+h} - g)^2 | F_n]$ where F_n is the information set available at time n.

Similarly to the univariate case, the point forecasts turn out to be the conditional expectation as : $Y_n(h) = E[y_{n+h} | F_n]$,

It turns out that the forecasts can be obtained recursively as h $i$ncreases.

$$Y_n(1) \ = \boldsymbol{\phi}_0 + \sum_{\boldsymbol{i=1}}^{\boldsymbol{p}} \boldsymbol{\phi}_i \, \boldsymbol{y}_{n+1-i}$$

and the associated forecast error is $\boldsymbol{e}_{n+1} = \hat{\boldsymbol{\varepsilon}}_{n+1}$. The covariance matrix of the forecast error is $\boldsymbol{\Sigma}$.

For 2-step ahead forecasts, we substitute $\boldsymbol{Y}_{n+1}$ by its forecast:

$$Y_n(2) = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 Y_n(1) + \sum_{\boldsymbol{j=2}}^{\boldsymbol{p}} \boldsymbol{\phi}_j \, \boldsymbol{y}_{n+2-j}$$

and the forecast error is $\boldsymbol{e}_{n+2} = \hat{\boldsymbol{\varepsilon}}_{n+2} + \hat{\boldsymbol{\phi}}_1 [\boldsymbol{Y}_{n+1} - \hat{\boldsymbol{Y}}_{n+1}] = \hat{\boldsymbol{\varepsilon}}_{n+2} + \hat{\boldsymbol{\phi}}_1 \hat{\boldsymbol{\varepsilon}}_{n+1}$.

For a stationary VAR model, the forecast $Y_n(h)$converges to $\boldsymbol{\mu_Y}$ as h→ ∞. This is the mean-reverting property of a stationary time series. In fact, one can show that $\boldsymbol{\phi}(\boldsymbol{B}) Y_n(h) = \boldsymbol{0}$, for $\mathrm{h} > p$.

Example: IBM and S&P 500 Index

Table: Forecasts of a VAR(3) model for the monthly log returns, in percentages, of IBM stock and the S&P 500 Index: forecast origin December 1999.

Step	1	2	3	4	5	6
IBM	1.40	1.12	0.82	1.21	1.27	1.31
SD	6.67	6.70	6.70	6.72	6.72	6.72
S&P	0.32	0.38	-0.02	0.53	0.56	0.61
SD	5.59	5.61	5.61	5.64	5.64	5.64

Estimation

Two methods are commonly used. They are the least squares method and maximum likelihood method. Under normality, the two methods are asymptotically equivalent.

Suppose that the sample $\{\boldsymbol{y}_t\}_{t=1}^{T}$ is available and a VAR(p) is entertained. That is,

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad t = p+1, \cdots, T.$$

<u>Least squares (LS) method</u>

The VAR(p) model can be written as

$$\boldsymbol{y}_t' = \boldsymbol{x}_t' \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t',$$

where $\boldsymbol{x}_t = (1, \boldsymbol{y}_{t-1}', \cdots, \boldsymbol{y}_{t-p}')'$ is a $(d \times p + 1)$-dimensional vector and $\boldsymbol{\beta}' = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p]$ is a $d \times (dp+1)$ matrix. The least squares estimate of $\boldsymbol{\beta}$ is

$$\widetilde{\boldsymbol{\beta}} = \left[\sum_{t=p+1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1} \sum_{t=p+1}^{n} \boldsymbol{x}_t \boldsymbol{y}_t'.$$

The LS residual is

$$\tilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{y}_t - \sum_{i=1}^{p} \widetilde{\boldsymbol{\phi}}_i \boldsymbol{y}_{t-i}, t = p+1, \cdots, T.$$

LS estimation

The LS estimate of noise covariance is

$$\tilde{\boldsymbol{\Sigma}} = \frac{1}{T-(d+1)p-1} \sum_{\boldsymbol{t=p+1}}^{\boldsymbol{T}} \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \tilde{\boldsymbol{\varepsilon}}'_{\boldsymbol{t}}.$$

For a stationary VAR(p) model with independent error terms $\boldsymbol{\varepsilon}_t$, it can be shown that the least squares estimate $\tilde{\boldsymbol{\phi}}$ is consistent.

Furthermore, let $\tilde{\boldsymbol{b}} = \mathrm{vec}(\tilde{\boldsymbol{\beta}})$, where $\mathrm{vec}(\boldsymbol{A})$ is the column stacking operator of matrix $\boldsymbol{A}$. Then $\tilde{\boldsymbol{b}}$ is asymptotically normal with mean $\mathrm{vec}(\boldsymbol{\beta})$ and covariance matrix

$$\mathrm{Cov}(\tilde{\boldsymbol{\beta}}) = \tilde{\boldsymbol{\Sigma}} \otimes \left[\sum_{t=p+1}^{T} \boldsymbol{x}_t \boldsymbol{x}'_t\right]^{-1},$$

where $\otimes$ denotes the Kronecker product; see Lütkepohl (1991) or Result 7.10 of Johnson and Wichern (2002, *Applied Multivariate Statistical Analysis*).

<u>Maximum likelihood method</u>

The coefficient estimates are the same as those of LS estimates. However, the estimate of $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-p} \sum_{t=p+1}^{T} \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \tilde{\boldsymbol{\varepsilon}}'_{\boldsymbol{t}}.$$

Example: GDP of UK,Canada,US

Consider the quarterly growth rates, in percentages, of real gross domestic product (GDP) of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The data were seasonally adjusted and downloaded from the database of Federal Reserve Bank at St. Louis. The GDP were in millions of local currency, and the growth rate denotes the differenced series of log GDP.

We employ a VAR(2) model: k=3, p=2, and T =125.

$\tilde{\Sigma}_\epsilon = \begin{bmatrix} 0.299 & 0.028 & 0.079 \\ & 0.309 & 0.148 \\ & & 0.379 \end{bmatrix}$ and $\hat{\Sigma}_\epsilon = \begin{bmatrix} 0.282 & 0.027 & 0.074 \\ & 0.292 & 0.139 \\ & & 0.357 \end{bmatrix}$

The two estimates of the covariance matrix differ by a factor of 116/123=0.943. From the output, the t-ratios indicate that some of the LS estimates are not statistically significant at the usual 5% level.

8_Est.R

Maximum Likelihood Estimators

Under the assumption of Gaussian innovations, ***maximum likelihood (ML)*** estimation methods coincide with LS estimation methods when we condition on the first p observations.

ML methods try to find the distribution parameters that maximize the likelihood function.

In the case of a multivariate mean-adjusted VAR(p) process, the given sample data are T observations of the d-variate variable $y_t, t = 1, \dots, T$ and a pre-sample of p initial conditions $y_{-p+1}, \dots, y_0$. If we assume that the process is stationary and that innovations are Gaussian white noise, the variables $y_t, t = 1, \dots, T$ will also be jointly normally distributed.

Maximum Likelihood Estimators

The noise terms $(\epsilon_1, \dots,\ \epsilon_T)$ are assumed to be independent with constant covariance matrix $\mathbf{\Sigma}$ and, therefore, $\boldsymbol{\epsilon} = vec(\epsilon_1, \dots,\ \epsilon_T)$ has covariance matrix $\boldsymbol{\Sigma_\epsilon} = \boldsymbol{I_T} \otimes \boldsymbol{\Sigma}$. Under the assumption of Gaussian noise, $\boldsymbol{\epsilon}$ has the following dT-variate normal density:

$$
\begin{aligned}
f_\epsilon &= (2\pi)^{-\frac{dT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon_t}'\boldsymbol{\Sigma^{-1}}\boldsymbol{\epsilon_t}\right) \\
&= (2\pi)^{-\frac{dT}{2}} |\boldsymbol{I_T} \otimes \boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}'\left(\boldsymbol{I_T} \otimes \boldsymbol{\Sigma^{-1}}\right)\boldsymbol{\epsilon}\right)
\end{aligned}
$$

Using

$$
\begin{gathered}
\epsilon_1 = y_1 - A_1 y_0 - A_2 y_{-1} - \cdots - A_p y_{1-p} \\
\epsilon_2 = y_2 - A_1 y_1 - A_2 y_0 - \cdots - A_p y_{2-p} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\epsilon_p = y_1 - A_1 y_{p-1} - A_2 y_{p-2} - \cdots - A_p y_0 \\
\epsilon_{p+1} = y_{p+1} - A_1 y_p - A_2 y_{p-1} - \cdots - A_p y_1 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\epsilon_{T-1} = y_{T-1} - A_1 y_{T-2} - A_2 y_{T-3} - \cdots - A_p y_{T-p-1} \\
\epsilon_T = y_T - A_1 y_{T-1} - A_2 y_{T-2} - \cdots - A_p y_{T-p}
\end{gathered}
$$

Maximum Likelihood Estimators

Given the model equation $\boldsymbol{y} = \boldsymbol{x}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$, we write the log-likelihood as follows:

$$\log(L) = -\frac{dT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{\boldsymbol{t}}'\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{\epsilon}_{\boldsymbol{t}}$$

$$= -\frac{dT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\alpha})'\left(\boldsymbol{I}_{\boldsymbol{T}} \otimes \boldsymbol{\Sigma}^{-\mathbf{1}}\right)(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\alpha})$$

$$= -\frac{dT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}| - \frac{1}{2}\mathrm{trace}\left((\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X})'\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-\mathbf{1}}(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X})\right)$$

Example: GDP of UK,Canada,US

In practice, we use some available packages in R to perform estimation. For example, we can use the command VAR in the MTS package to estimate a VAR model.

From the output, the fitted VAR(2) model for the percentage growth rates of quarterly GDP of United Kingdom, Canada, and United States is

$$\boldsymbol{x}_t = \begin{bmatrix} 0.13 \\ 0.12 \\ 0.29 \end{bmatrix} + \begin{bmatrix} 0.38 & 0.10 & 0.05 \\ 0.35 & 0.34 & 0.47 \\ 0.49 & 0.24 & 0.24 \end{bmatrix} \boldsymbol{x}_{t-1} + \begin{bmatrix} 0.06 & 0.11 & 0.02 \\ -0.19 & -0.18 & -0.01 \\ -0.31 & -0.13 & 0.09 \end{bmatrix} \boldsymbol{x}_{t-2} + \boldsymbol{\epsilon}_t,$$

8_VAR.R

Building VAR(p) Models

- ❑ Order selection: use AIC.
- ❑ Estimation: use maximum likelihood estimation or ordinary least squares method
- ❑ Model checking: as univariate case
- ❑ Forecasting: similar to univariate case

Remark: Simple VAR models turn to be sufficient to model asset returns.

Order identification – information criteria

The log-likelihood function is:

$$\log(L) = -\frac{dT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma_\epsilon}| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\epsilon_t'\Sigma^{-1}\epsilon_t}$$

Akaike proposed a criterion based on information theoretic considerations. It is commonly called the ***Akaike information criterion (AIC)***, proposes to determine the model order by minimizing the following expression:

$$AIC(p) = \log\left|\widehat{\boldsymbol{\Sigma}}(\boldsymbol{p})\right| + \frac{2pd^2}{T}$$

Neither the FPE nor the AIC estimators are consistent estimators in the sense that they determine the correct model order in the limit of an infinite sample.

Different but consistent criteria have been proposed. Among them, the ***Bayesian information criterion (BIC)*** is quite popular. The BIC chooses the model that minimizes the following expression:

$$BIC(p) = \log\left|\widehat{\boldsymbol{\Sigma}}(\boldsymbol{p})\right| + \frac{\log T}{T}pd^2$$

AIC penalizes each parameter by a factor of 2. BIC penalizes more complicated model more heavily, e.g. when log(T)>2.

Order determination – LR test

Sequential Chi-square test: Let P be a positive integer, denoting the maximum order entertained. For $\ell > 0$, consider the hypothesis

$$H_0: VAR(\ell - 1) \text{ vs } H_a: VAR(\ell).$$

This is to test H_0: $\boldsymbol{\phi}_\ell = \mathbf{0}$ vs H_a: $\boldsymbol{\phi}_\ell \neq \mathbf{0}$

in the autoregression

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\phi}_{\ell-1} \boldsymbol{y}_{t-\ell+1} + \boldsymbol{\phi}_\ell \boldsymbol{y}_{\boldsymbol{t}-\boldsymbol{\ell}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}}, \qquad t = P + 1, \cdots, T.$$

How to proceed? Likelihood ratio test in multivariate linear regression.

If $\boldsymbol{\varepsilon}_{\boldsymbol{t}}$ is Gaussian, we have

$$\log(L) = -\frac{dT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\epsilon}_{\boldsymbol{t}}' \boldsymbol{\Sigma}^{-\mathbf{1}} \boldsymbol{\epsilon}_{\boldsymbol{t}}$$

then we can use the Likelihood Ratio (LR) test; see Result 7.1 on page 393 of Johnson and Wichern (2002, *Applied Multivariate Statistical Analysis*).

Let $\widehat{\boldsymbol{\Sigma}}_\ell$ be ML estimate of $\boldsymbol{\Sigma}$ of fitting a VAR(ℓ) model to $\boldsymbol{y}_t$. Then, the LR test statistic is

$$M(\ell) = -\left(T - P - \frac{3}{2} - \ell d\right) \ln\left[\frac{|\widehat{\boldsymbol{\Sigma}}_\ell|}{|\widehat{\boldsymbol{\Sigma}}_{\ell-1}|}\right],$$

where $\widehat{\boldsymbol{\Sigma}}_\ell$ are ML estimate of $\boldsymbol{\Sigma}$ using $t = P + 1, \cdots, T$. For large T, $M(\ell)$ is approximately a chi-square distribution with d^2 degrees of freedom.

Example: CPI and IP

The cross-correlation function between changes in CPI (Consumer Price Index) and IP (industrial production) is plotted in Figure 10.6, which was created by the ccf function in R.

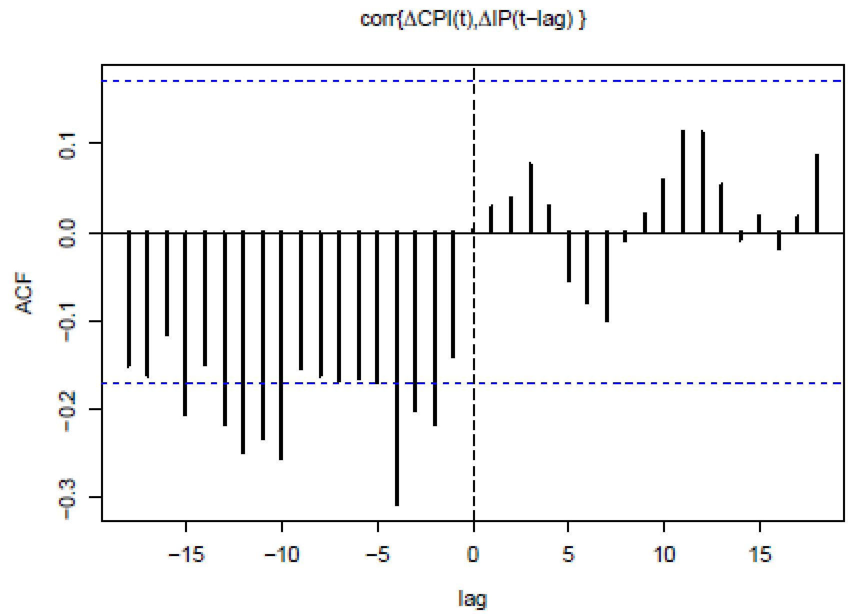


Fig. 10.6. *CCF for ΔCPI and ΔIP. Note the negative correlation at negative lags, that is, between the CPI and future values of IP.*

- ❑ Data: 7_CPI.dat.csv, 7_IP.dat.csv
- ❑ R: 7_Cross-correlation.R

Example: VAR model for ΔCPI and ΔIP

VAR processes were fit to the two series using **R's function ar**. AIC as a function of p is shown below. The two best fitting models are VAR(1) and VAR(5), with the latter being slightly better by AIC.

Although BIC is not part of AR's output, it can be calculated. For this example, BIC is much smaller for the VAR(1) model than for the VAR(5) model. For this reason and because the VAR(1) model is so much simpler to analyze, we will use the VAR(1) model.

p	0	1	2	3	4
AIC	127.99	0.17	1.29	5.05	3.40

5	6	7	8	9	10
0.00	6.87	9.33	10.83	13.19	14.11

$$\widehat{\boldsymbol{\Phi}} = \begin{pmatrix} 0.767 & 0.0112 \\ -0.330 & 0.3014 \end{pmatrix} \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 5.68e-06 & 3.33e-06 \\ 3.33e-06 & 6.73e-06 \end{pmatrix}$$

It is useful to look at the two off-diagonals of $\widehat{\boldsymbol{\Phi}}$. Since $\phi_{1,2} = 0.01 \approx 0$, $Y_{2,t-1}$ (lagged IP) has little influence on $Y_{1,t}$(CPI), and since $\phi_{2,1} = -0.330$, $Y_{1,t-1}$ (lagged CPI) has a substantial negative effect on $Y_{2,t}$ (IP).

- R: 7_BivariateARModel.R

Problems with VAR models in practice

We have considered only a simple model with two variables and only a small number of lags for each.

In practice, since we are not considering any moving average errors, the autoregressions would probably have to have more lags to be useful for prediction. Otherwise, univariate ARMA models would do better.

Suppose that we consider say six lags (p=6) for each variable and we have a small system with four variables (d=4). Then each equation would have 24 parameters to be estimated and we thus have 96 parameters to estimate overall.

This **overparameterization** is one of the major problems with VAR models. Unless the sample size is large, estimating that many parameters will consume a lot of degree of freedom with all the problems associated with that.

Problems with VAR models in practice

Strictly speaking, in a d-variable VAR model, all the d variables should be (joint) stationary. If they are not stationary, we have to transform (e.g., by first-differencing) the data appropriately. If some of the variables are non-stationary, and the model contains a mix of I(0) and I(1), then the transforming of data will not be easy.

Since the individual coefficients in the estimated VAR models are often difficult to interpret, the practitioners of this technique often estimate the so-called impulse response function. The impulse response function traces out the response of the dependent variable in the VAR system to shocks in the error terms, and traces out the impact of such shocks for several periods in the future.

Vector moving average model of order q

VMA(q) model:

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} - \cdots - \boldsymbol{\Theta}_q \boldsymbol{\varepsilon}_{t-q}$$

where $\boldsymbol{\mu}$ is a constant vector denoting the mean of $\boldsymbol{Y}_t$, $\boldsymbol{\theta}_i$ are $k \times k$ matrices with $\boldsymbol{\theta}_q \neq \mathbf{0}$ indicating dynamic dependence, and $\{\boldsymbol{\varepsilon}_t\}$ is defined as before. Using the backshift operator the model becomes $\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{\theta}(B)\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\theta}(B) = \boldsymbol{I} - \sum_{i=1}^{q} \boldsymbol{\theta}_i B^i$ is a matrix polynomial of degree q.

$$\boldsymbol{\Theta}_0 = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix}, \boldsymbol{\Theta}_q = \begin{bmatrix} \boldsymbol{\Theta}_{11}^q & \cdots & \boldsymbol{\Theta}_{1d}^q \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Theta}_{d1}^q & \cdots & \boldsymbol{\Theta}_{dd}^q \end{bmatrix}$$

VMA(1) model:

Again, we start with the 2-dimensional vector moving-average (VMA) model,

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t - \theta_1 \boldsymbol{\varepsilon}_{t-1}.$$

The model can be written as

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} \theta_{1,11} & \theta_{1,12} \\ \theta_{1,21} & \theta_{1,22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix},$$

or equivalently,

$$Y_{1t} = \mu_1 + \boldsymbol{\varepsilon}_{1t} - \theta_{1,11} \boldsymbol{\varepsilon}_{1,t-1} - \theta_{1,12} \boldsymbol{\varepsilon}_{2,t-1}$$
$$Y_{2t} = \mu_2 + \boldsymbol{\varepsilon}_{2t} - \theta_{1,21} \boldsymbol{\varepsilon}_{1,t-1} - \theta_{1,22} \boldsymbol{\varepsilon}_{2,t-1}$$

Thus, $\theta_{1,12}$ is the impact of $\varepsilon_{2,t-1}$ on Y_{1t} in the presence of $\varepsilon_{1,t-1}$.

Vector ARMA models

A d-dimensional time series $\boldsymbol{Y}_t$ follows a vector ARMA, or VARMA(p,q), model

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\phi}_j \boldsymbol{Y}_{t-j} + \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \boldsymbol{\Theta}_j \boldsymbol{\varepsilon}_{t-j}$$

$$\boldsymbol{\phi}(B)\boldsymbol{Y_t} = \boldsymbol{\mu} + \boldsymbol{\theta}(B)\boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\mu}$ is a constant vector,$\phi(B) = \boldsymbol{I} - \sum_{i=1}^{p} \phi_i B^i$ and $\boldsymbol{\theta}(B) = \boldsymbol{I} - \sum_{i=1}^{q} \boldsymbol{\theta}_i B^i$ are two matrix polynomials, and $\boldsymbol{\varepsilon}_t$ is a sequence of independent and identically distributed random vectors with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}$. We require two additional conditions:

1. $\phi(B)$ and $\boldsymbol{\theta}(B)$ are left co-prime, i.e. if $\boldsymbol{u}(B)$ is a left common factor of $\phi(B)$ and $\boldsymbol{\theta}(B)$, then $|\boldsymbol{u}(B)|$ is a non-zero constant. Such a polynomial matrix is called a *uni-modular* matrix. In theory, $\boldsymbol{u}(B)$ is uni-modular if and only if $\boldsymbol{u}^{-1}(B)$ exits and is a matrix polynomial.
2. The MA order q is as small as possible and the AR order p is as small as possible for that q, and the matrices ϕ_p and $\boldsymbol{\theta}_q$ satisfy the condition that Rank[ϕ_p, $\boldsymbol{\theta}_q$]= dim($\boldsymbol{Y}_t$).

These two conditions are sufficient conditions for VARMA models to be identifiable. In the literature, these conditions are referred to as *block identifiability*.

Identifiability

Unlike the VAR or VMA models, VARMA models encounter the problem of identifiability. For a given linear vector process,

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\varphi}_i \boldsymbol{\varepsilon}_{t-i}$$

where $\boldsymbol{\varphi}_0 = \boldsymbol{I}$, and $\{\boldsymbol{\varepsilon}_t\}$ is an IID sequence of random vectors with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}$.

A VARMA model is said to be identifiable if the matrix polynomials $\phi(B)$ and $\boldsymbol{\theta}(B)$ are uniquely determined by $\boldsymbol{\varphi}_i$.

There are cases for which multiple pairs of AR and MA matrix polynomials give rise to the same $\boldsymbol{\varphi}_i$s. We use simple bivariate models to discuss the issue.

Example: VMA or VAR?

Consider the VMA(1) model

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix}.$$

This is a well-defined VMA(1) model. However, it can also be written as the VAR(1) model

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

To see this, the VMA(1) model implies that

$Y_{1t} = \varepsilon_{1t} - 2\varepsilon_{2,t-1}$ and $Y_{2t} = \varepsilon_{2t}$.

In other words, Y_{2t} is a white noise series. As such, we have

$$Y_{1t} = \varepsilon_{1t} - 2\varepsilon_{2,t-1} = \varepsilon_{1t} - 2Y_{2,t-1}.$$

Consequently, we have $Y_{1t} + 2Y_{2,t-1} = \varepsilon_{1t}$ and $Y_{2t} = \varepsilon_{2t}$, which is precisely the VAR(1) model. This type of non-uniqueness in model specification is harmless because either model can be used in a real application.

Example: Parameters?

Consider the VARMA(1,1) model

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} 0.3 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix}.$$

It is easy to see that the model is identical to

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & 2+\omega \\ 0 & \beta \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} 0.3 & \omega \\ 0 & \beta \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix},$$

for any $\omega \neq 0$ and $\beta \neq 0$. From above, we have

$$Y_{1t} = 0.8Y_{1,t-1} + 2Y_{2,t-1} + \varepsilon_{1t} - 0.3\varepsilon_{1,t-1}$$
$$Y_{2t} = \varepsilon_{2t}.$$

R lab

Consider the quarterly growth rates, in percentages, of real gross domestic product (GDP) of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The data were seasonally adjusted and downloaded from the database of Federal Reserve Bank at St. Louis.

The GDP were in millions of local currency, and the growth rate denotes the differenced series of log GDP.

```
> library("MTS")
> da=read.table("q-gdp-ukcaus.txt",header=T)
> gdp=log(da[,3:5])
> dim(gdp)
> z=gdp[2:126,]-gdp[1:125,] ## Growth rate
> z=z*100 ## Percentage growth rates
> dim(z)
> MTSplot(z)
```

Problem 1 Plot the three series. Discuss the behaviours of them.

R lab

Problem 2 Estimate CCM of the multiple series. Interpret the cross-dependence among them. (R function ccm)

Problem 3 Among others, we decided to use VAR model for the series. Select an appropriate order using AIC. Justify your choice through the LR test.

```
> z1=z/100 ### Original growth rates
> m0=VARorder(z1)
```

Problem 4 Report the fitted model and interpret. Use R function refVAR to obtain a simplified VAR with coefficients are significant at a given level. Interpret the dependence among the GDP growth rates of the three countries.

```
> m1=VAR(z1,2)
> m2=refVAR(m1,thres=1.96)
```

R lab

```
> MTSdiag(m1,adj=12)
```

Problem 5 Plot residuals of the fitted VAR model (m1$residuals without simplified). Are they stationary?

Problem 6 Estimate CCM of the residuals. Is the fitted VAR model adequate?

Problem 7 Apply the multivariate Portmanteau test statistics to the residuals of the fitted VAR(2) model. The results are given in the following R demonstration.

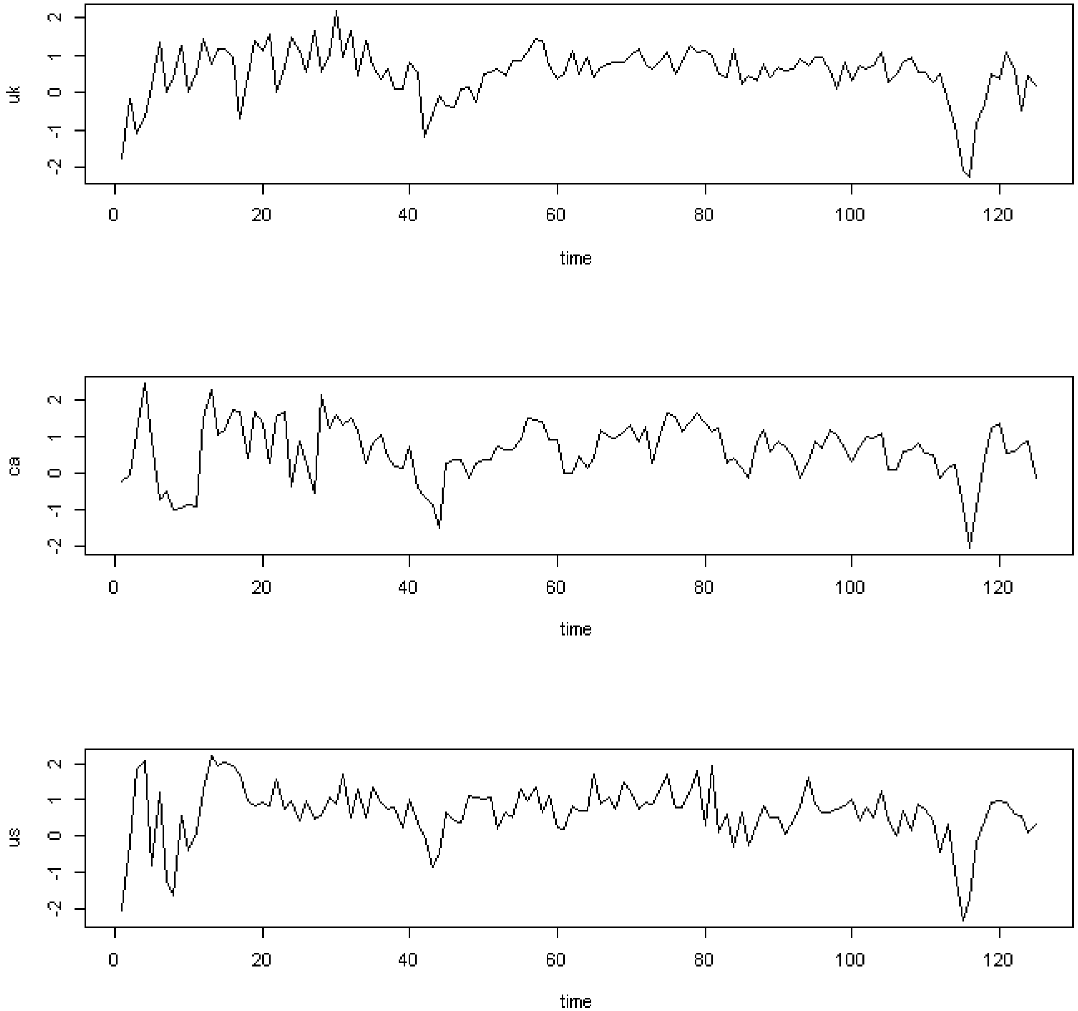
```
> VARpred(m1,8)
> colMeans(z)
> sqrt(apply(z,2,var))
```

Problem 8 Compute forecasts and interpret. Using the fitted model, we consider one-step to eight-step ahead forecasts of the GDP growth rates at the forecast origin 2011.II. Provide the standard errors and root mean-squared errors of the predictions. The root mean squared errors include the uncertainty due to the use of estimated parameters. Analyze the results.

R lab – Results & Discussions

Problem 1 Plot the three series. Discuss the behaviours of them.

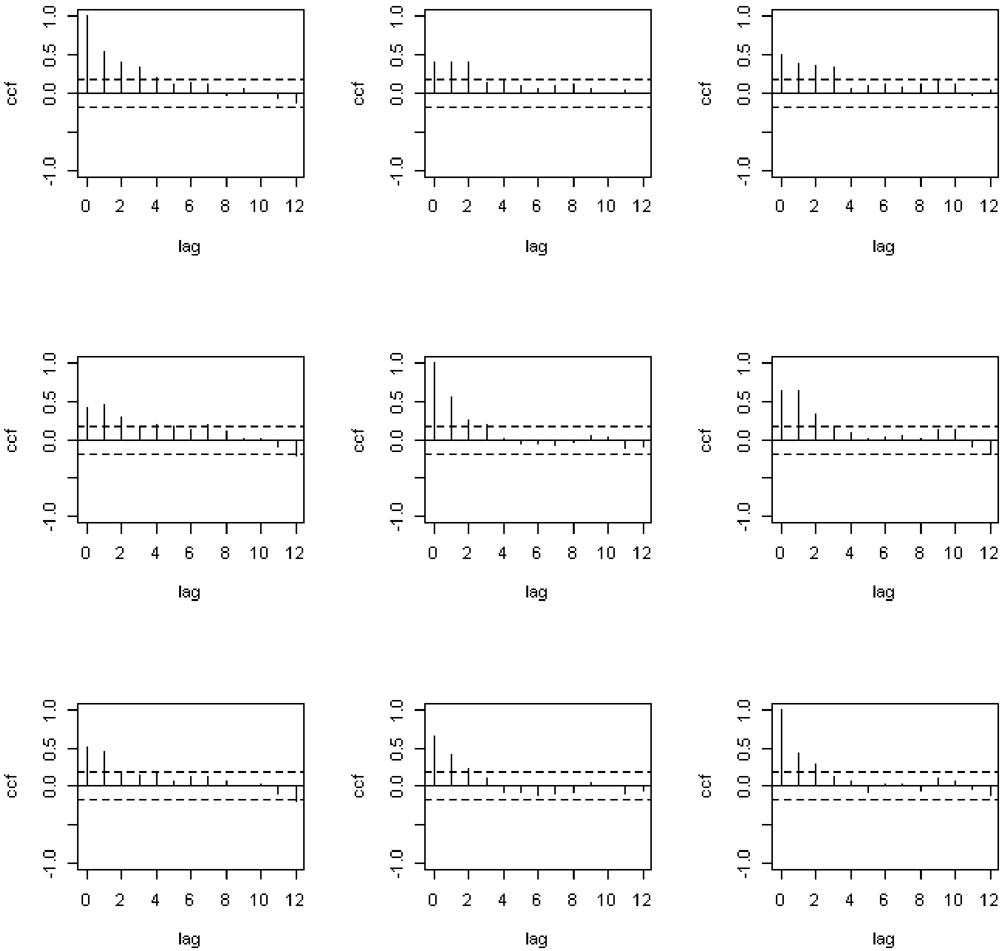
The time series of the three GDP growth rates display similar movement, with e.g. simultaneous drops around 2009.



R lab – Results & Discussions

Problem 2 Estimate CCM of the multiple series. Interpret the cross-dependence among them.

Each of the three GDP growth rates display autocorrelations, indicated by the significant values in the CCM plots on the diagonal. Meanwhile, there are also cross-correlations with feedback features, indicated by the significant values in the CCM plots off the diagonal.

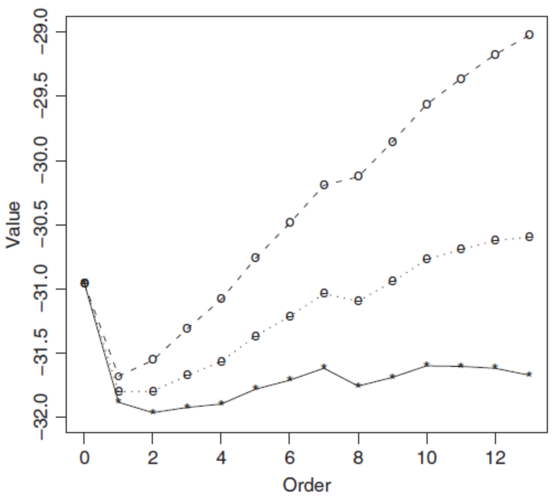


R lab – Results & Discussions

Problem 3 Select an appropriate order using AIC. Justify your choice through the LR test.

.

```
selected order: aic =  2
selected order: bic =  1
selected order: hq =  1
Summary table:
      p     AIC      BIC       HQ     M(p) p-value
[1,]  0 -30.9560 -30.9560 -30.9560   0.0000  0.0000
[2,]  1 -31.8830 -31.6794 -31.8003 115.1329  0.0000
[3,]  2 -31.9643 -31.5570 -31.7988  23.5389  0.0051
[4,]  3 -31.9236 -31.3127 -31.6754  10.4864  0.3126
[5,]  4 -31.8971 -31.0826 -31.5662  11.5767  0.2382
[6,]  5 -31.7818 -30.7636 -31.3682   2.7406  0.9737
[7,]  6 -31.7112 -30.4893 -31.2148   6.7822  0.6598
[8,]  7 -31.6180 -30.1925 -31.0389   4.5469  0.8719
```



GURE 2.2 Information criteria for the quarterly growth rates, in percentages, of real gross domestic ducts of United Kingdom, Canada, and United States from the second quarter of 1980 to the second rter of 2011. The solid, dashed, and dotted lines are for AIC, BIC, and HQ, respectively.

R lab – Results & Discussions

Problem 4 Report the fitted model and interpret. Use R function refVAR to obtain a simplified VAR with coefficients are significant at a given level.

The unconstraint fitted VAR(2) model for the growth rates of quarterly GDP of UK, Canada, and US is:

$$y_t = \begin{bmatrix} 0.0013 \\ 0.0012 \\ 0.0029 \end{bmatrix} + \begin{bmatrix} 0.38 & 0.10 & 0.05 \\ 0.35 & 0.34 & 0.47 \\ 0.49 & 0.24 & 0.24 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.06 & 0.11 & 0.02 \\ -0.19 & -0.18 & -0.01 \\ -0.31 & -0.13 & 0.09 \end{bmatrix} y_{t-2} + \epsilon_t$$

where some of the estimates are not statistically significant at the usual 5% level.

The simplified VAR(2) is:

$$y_t = \begin{bmatrix} 0.0016 \\ - \\ 0.0028 \end{bmatrix} + \begin{bmatrix} 0.47 & 0.21 & - \\ 0.33 & 0.27 & 0.50 \\ 0.47 & 0.23 & 0.23 \end{bmatrix} y_{t-1} + \begin{bmatrix} - & - & - \\ -0.20 & - & - \\ -0.30 & - & - \end{bmatrix} y_{t-2} + \epsilon_t$$

All estimates are now significant at the usual 5% level.

R lab – Results & Discussions

Problem 4 Report the fitted model and interpret. Use R function refVAR to obtain a simplified VAR with coefficients are significant at a given level.

The fitted VAR(2) model is equivalent to:

$UK : UK_t = 0.0016 + 0.47UK_{t-1} + 0.21CA_{t-1} + \epsilon_{UK,t},$

$CA : CA_t = 0.33UK_{t-1} + 0.27CA_{t-1} + 0.5US_{t-1} - 0.2UK_{t-2} + \epsilon_{CA,t},$

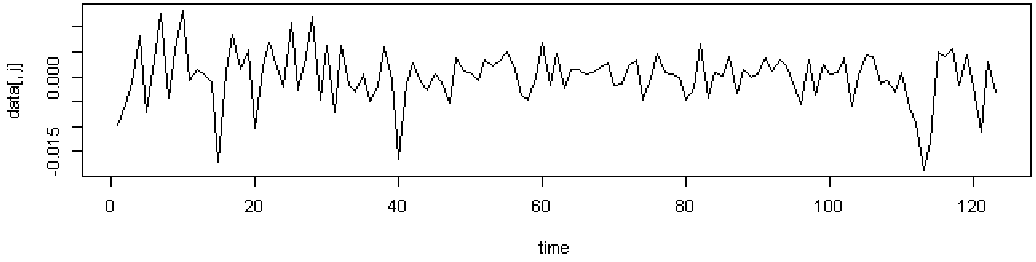
$US : US_t = 0.28 + 0.47UK_{t-1} + 0.23CA_{t-1} + 0.23US_{t-1} - 0.3UK_{t-2} + \epsilon_{US,t}.$

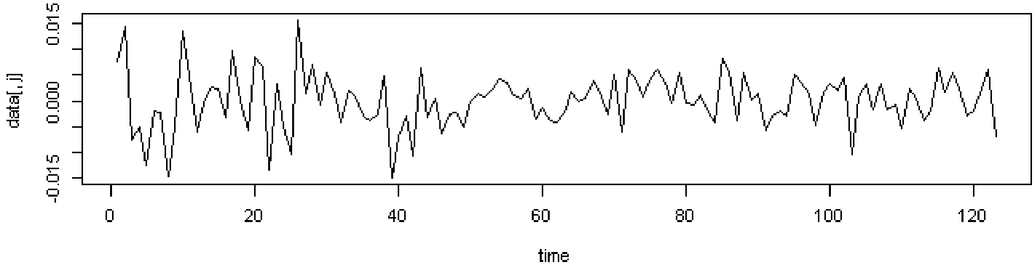
It shows that the GDP growth rate of United Kingdom does not depend on the lagged growth rates of the United States in the presence of lagged Canadian GDP growth rates, but the United Kingdom growth rate depends on the past growth rate of Canada. On the other hand, the GDP growth rate of Canada is dynamically related to the growth rates of United Kingdom and United States. Similarly, the GDP growth rate of the United States depends on the lagged growth rates of United Kingdom and Canada.

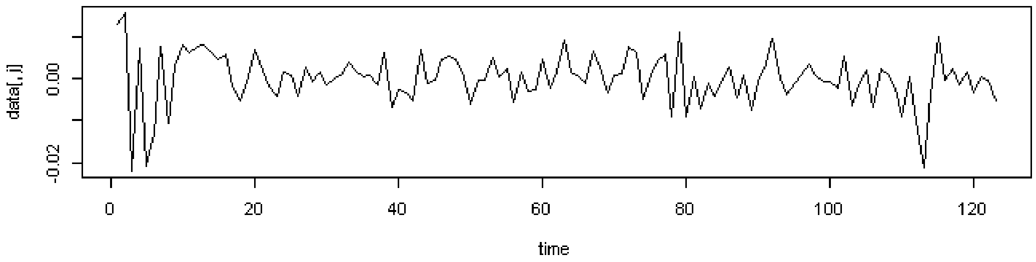
R lab – Results & Discussions

Problem 5 Are the residuals stationary?

The residuals look stationary fluctuating around zero with almost constant spread.



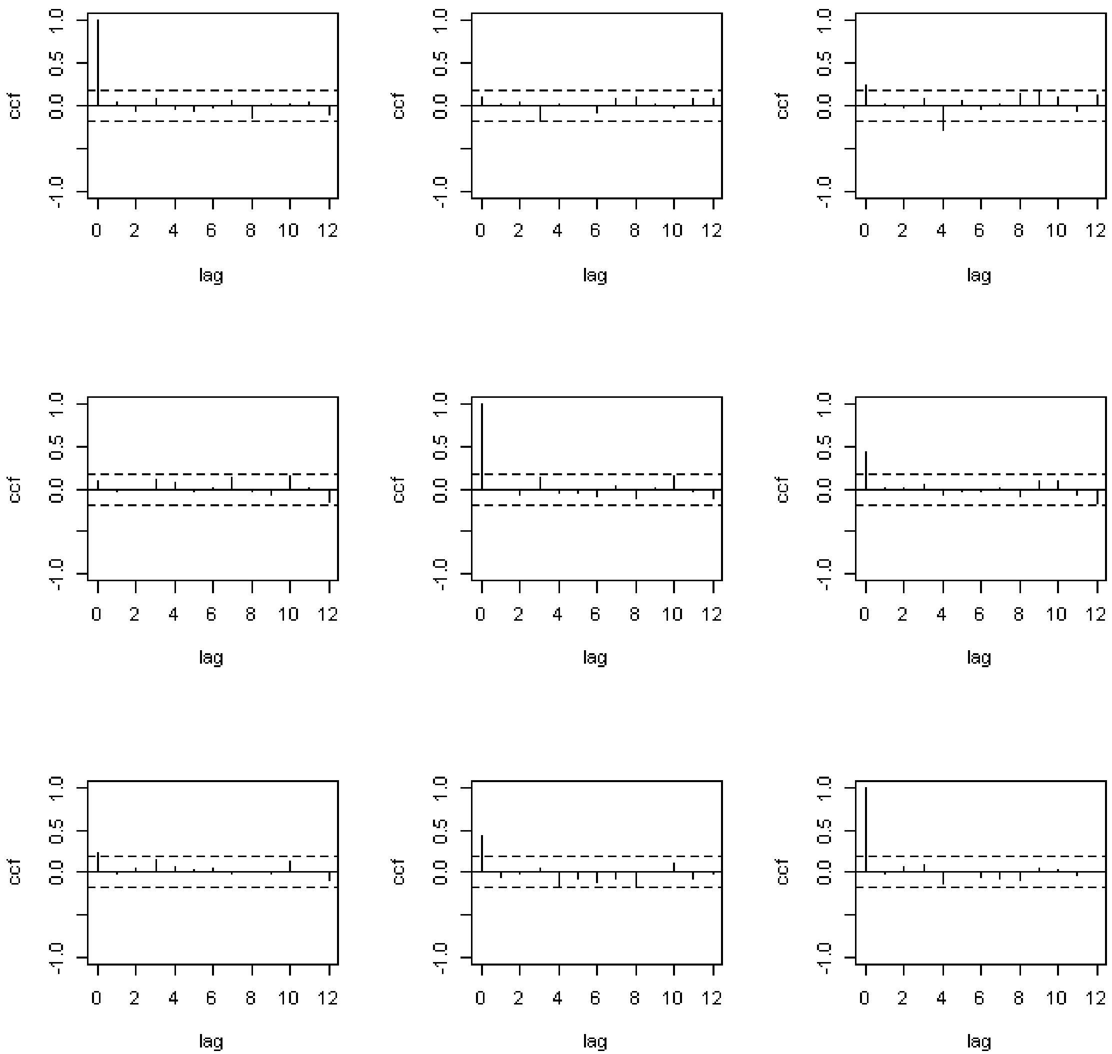




R lab – Results & Discussions

Problem 6 Estimate CCM of the residuals. Is the fitted VAR model adequate?

The residuals have similar features to white noise, with no serial autocorrelations and meanwhile insignificant lead-lag cross-dependence.



R lab – Results & Discussions

Problem 7 Apply the multivariate Portmanteau test statistics to the residuals of the fitted VAR(2) model.

All the statistics are statistically insignificant at 5% level, indicating by the large P-values. We don't reject the null. It implies the VAR(2) model is adequate.

```
> m0=VARorder(z1)
selected order: aic =  2
selected order: bic =  1
selected order: hq =  1
Summary table:
       p     AIC      BIC      HQ      M(p) p-value
 [1,]  0 -30.9560 -30.9560 -30.9560   0.0000  0.0000
 [2,]  1 -31.8830 -31.6794 -31.8003 115.1329  0.0000
 [3,]  2 -31.9643 -31.5570 -31.7988  23.5389  0.0051
 [4,]  3 -31.9236 -31.3127 -31.6754  10.4864  0.3126
 [5,]  4 -31.8971 -31.0826 -31.5662  11.5767  0.2382
 [6,]  5 -31.7818 -30.7636 -31.3682   2.7406  0.9737
 [7,]  6 -31.7112 -30.4893 -31.2148   6.7822  0.6598
 [8,]  7 -31.6180 -30.1925 -31.0389   4.5469  0.8719
 [9,]  8 -31.7570 -30.1279 -31.0952  24.4833  0.0036
[10,]  9 -31.6897 -29.8569 -30.9451   6.4007  0.6992
[11,] 10 -31.5994 -29.5630 -30.7721   4.3226  0.8889
[12,] 11 -31.6036 -29.3636 -30.6936  11.4922  0.2435
[13,] 12 -31.6183 -29.1746 -30.6255  11.8168  0.2238
[14,] 13 -31.6718 -29.0245 -30.5964  14.1266  0.1179
```

R lab – Results & Discussions

Problem 8 Compute forecasts and interpret

TABLE 2.2 Forecasts of Quarterly GDP Growth Rates, in Percentages, for United Kingdom, Canada, and United States via a VAR(2) Model

	Forecasts			Standard Errors			Root MSE		
Step	United Kingdom	Canada	United States	United Kingdom	Canada	United States	United Kingdom	Canada	United States
1	0.31	0.05	0.17	0.53	0.54	0.60	0.55	0.55	0.61
2	0.26	0.32	0.49	0.58	0.72	0.71	0.60	0.78	0.75
3	0.31	0.48	0.52	0.62	0.77	0.73	0.64	0.79	0.75
4	0.38	0.53	0.60	0.65	0.78	0.74	0.66	0.78	0.75
5	0.44	0.57	0.63	0.66	0.78	0.75	0.67	0.78	0.75
6	0.48	0.59	0.65	0.67	0.78	0.75	0.67	0.78	0.75
7	0.51	0.61	0.66	0.67	0.78	0.75	0.67	0.78	0.75
8	0.52	0.62	0.67	0.67	0.78	0.75	0.67	0.78	0.75
Data	0.52	0.62	0.65	0.71	0.79	0.79	0.71	0.79	0.79

The forecast origin is the second quarter of 2011. The last row of the table gives the sample means and sample standard errors of the series.

R lab – Results & Discussions

Problem 8 Compute forecasts and interpret

- The point forecasts of the three series move closer to the sample means of the data as the forecast horizon increases, showing evidence of mean reverting.
- The standard errors and root mean-squared errors of forecasts increase with the forecast horizon.
- The standard errors should converge to the standard errors of the time series as the forecast horizon increases.
- The effect of using estimated parameters is evident when the forecast horizon is small. The effect vanishes quickly as the forecast horizon increases. This is reasonable because a stationary VAR model is mean-reverting.
- The standard errors and mean-squared errors of prediction should converge to the standard errors of the series.
- To construct interval predictions: a two-step ahead 95% interval forecast for U.S. GDP growth rate is $0.49 \pm 1.96 \times 0.71$.

Appendix: Eigenvectors and eigenvalues

Consider a square $n \times n$ matrix $\boldsymbol{A}$ and a n-vector $\boldsymbol{x}$. We call ***eigenvectors*** of the matrix $\boldsymbol{A}$ those vectors such that the following relationship holds

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

for some real number λ. Given an eigenvector $\boldsymbol{x}$ the corresponding λ is called an ***eigenvalue***. Zero is a trivial eigenvalue. Nontrivial eigenvalues are determined by finding the solutions of the equation

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$$

where $\boldsymbol{I}$ is the identity matrix. A $n \times n$ matrix has at most n distinct eigenvalues and eigenvectors.

- ❑ 8_matrix.R

Appendix: Vectoring operators

Given an $m \times n$ matrix,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

the vec operator, written as $vec(\boldsymbol{A})$, stacks the matrix columns in an $mn \times 1$ vector as follows:

$$vec(\boldsymbol{A}) = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

Appendix: Kronecker Product

Given the $m \times n$ matrix $\boldsymbol{A}$ and $p \times q$ matrix $\boldsymbol{B}$,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \qquad \boldsymbol{B} = \begin{pmatrix} b_{11} & \cdots & b_{1q} \\ \vdots & \ddots & \vdots \\ b_{p1} & \cdots & b_{pq} \end{pmatrix}$$

We define the Kronecker product $\boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{B}$ as follows:

$$\boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{pmatrix}$$

The Kronecker product, also called the direct product or the tensor product, gives an $(mp) \times (nq)$ matrix. It can be demonstrated that the tensor product satisfies the associative and distributive property and that, given any four matrices A, B, C, D of appropriate dimensions, the following properties hold:

$$(\boldsymbol{C}' \otimes \boldsymbol{A})\boldsymbol{vec}(\boldsymbol{B}) = \boldsymbol{vec}(\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C})$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})' = (\boldsymbol{A}') \otimes (\boldsymbol{B}')$$

$$Trace(\boldsymbol{A}'\boldsymbol{BCD}') = \left(vec(\boldsymbol{A})\right)'(\boldsymbol{D} \otimes \boldsymbol{B})\boldsymbol{vec}(\boldsymbol{C})$$