Lecture 10: Multiple Time Series (MTS)

Outline

10.1 MTS Basic Concepts

Reference: Chapters 8 and 10 of AFTS.
We shall focus on two series (i.e., the bivariate case):

- Focus on two asset return series for ease of demonstration:

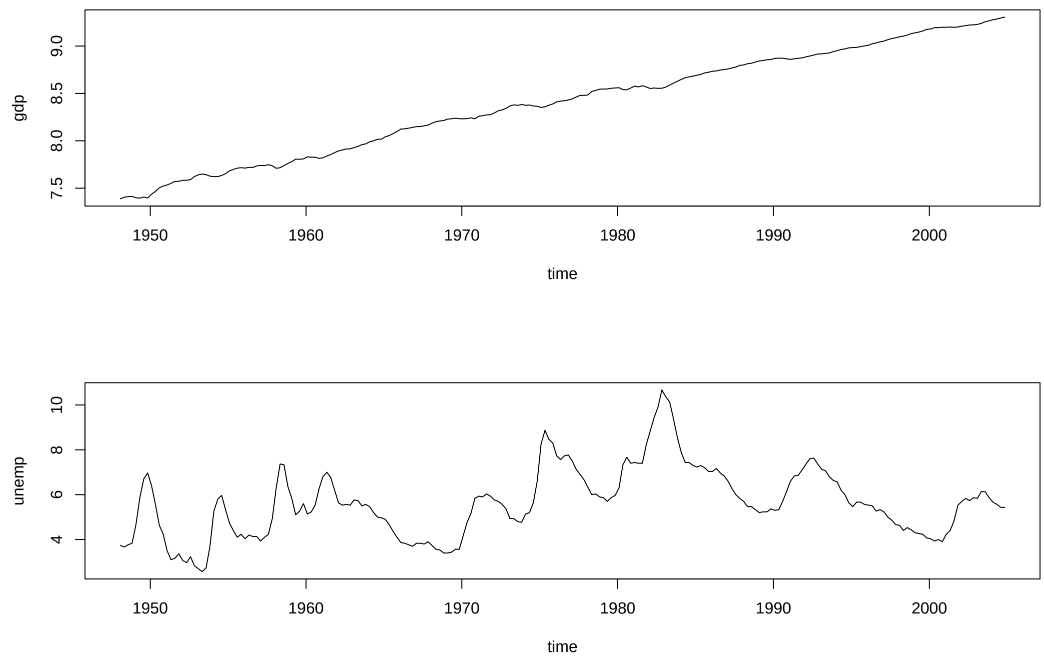
$$\boldsymbol{X}_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}$$

- Data: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$.
- Examples:
  (a) US quarterly GDP and unemployment rate. (See Fig 1.)
  (b) Daily closing prices of oil related ETFs, e.g., oil services holdings (OIH) and energy select section SPDR (XLE).
  (c) Quarterly GDP grow rates of Canada, United Kingdom, and United States.

Why consider two series jointly?

- model the relationship between the series;
- improve the accuracy of forecasts (use more information).

Figure 1: US Quarterly GDP and Unemployment (1948–2004).



Weak stationarity:

Both

$$\mathbb{E}(\boldsymbol{X}_t) = \begin{bmatrix} \mathbb{E}(x_{1,t}) \\ \mathbb{E}(x_{2,t}) \end{bmatrix} = \boldsymbol{\mu}$$

and

$$\begin{aligned} \text{Cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-h}) &= \begin{bmatrix} \text{Cov}(x_{1,t}, x_{1,t-h}) & \text{Cov}(x_{1,t}, x_{2,t-h}) \\ \text{Cov}(x_{2,t}, x_{1,t-h}) & \text{Cov}(x_{2,t}, x_{2,t-h}) \end{bmatrix} \\ &= \begin{bmatrix} \Gamma_{11}(h) & \Gamma_{12}(h) \\ \Gamma_{21}(h) & \Gamma_{22}(h) \end{bmatrix} = \Gamma_h \end{aligned}$$

are time invariant.
Note: Γ_h is not symmetric if $h \neq 0$. E.g., consider Γ_1:

- $\Gamma_{12}(1) = \text{Cov}(x_{1,t}, x_{2,t-1}) \implies x_{1,t}$ depends on past of $x_{2,t}$
- $\Gamma_{21}(1) = \text{Cov}(x_{2,t}, x_{1,t-1}) = \Gamma_{12}(-1) \implies x_{2,t}$ depends on past of $x_{1,t}$

Cross-Correlation matrix:

Define the diagonal matrix D

$$D = \begin{bmatrix} \sqrt{\mathbb{V}(x_{1,t})} & 0 \\ 0 & \sqrt{\mathbb{V}(x_{2,t})} \end{bmatrix} = \begin{bmatrix} \sqrt{\Gamma_{11}(0)} & 0 \\ 0 & \sqrt{\Gamma_{22}(0)} \end{bmatrix}$$

The cross-correlation matrix at lag h is then defined as:

$$\boldsymbol{\rho}_h = D^{-1}\Gamma_h D^{-1}$$

Thus $\rho_{ij}(h)$ is the cross-correlation between $x_{i,t}$ and $x_{j,t-h}$.
From stationarity:

$$\Gamma_h = \Gamma'_{-h} \qquad \text{and} \qquad \boldsymbol{\rho}_h = \boldsymbol{\rho}'_{-h}$$

E.g., $\text{cor}(x_{1,t}, x_{2,t-1}) = \text{cor}(x_{2,t}, x_{1,t+1})$.

Testing for serial dependence:

Multivariate version of Ljung-Box $Q(m)$ statistics are available to test:

$$H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \mathbf{0}, \qquad \text{vs.} \qquad H_a : \boldsymbol{\rho}_i = \mathbf{0}, \text{ for some } i.$$

For a series $\boldsymbol{x}_t$ of dimension k, use the test statistic:

$$Q(m) = T^2 \sum_{h=1}^{m} \frac{1}{T-h} \text{tr}\left(\widehat{\Gamma}'_h \widehat{\Gamma}_0^{-1} \widehat{\Gamma}_h \widehat{\Gamma}_0^{-1}\right) \quad \sim \quad \chi^2(mk^2), \quad \text{under } H_0.$$

Analysis with R

Use the package `MTS`. Some useful commands are:

(a) `MTSplot`: draws a multiple time series plot;

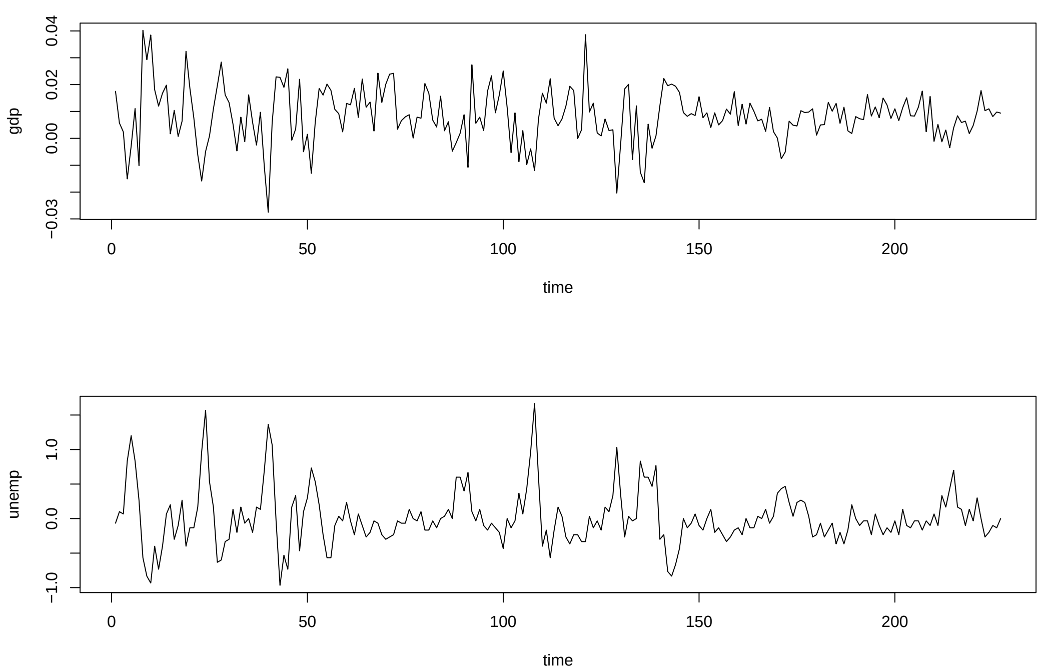
(a) `ccm`: computes the cross-correlation matrices;

(a) `mq`: computes the Ljung-Box statistics.

Example 1. *Consider the quarterly series of US GDP and unemployment data:*

```
> require(MTS)
> x=read.table("q-gdpun.txt",header=T)
> x[1,]
year mon day    gdp   unemp
1948   1   1 7.3878  3.7333
> z=x[,4:5]
> MTSplot(xt)
> ### Examine differenced series
> dz=diffM(z)
> MTSplot(dz)
> mq(zt,lag=5)
Ljung-Box Statistics:
        m       Q(m)     df    p-value
 [1,]     1       105       4        0
 [2,]     2       153       8        0
 [3,]     3       177      12        0
 [4,]     4       196      16        0
 [5,]     5       208      20        0
```

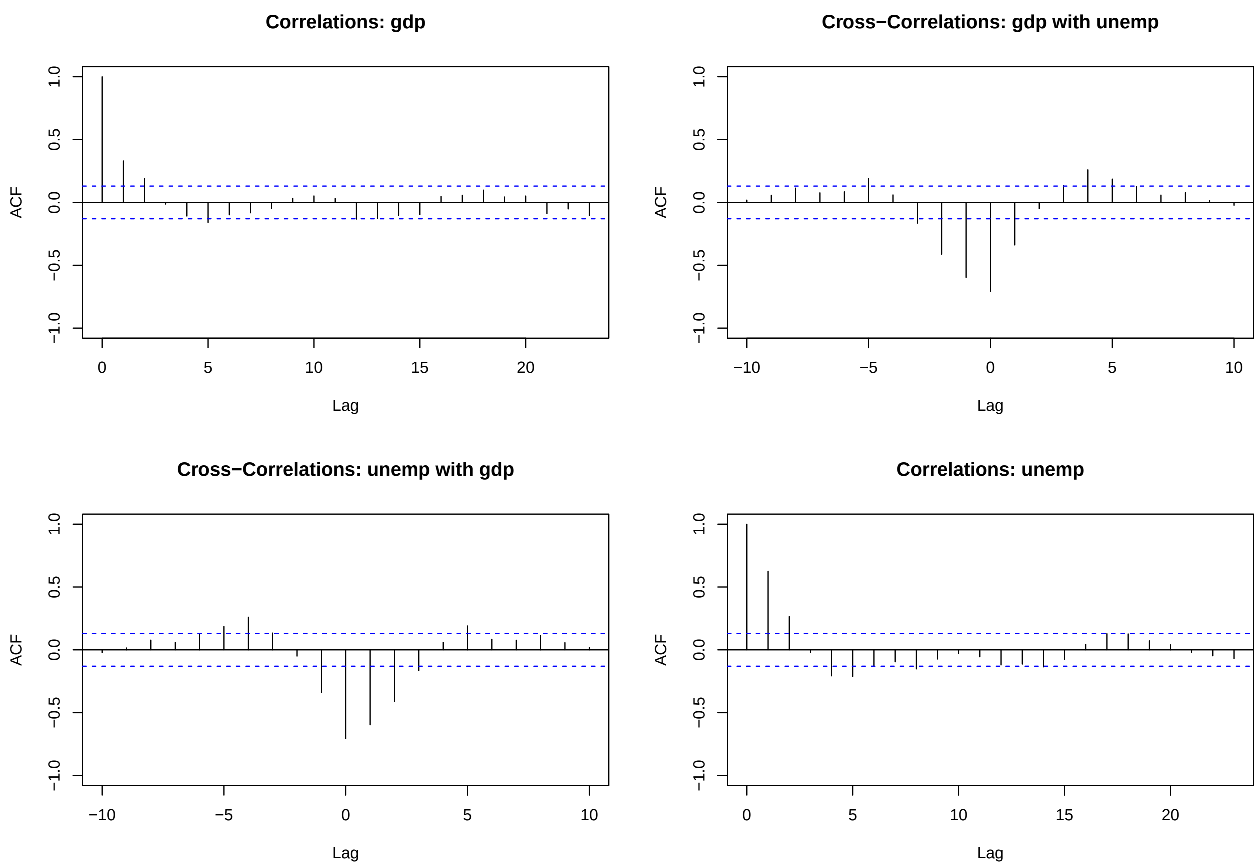
Figure 2: Differenced US Quarterly GDP and Unemployment (1948–2004).



The differenced series are plausibly stationary, and there is strong evidence of serial correlation.

```
##########################################################################
> ### Examine correlations
> ccm(dz)
##########################################################################
```

Figure 3: Autocorrelations and cross-correlations for the differenced GDP and Unemployment series.



The differenced series have rapidly decaying ACFs consistent with ARMAs, and are most strongly (negatively) cross-correlated at lag 0.

10.2 Vector AutoRegressive (VAR) models

VAR(1) model for two return series:

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \qquad \Longleftrightarrow \qquad \boldsymbol{r}_t = \boldsymbol{\phi}_0 + \Phi \boldsymbol{r}_{t-1} + \boldsymbol{a}_t$$

where $\boldsymbol{a}_t = (a_{1t}, a_{2t})'$ is a sequence of iid bivariate normal random vectors with mean zero and covariance matrix

$$\mathrm{Cov}(\boldsymbol{a}_t) = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \qquad \text{where } \sigma_{21} = \sigma_{12},$$

and is serially uncorrelated

$$\mathrm{Cov}(\boldsymbol{a}_t, \boldsymbol{a}_{t-h}) = 0, \qquad \text{for } h \neq 0.$$

That is, $\{\boldsymbol{a}_t\}$ is multivariate WN.

Granger causality

We can rewrite the VAR(1) model as:

$$\begin{aligned} r_{1t} &= \phi_{10} + \phi_{11} r_{1,t-1} + \phi_{12} r_{2,t-1} + a_{1,t} \\ r_{2t} &= \phi_{20} + \phi_{21} r_{1,t-1} + \phi_{22} r_{2,t-1} + a_{2,t} \end{aligned}$$

so that in general r_{1t} depends on lagged values of both r_{1t} and r_{2t}. But, if $\phi_{12} = 0$ and $\phi_{21} \neq 0$, then we have an example of a *Granger causality* relation:

- r_{1t} does not depend on $r_{2,t-1}$, but
- r_{2t} does not depend on $r_{1,t-1}$

implying that knowing $r_{1,t-1}$ is helpful in predicting $r_{2,t}$, but $r_{2,t-1}$ is not helpful in forecasting $r_{1,t}$. (We say that $r_{1,t}$ Granger causes $r_{2,t}$.)

Note: if $\sigma_{12} = 0$, then r_{1t} and r_{2t} are not concurrently correlated.

Stationarity condition

The VAR(1) model:

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \Phi \boldsymbol{r}_{t-1} + \boldsymbol{a}_t$$

is stationary if all zeros of the polynomial $|I - \Phi x|$ are greater than 1 in modulus, equivalently, if all eigenvalues of Φ are smaller than 1 in modulus:

- if x solves $|I - \Phi x| = 0$ then $|x| > 1$
- if λ is an eigenvalue of Φ then $|\lambda| < 1$

Moments

For the VAR(1) model:

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \Phi \boldsymbol{r}_{t-1} + \boldsymbol{a}_t$$

- The mean is $\boldsymbol{\mu} = (I - \Phi)^{-1}\boldsymbol{\phi}_0$, if the inverse exists.
- Using the MA(∞) representation

  $$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \sum_{i=0}^{\infty} \Phi^i \boldsymbol{a}_{t-i}$$

  the covariance function can be shown to be:

  $$\mathrm{Cov}(\boldsymbol{r}_t) = \sum_{i=0}^{\infty} \Phi^i \Sigma (\Phi^i)'$$

- This leads to the recursions (generalizes the AR(1)):

  $$\Gamma_h = \Phi \Gamma_{h-1}, \quad h > 0, \quad \text{which implies} \quad \Gamma_h = \Phi^h \Gamma_0$$

Building VAR Models

Let $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T$ be a sample from a model of dimension k.

(i) When $p > 0$ and $q > 0$ in a VARMA(p, q):

$$\boldsymbol{r}_t = \sum_{i=1}^{p} \Phi_i \boldsymbol{r}_{t-i} + \sum_{j=1}^{q} \Theta_j \boldsymbol{a}_{t-j} + \boldsymbol{a}_t$$

the parameters are not identifiable (from covariance of process); additional condititons are needed. For this reason one usually fits only a VAR(p) or a VMA(q), typically the former.

(ii) For a VAR(p), parameters can be estimated via either LSE or MLE. Fit models of increasing order: VAR(ℓ), $\ell = 1, 2, \ldots$, and let $\hat{\Sigma}_\ell$ be the estimated residual covariance matrix:

- Under LSE, a test of VAR($\ell - 1$) vs. VAR(ℓ) is:

  $$M(\ell) = -(T - k - \ell - 3/2) \log \left(|\hat{\Sigma}_\ell| / |\hat{\Sigma}_{\ell-1}| \right) \quad \dot{\sim} \quad \chi^2(k^2)$$

- Under MLE, can use AIC or BIC:

  $$\begin{aligned} AIC(\ell) &= \log |\hat{\Sigma}_\ell| + \frac{2k^2\ell}{T} \\ BIC(\ell) &= \log |\hat{\Sigma}_\ell| + \frac{k^2\ell \log(T)}{T} \end{aligned}$$

(iii) If $\{\boldsymbol{a}_t\} \sim$ iid WN, then LSEs and MLEs are both asymptotically normal with the same limiting distribution. Adequacy of fitted model can be tested by inspecting the residuals for serial correlation:

$$\hat{\boldsymbol{a}}_t = \boldsymbol{r}_t - \sum_{i=1}^{p} \hat{\Phi}_i \boldsymbol{r}_{t-i}$$

using the LB-statistic:

$$Q_k(m) = T^2 \sum_{i=1}^{m} \frac{1}{T-i} \text{tr}\left(\hat{A}_i' \hat{A}_0^{-1} \hat{A}_i \hat{A}_0^{-1}\right) \quad \dot{\sim} \quad \chi^2((m-p)k^2)$$

where

$$\hat{A}_i = \frac{1}{T} \sum_{t=i+1}^{T} \hat{\boldsymbol{a}}_t \hat{\boldsymbol{a}}_{t-i}'$$

Note: Need to adjust df in the R function `MTSdiag` by setting the argument `adj` $= d$, where d is the total number of non-zero coefficients in the fitted model (for a VAR(p) usually $d = pk^2$).

Impulse Response Function (IRF)

Similar to univariate case, can write a VAR(p) as VMA(∞):

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Psi_j \boldsymbol{a}_{t-j}, \qquad \Psi_0 = I$$

where the $\{\Psi_j\}$ are the IRF of $\boldsymbol{r}_t$, and Ψ_j is the effect of $\boldsymbol{a}_t$ on $\boldsymbol{r}_{t+j}$.

- **Problem:** interpretation of Ψ_j as the effect of $\boldsymbol{a}_t$ on $\boldsymbol{r}_{t+j}$ is confounded because the elements of $\boldsymbol{a}_t$ are serially correlated (recall $\mathbb{V}(\boldsymbol{a}_t) = \Sigma$).
- **Solution:** Choleski decompose $\Sigma = LGL'$, where L is lower triangular and G is diagonal, and define the orthogonal innovations:

$$\boldsymbol{b}_t = L^{-1}\boldsymbol{a}_t, \qquad \Longrightarrow \qquad \mathbb{V}(\boldsymbol{b}_t) = L^{-1}\Sigma(L^{-1})' = G$$

so that

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Psi_j L L^{-1} \boldsymbol{a}_{t-j} \equiv \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Psi_j^* \boldsymbol{b}_{t-j}$$

- **Interpretation:**

$$\Psi_\ell^*(i,j) = \text{impact of shock } \boldsymbol{b}_{j,t} \text{ on return } \boldsymbol{r}_{i,t+\ell}$$

(But: this interpretation is somewhat arbitrary; depends on the ordering of the elements of $\boldsymbol{r}_t \ldots$)

Example 2. *Consider monthly log returns of IBM stock ($r_{1,t}$) and the S&P 500 index ($r_{2,t}$). BIC suggests the VAR(1):*

$$\begin{aligned} r_{1t} &= 1.06 - 0.03 r_{1,t-1} + 0.15 r_{2,t-1} + a_{1,t} \\ r_{2t} &= 0.41 - 0.02 r_{1,t-1} + 0.10 r_{2,t-1} + a_{2,t} \end{aligned}$$

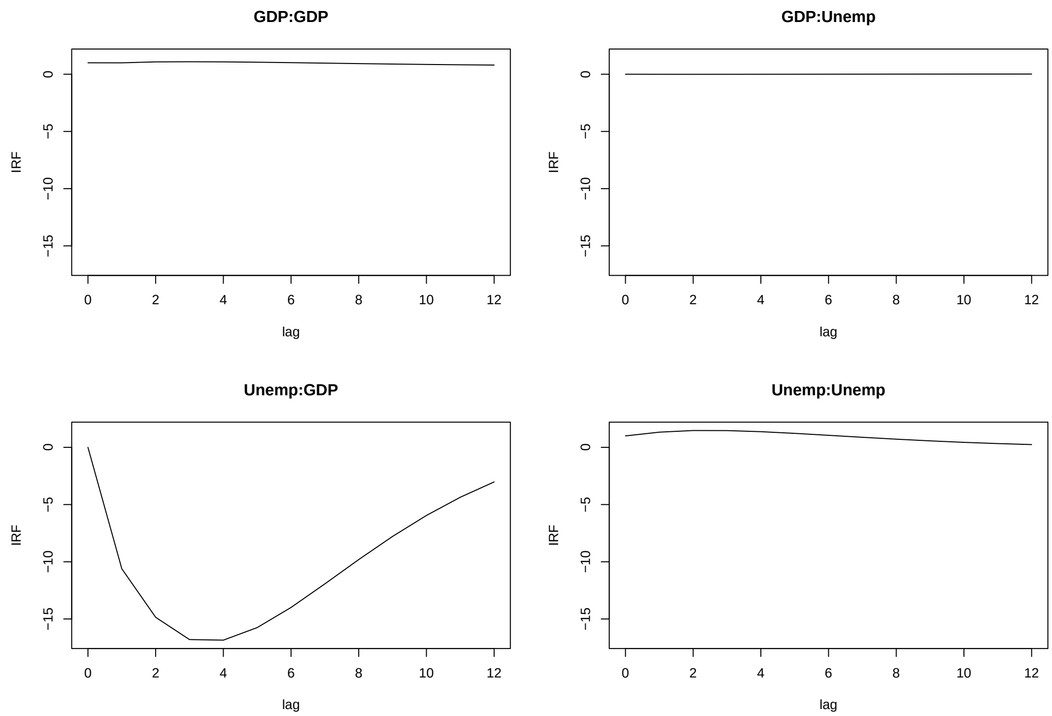
See Fig 8.7 in AFTS: since the dynamic dependence of the returns is weak, the IRFs exhibit simple patterns and decay quickly. But: the top right panel shows a peak at lag 1; thus a shock (impulse) in S&P 500 at time t has a maximum positive effect on IBM at time $t+1$.

Example 3. *Consider again the quarterly series of GDP and unemployment from Example 1. The following code Illustrates model fitting, checking, prediction, and IRFs.*

```
> require(MTS)
> x=read.table("Datasets/q-gdpun.txt",header=T)
> z=x[,4:5]

### plot the bivariate series
> MTSplot(z)

### find optimal VAR order
> VARorder(z, maxp = 13)
#selected order: aic =  4
#selected order: bic =  2
#selected order: hq =  3

### Fit VAR(2) and then refine it by omitting coeffs with t-ratio<1.96,
> m1=VAR(z, p = 2, output = T, include.mean = T)
> m2=refVAR(m1,thres=1.96)

### Check LB goodness of fit.
### Must adjust df by setting: adj=p*k^2 for a k-dim VAR(p).
### In general: adj is the number of non-zero VAR coeffts.
### Here: p=2=k implies adj=2^3=8, but one coef=0, so adj=7.
### All p-values low, so not a good fit...
> MTSdiag(m2, adj= 7)

### Forecast 3-steps ahead.
> VARpred(m2, h=3)

### Get irf's and plot them manually (canned plot hard to control...)
### Summary: 1 unit change in GDP makes Unemp dip 3-4 quarters later.
> out=VARMAirf(m2$Phi,m2$Sigma)
> psi11=out$irf[1,]; psi21=out$irf[2,]; psi12=out$irf[3,]; psi22=out$irf[4,]
#pdf(file="Plots/GDP-Unemp.pdf", pointsize=12, paper="a4r",width=0,height=0)
> par(mfcol=c(2,2))
> x=seq(0, length(psi11)-1); miny=min(out$irf); maxy=max(out$irf)
> plot(x,psi11, type="l", ylab="IRF", xlab="lag", main="GDP:GDP", ylim=c(miny,maxy))
> plot(x,psi21, type="l", ylab="IRF", xlab="lag", main="Unemp:GDP", ylim=c(miny,maxy))
> plot(x,psi12, type="l", ylab="IRF", xlab="lag", main="GDP:Unemp", ylim=c(miny,maxy))
> plot(x,psi22, type="l", ylab="IRF", xlab="lag", main="Unemp:Unemp", ylim=c(miny,maxy))
> dev.off()
######################################################################
```

Fig. 4 (lower left panel): an impulse (unit increase) in `GDP` *has a maximum (negative) effect in* `Unemp` *4 time periods (1 year) later.*

Figure 4: IRF plot for US Quarterly GDP and Unemployment (1948–2004).



10.3 Co-integration and VECM

Basic ideas:

- x_{1t} and x_{2t} are unit-root nonstationary;
- a linear combination of x_{1t} and x_{2t} is unit-root stationary.

That is, x_{1t} and x_{2t} share a single unit root!

Why is this of interest?

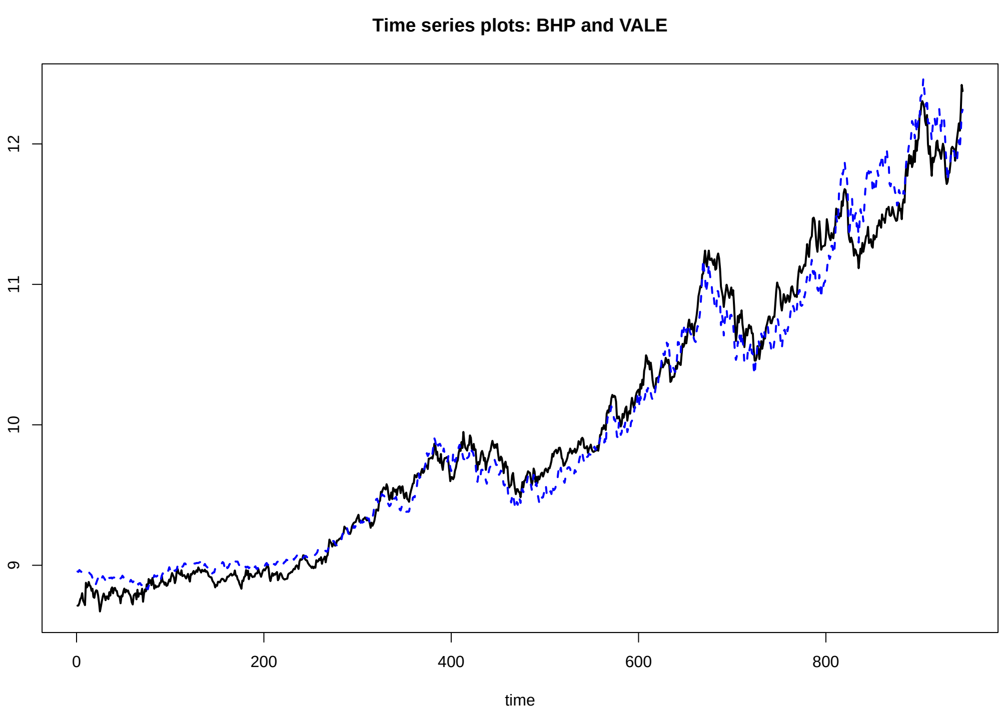
- A stationary series is mean reverting.
- Long term forecasts of the “linear” combination converge to a mean value, implying that the long-term forecasts of x_{1t} and x_{2t} must be linearly related.
- This mean-reverting property has many applications, e.g., pairs trading in finance.

Example 4. *Consider the daily adjusted closing stock prices of BHP Billiton Limited of Australia and Vale S.A. of Brazil. These are two natural resources companies. Both stocks are also listed in the New York Stock Exchange with tick symbols BHP and Vale, respectively. The sample period is from July 1, 2002 to March 31, 2006. (See Fig 5.)*

- *What can be said about the two prices? Is there any arbitrage opportunity between the two funds?*
- *Both series have a unit root (based on ADF test). Are they co-integrated?*

```
da=read.table("Datasets/d-bhp0206.txt",header=T)
da1=read.table("Datasets/d-vale0206.txt",header=T)
bhp=log(da[,9])
vale=log(da1[,9])
zt=10+scale(cbind(bhp,vale))
xt=seq(1:length(bhp))
plot(xt, zt[,1], lty=1, type="l", ylab="", xlab="time", lwd=2,
     main="Time series plots: BHP and VALE")
lines(xt, zt[,2], lty=2, col="blue", lwd=2)
```

Figure 5: Daily adjusted closing stock prices of BHP and Vale stocks (adjusted to have the same mean).



Co-integration tests

Several tests available, e.g., Johansen's test (Johansen, 1988). To get the basic idea, let $\Delta x_t = x_t - x_{t-1}$, and consider a univariate AR(2):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t$$

Subtract x_{t-1} from both sides and rearrange terms to obtain:

$$\Delta x_t = \gamma x_{t-1} + \phi_1^* \Delta x_{t-1} + a_t$$

where $\phi_1^* = -\phi_2$ and $\gamma = \phi_1 + \phi_2 - 1$.

Now, x_t is unit-root nonstationary if and only if $\gamma = 0$. Testing that x_t has a unit root is equivalent to testing that $\gamma = 0$ in the above model. This extends to general AR(p) models.

Turn now to the VAR(p) case:

$$\boldsymbol{X}_t = \Phi_1 \boldsymbol{X}_{t-1} + \cdots + \Phi_p \boldsymbol{X}_{t-p} + \boldsymbol{a}_t$$

Letting $Y_t = X_t - X_{t-1}$, subtract $\boldsymbol{X}_{t-1}$ from both sides and re-group the coefficient matrices to rewrite the model as:

$$\boldsymbol{Y}_t = \Pi \boldsymbol{X}_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \boldsymbol{Y}_{t-i} + \boldsymbol{a}_t \qquad (1)$$

where:

$$\begin{aligned}
\Phi_{p-1}^* &= -\Phi_p \\
\Phi_{p-2}^* &= -\Phi_{p-1} - \Phi_p \\
\vdots &= \vdots \\
\Phi_1^* &= -\Phi_2 - \cdots - \Phi_p \\
\Pi &= \Phi_p + \cdots + \Phi_1 - I
\end{aligned}$$

This is called a *Vector Error-Correction Model* (VECM).

Note: The matrix Π is a zero matrix if there is no co-integration.

Steps in testing for co-integration:

- Fit the model in equation (1),
- Test for the rank of Π.

If $\boldsymbol{X}_t$ is k-dimensional, and rank(Π) $= m$, then we have $k - m$ unit roots in $\boldsymbol{X}_t$, implying that there are m linear combinations of $\boldsymbol{X}_t$ that are unit-root stationary. In this case, we can write:

$$\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are full-rank $k \times m$ matrices. Furthermore:

- $\boldsymbol{W}_t = \boldsymbol{\beta}' \boldsymbol{X}_t$ is unit-root stationary,
- $\boldsymbol{\beta}$ is the co-integrating vector.

Co-integration Notes (* done on board ***)**

Implementation of co-integration tests (R package `urca`)

There are two types of co-integration test implemented in function **ca.jo** (both based on a likelihood ratio test idea):

- maximal eigenvalue (`type='eigen'`)
- trace (`type='trace'`)

The tests require one to specify whether the underlying VAR(p) has a "constant" or "trend" (drift) term:

$$\boldsymbol{X}_t = \boldsymbol{\mu}_t + \Phi_1 \boldsymbol{X}_{t-1} + \cdots + \Phi_p \boldsymbol{X}_{t-p} + \boldsymbol{a}_t$$

In function **ca.jo** , use following options:

- `ecdet='none'`: if $\boldsymbol{\mu}_t = \boldsymbol{0}$ (no constant)
- `ecdet='const'`: if $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ (constant)
- `ecdet='trend'`: if $\boldsymbol{\mu}_t = \boldsymbol{a} + \boldsymbol{b}t$ (linear trend/drift)

(Usage of `ecdet='const'` is generally recommended.)

Fitting and forecasting a VECM (R package `tsDyn`)

The most general form of a VECM is:

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \Pi \boldsymbol{X}_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \boldsymbol{Y}_{t-i} + \boldsymbol{a}_t$$

Function **VECM** can be used to fit this general VECM. Use the following options:

- `include='none'`: if $\boldsymbol{\mu}_t = \boldsymbol{0}$ (no constant)
- `include='const'`: if $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ (constant)
- `include='trend'`: if $\boldsymbol{\mu}_t = \boldsymbol{a} + \boldsymbol{b}t$ (linear trend/drift)

(There are also restricted forms of `'const'` and `'trend'`, but these don't seem to be implemented...)

After fitting, use function **forecast** to predict a VECM.

10.4 Pairs Trading

General idea: sell overvalued securities and buy undervalued ones, but the true value of the security is hard to determine in practice! Pairs trading attempts to resolve this difficulty by using relative pricing. Basically, if two securities have similar characteristics, then the prices of both securities must be more or less the same. Here the true price is not important.

Statistical concepts: the prices behave like a random-walk, but a linear combination of them is stationary (and hence mean-reversting). Deviations from the mean lead to trading opportunities.

Theory in Finance (Arbitrage Pricing Theory, or APT): If two securities have exactly the same risk factor exposures, then the expected returns of the two securities over a given time period are the same. (The key here is that the returns must be the same for all times.)

More details: Consider Stock 1 and Stock 2. Let $p_{i,t}$ be the log-price of Stock i at time t. It is reasonable to assume that both series $\{p_{1t}\}$ and $\{p_{2t}\}$ contain a unit root when analyzed individually.

- Assume that $\{p_{1t}\}$ and $\{p_{2t}\}$ are co-integrated, i.e., there exists a linear combination $c_1p_{1t} - c_2p_{2t}$ that is stationary. Dividing the linear combination by c_1, we have:

  $$w_t = p_{1t} - \gamma p_{2t}$$

  which is stationary, and hence mean-reverting.

- Now, form the portfolio Z by buying 1 share of Stock 1 and selling short on γ shares of Stock 2. The return of the portfolio over a time period h is

  $$\begin{aligned} r(h) &= (p_{1,t+h} - p_{1,t}) - \gamma(p_{2,t+h} - p_{2,t}) \\ &= p_{1,t+h} - \gamma p_{2,t+h} - (p_{1,t} - \gamma p_{2,t}) \\ &= w_{t+h} - w_t \end{aligned}$$

  which is the increment of the stationary series w_t from $t \mapsto t+h$. Since w_t is stationary, we have obtained a direct link of the portfolio to a stationary time series whose forecasts we can predict.

A trading strategy: Assume that $\mathbb{E}(w_t) = \mu$, select a threshold δ, and proceed as follows:

- Buy Stock 1 and short γ shares of Stock 2 when $w_t = \mu - \delta$.
- Unwind the position, i.e., sell Stock 1 and buy γ shares of Stock 2, when $w_{t+h} = \mu + \delta$.
- Profit: $r(h) = w_{t+h} - w_t = 2\delta$.

Practical considerations:

- Threshold δ is chosen so that the profit exceeds the cost of the two trades. In a high frequency regime, δ must be greater than the *trading slippage*, which is the same linear combination of bid-ask spreads of the two stocks, i.e.,

  $$(\text{bid-ask spread of Stock 1}) + \gamma(\text{bid-ask spread of Stock 2})$$

- The speed of mean-reversion of w_t plays an important role as h is directly related to the speed of mean-reversion.
- There are many possible ways to search for co-integrating pairs of stocks, e.g., via fundamentals, risk factors, etc.

Example 5. *Consider again the daily adjusted closing stock prices of BHP and Vale of Example 4. First regress one variable on the other to assess suitability for Pairs Trading:*

```
> x = cbind(bhp,vale)
> m1=lm(bhp~vale)
> summary(m1)

Call:
lm(formula = bhp ~ vale)

Residuals:
      Min        1Q    Median        3Q       Max
-0.151818 -0.028265  0.003121  0.029803  0.147105

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.822648   0.003662   497.7   <2e-16 ***
vale        0.716664   0.002354   304.4   <2e-16 ***
---
Residual standard error: 0.04421 on 944 degrees of freedom
Multiple R-squared: 0.9899,     Adjusted R-squared: 0.9899
F-statistic: 9.266e+04 on 1 and 944 DF,  p-value: < 2.2e-16
```

Now carry out Johansen's co-integration test (default is type "maximal eigenvalue"):

```
> library("urca"); library("tsDyn")
> var.order = ar(x)$order
> m2 = ca.jo(x, K=var.order, ecdet = "none")
> summary(m2)

######################
# Johansen-Procedure #
######################
Test type: maximal eigenvalue statistic (lambda max) , with linear trend
Eigenvalues (lambda):
[1] 0.0406019854 0.0000101517

Values of teststatistic and critical values of test:
          test 10pct  5pct  1pct
r <= 1 |  0.01  6.50  8.18 11.65
r = 0  | 39.13 12.91 14.90 19.19

Eigenvectors, normalised to first column:
(These are the cointegration relations)

           bhp.l2  vale.l2
bhp.l2   1.000000 1.000000
vale.l2 -0.717784 2.668019
```

```
Weights W:
(This is the loading matrix)

            bhp.l2       vale.l2
bhp.d  -0.06272119 -2.179372e-05
vale.d  0.03303036 -3.274248e-05
```

Conclude: there exists 1 cointegrating relation given by the 1st column of the (`bhp.l2`,`vale.l2`) matrix:

$$w_t = bhp - 0.717784(vale)$$

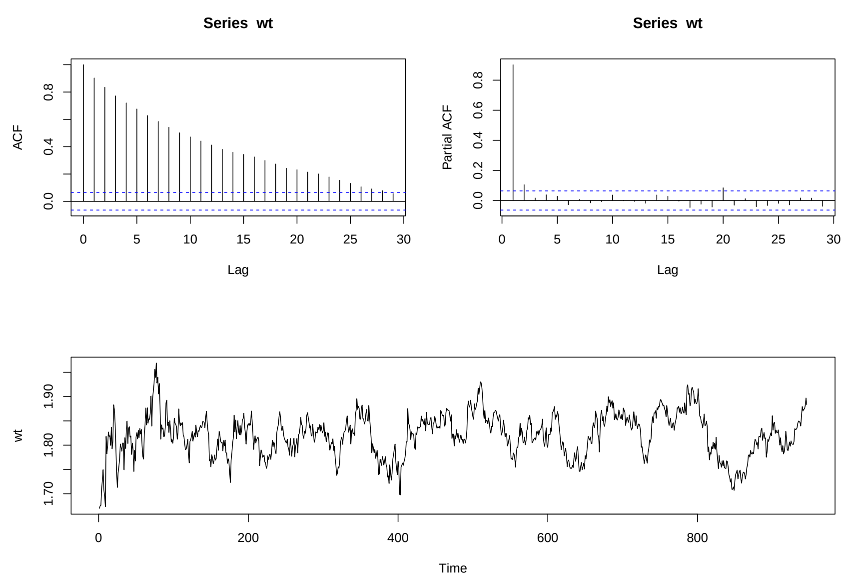
Repeating Johansen's co-integration test (with type "trace"), gives same conclusion:

```
> m3 = ca.jo(x, K=var.order, type=c("trace"))
> summary(m3)

######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.0406019854 0.0000101517

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  0.01  6.50  8.18 11.65
r = 0  | 39.14 15.66 17.95 23.52

Eigenvectors, normalised to first column:
(These are the cointegration relations)

           bhp.l2  vale.l2
bhp.l2   1.000000 1.000000
vale.l2 -0.717784 2.668019

Weights W:
(This is the loading matrix)

            bhp.l2       vale.l2
bhp.d  -0.06272119 -2.179372e-05
vale.d  0.03303036 -3.274248e-05
```

Inspection of the ACF/PACF of w_t suggests an AR(2) model (Fig 6):

```
### Plots
> wt=bhp-0.718*vale
> split.screen(figs=c(2,1))
> screen(1)
> split.screen(figs=c(1,2))
> screen(3)
> acf(wt)
> screen(4)
> pacf(wt)
> screen(2)
> ts.plot(wt)

### Fit AR(2)
> m4=arima(wt,order=c(2,0,0))

Coefficients:
         ar1     ar2  intercept
      0.8050  0.1215      1.820
s.e.  0.0323  0.0325      0.008

sigma^2 estimated as 0.000333:  log likelihood = 2444.26,  aic = -4880.52

### No problems detected in resids
> tsdiag(m4)
```

Figure 6: Time series plot and ACF/PACF for the cointegrated BHP and Vale series w_t.



*Now use **VECM** (package **tsDYn**) to estimate and predict the underlying VECM. Note:*

- *Set **lag=p-1** from the fitted VAR(p) (since $p = 2$ here, we should use **lag=1**).*
- ***r** is the number of cointegrating relations (**r**=1 here).*

```
> m5 = VECM(x, lag=1, r=1, estim="ML", include ="const")
> summary(m5)
#############
### Model VECM
#############
Full sample size: 946  End sample size: 944
Number of variables: 2  Number of estimated slope parameters 8
AIC -14875.77  BIC -14832.12  SSR 0.8188267
Cointegrating vector (estimated by ML):
   bhp      vale
r1   1 -0.717784

              ECT                 Intercept           bhp -1
Equation bhp  -0.0627 (0.0146)***  0.1159 (0.0266)***  -0.1149 (0.0367)**
Equation vale  0.0330 (0.0169).   -0.0584 (0.0308).     0.0528 (0.0425)
              vale -1
Equation bhp  0.0692 (0.0320)*
Equation vale 0.0452 (0.0371)

### predict fitted VECM 3 steps ahead
> predict(m5, n.ahead=3)
         bhp     vale
947 3.650391 2.465768
948 3.648675 2.469617
949 3.647362 2.473280
```

Let $\boldsymbol{x}_t = (bhp, vale)'$, and recall that $\Delta \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$. The fitted VECM is:

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \Phi_1^* \Delta \boldsymbol{x}_{t-1} + \boldsymbol{a}_t$$

where

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 0.1159 \\ -0.0584 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} -0.0627 \\ 0.0330 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} 1 \\ -0.7179 \end{bmatrix}, \quad \Phi_1^* = \begin{bmatrix} -0.1149 & 0.0692 \\ 0.0528 & 0.0452 \end{bmatrix}$$

and the cointegrating relation is: $w_t = \boldsymbol{\beta}'\boldsymbol{x}_t = (1, -0.7179)\boldsymbol{x}_t$.