Covariance-Stationary Vector Processes

(Hamilton: Chapters 10-11, Tsay: Chapter 8)

Isabel Casas

- Introduction
- Vector Autoregressive of order p
- Maximum Likelihood Estimation for VAR(p)
- Granger's Causality

SYDDANSK UNIVERSITET

Introduction

The univarite AR(p) process

$$y_t = c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t \qquad \epsilon_t \sim IID(0, \sigma^2)$$

explains the interaction of y_t with its previous p lags.

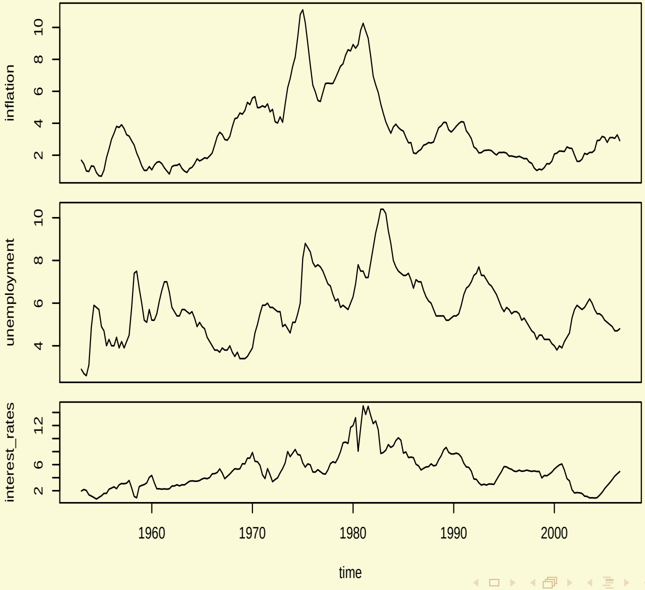
Q: Might y_t depend not only on its own lags but on the lags of other variables?

Q: Should we model the co-movements of sereval variables?

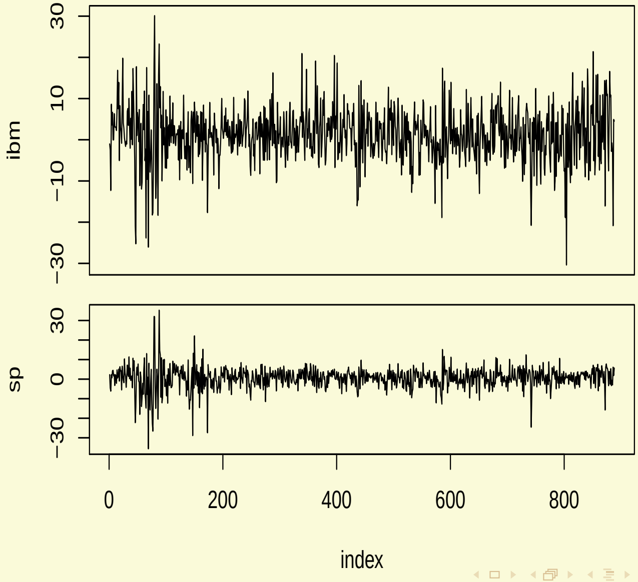
Introduction

- Financial markets are dependant of each other and knowing the financial market are interrelated is very important
- Consumption and income
- Stock prices and dividends
- Forward and spot exchange rates
- Interest rates, money growth, income, inflation

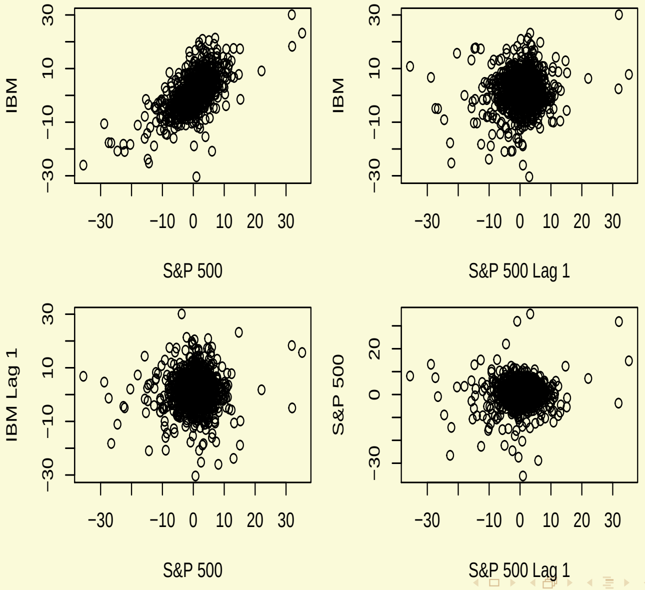
Examples


inflation
2
4
6
8
10
unemployment
4
6
8
10
interest_rates
2
6
12
1960
1970
1980
1990
2000
time

Examples



Examples


IBM
S&P 500
IBM
S&P 500 Lag 1
IBM Lag 1
S&P 500
S&P 500
S&P 500 Lag 1
30
10
−10
−30
20
0
−30
−10
0
10
20
30

Examples

- There is correlation (linear dependency) between the two series at time t (concurrent correlation)
- The cross correlation at lag 1 is weak
- Can we obtain the sample cross correlation matrix (CCM)?

pth order Vector Autoregression: VAR(p)

Consider n time series variables $\mathbf{y_t} = (y_{1t}, \ldots, y_{nt})'$.

For example $n = 2$ where $\mathbf{y_t} = (y_{1t}, y_{2t})'$ where y_{1t} might be the log of the GNP in year t and y_{2t} is the interest rate paid on Treasury bills in year t.

The VAR(p) model:

$$\mathbf{y_t} = \mathbf{c} + \mathbf{\Phi_1 y_{t-1}} + \ldots + \mathbf{\Phi}_p \mathbf{y_{t-p}} + \boldsymbol{\epsilon}_t$$

Note: The bold notation means that the object is a vector or a matrix.

VAR(p)

$$
\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} + \begin{pmatrix} \phi_{11}^1 & \dots & \phi_{1n}^1 \\ \phi_{21}^1 & \dots & \phi_{2n}^1 \\ \vdots & & \vdots \\ \phi_{n1}^1 & \dots & \phi_{nn}^1 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{n,t-1} \end{pmatrix} + \dots
$$

$$
+ \begin{pmatrix} \phi_{11}^p & \phi_{12}^p & \dots & \phi_{1n}^p \\ \vdots & & \vdots & \\ \phi_{21}^p & \phi_{22}^p & \dots & \phi_{nn}^p \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{n,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{nt} \end{pmatrix}
$$

$$
\boldsymbol{\epsilon}_t \sim IID_n(\mathbf{0}, \boldsymbol{\Omega}) \qquad \boldsymbol{\Omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}
$$

VAR(p)

So, the first equation is:

$$
\begin{aligned}
y_{1t} = & c_1 + \phi^1_{11} y_{1,t-1} + \phi^1_{12} y_{2,t-1} + \ldots + \phi^1_{1n} y_{n,t-1} \\
& + \phi^2_{11} y_{1,t-2} + \phi^2_{12} y_{2,t-2} + \ldots + \phi^2_{1n} y_{n,t-2} \\
& \ldots \\
& + \phi^p_{11} y_{1,t-p} + \phi^p_{12} y_{2,t-p} + \ldots + \phi^p_{1n} y_{n,t-p} \\
& + \epsilon_{1t}
\end{aligned}
$$

Q1: Is there simultaneity?

Q2: How is the second equation?

VAR(p)

We can use the operator notation with matrices:

$$[\mathbf{I_n} - \mathbf{\Phi_1 L} - \mathbf{\Phi_2 L^2} - \ldots - \mathbf{\Phi_p L^p}]\mathbf{y_t} = \mathbf{c} + \boldsymbol{\epsilon_t}$$

or

$$\mathbf{\Phi(L)y_t} = \mathbf{c} + \boldsymbol{\epsilon_t}$$

where $\mathbf{\Phi}(L)$ is the $n \times n$ matrix polynomial in the lag operator:

Covariance-stationarity of VAR(p)

Q: Do you remember when an AR(p) process was weakly stationary?

Covariance-stationarity of VAR(p)

Q: Do you remember when an AR(p) process was weakly stationary?

A1: First and second moments are all independent of t

A2: The polynomial $1 - \phi(z)$ have roots outside the inner circle.

A3: Eigenvalues from $det(\mathbf{I}\lambda - \mathbf{F}) = 0$ are inside the inner circle.

Covariance-stationarity of VAR(p)

Q: Do you remember when an AR(p) process was weakly stationary?

A1: First and second moments are all independent of t

A2: The polynomial $1 - \phi(z)$ have roots outside the inner circle.

A3: Eigenvalues from $det(\mathbf{I}\lambda - \mathbf{F}) = 0$ are inside the inner circle.

The same applies to the VAR(p):

1. $\boldsymbol{\mu} = \mathbf{c} + \boldsymbol{\Phi}_1\boldsymbol{\mu} + \boldsymbol{\Phi}_2\boldsymbol{\mu} + \ldots + \boldsymbol{\Phi}_p\boldsymbol{\mu}$
2. The second moment:

$$\boldsymbol{\Gamma}_j = E((\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-j} - \boldsymbol{\mu})')$$

SYDDANSK UNIVERSITET

Second moment of VAR(p)

$$Var(\mathbf{y}_t) = \mathbf{\Gamma}_0 = [\gamma_{ij}^0] = \begin{pmatrix} Var(y_{1t}) & Cov(y_{1t}, y_{2t}) & \dots & Cov(y_{1t}, y_{nt}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(y_{nt}, y_{1t}) & Cov(y_{nt}, y_{2t}) & \dots & Var(y_{nt}) \end{pmatrix}$$

The correlation matrix of $\mathbf{y}_t$

$$Corr(\mathbf{y}_t) = \mathbf{R}_0 = \mathbf{D}^{-1}\mathbf{\Gamma_0}\mathbf{D}^{-1}$$

where $\mathbf{D}$ is a $n \times n$ diagonal matrix with $\sqrt{\gamma_{ii}^0}$ as elements.

VAR(p) moments estimated from the data

Ergodicity:

Covariance stationary $\Rightarrow$ ergodic in the mean

$$\overline{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t \rightarrow^p E(\mathbf{y}_t) = \boldsymbol{\mu}$$

Stationary Gaussian proccess $\Rightarrow$ ergodic in the second moment

$$\hat{\mathbf{\Gamma}}_{\mathbf{0}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t - \overline{\mathbf{y}})(\mathbf{y}_t - \overline{\mathbf{y}})' \rightarrow^p Var(\mathbf{y}_t) = \mathbf{\Gamma}_0$$

$$\hat{\mathbf{R}}_{\mathbf{0}} = \hat{\mathbf{D}}^{-\mathbf{1}}\hat{\mathbf{\Gamma}}_{\mathbf{0}}\hat{\mathbf{D}}^{-\mathbf{1}} \rightarrow^p Corr(\mathbf{y}_t) = \mathbf{R}_0$$

Cross covariance and correlation matrices

- $y_{1t}, y_{2t}, \ldots, y_{nt}$, each have each own autocovariances and autocorrelation matrices.
- There is also a cross lead-lag covariances and correlations between all possible pairs of components.

$$\begin{aligned}
\boldsymbol{\Gamma}_k =& [\gamma_{ij}^k] = E((\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-k} - \boldsymbol{\mu})') \\
\mathbf{R_k} =& [\rho_{ij}^k] = \mathbf{D^{-1}\Gamma_k D^{-1}} \\
\rho_{ij}^k =& \frac{\gamma_{ij}^k}{\sqrt{\gamma_{ii}^0 \gamma_{jj}^0}} = \frac{Cov(y_{it}, y_{j,t-k})}{std(y_{it})std(y_{jt})}
\end{aligned}$$

SYDDANSK UNIVERSITET

Cross covariance and correlation matrices

- When $k > 0$, this correlation measures the linear dependency of y_{it} on $y_{j,t-k}$
- If $\rho_{ij}^{k} \neq 0$ and $k > 0$ we say that the series y_{jt} leads the series y_{it} at lag k
- ρ_{ii}^{k} is the lag-k autocorrelation coefficient of y_{it}
- $\rho_{ij}^{k} \neq \rho_{ji}^{k}$ for $i \neq j$. The two correlations measure two different linear dependencies.
- So $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are generally non-symmetric
- However $\rho_{ij}^{k} = \rho_{ji}^{-k}$

What does it mean $\rho_{ij}^{k} = \rho_{ji}^{-k}$ but $\rho_{ij}^{k} \neq \rho_{ji}^{k}$ for $i \neq j$?

State Space form of VAR(p)

It can be written as a VAR(1): $\boldsymbol{\xi}_t = F\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$

$$\boldsymbol{\xi}_t = \begin{pmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p+1} - \boldsymbol{\mu} \end{pmatrix}$$

$$\mathbf{F} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_n & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & & & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_n & \mathbf{0} \end{pmatrix}$$

$$\mathbf{v}_t = \begin{pmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

Covariance-stationarity of VAR(p)

$$\boldsymbol{\xi}_{t+s} = \mathbf{v}_{t+s} + \mathbf{F}\mathbf{v}_{t+s-1} + \mathbf{F}^2\mathbf{v}_{t+s-2} + \ldots + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^s\boldsymbol{\xi}_t$$

The VAR is weakly stationary if:

1. The eigenvalues of $\mathbf{F}$ all lie inside the unit circle:

$$\det(\mathbf{I}_n\lambda^p - \boldsymbol{\Phi}_1\lambda^{p-1} - \ldots - \boldsymbol{\Phi}_p) = 0$$

or

2. The roots of

$$(\mathbf{I}_n - \boldsymbol{\Phi}_1 z - \ldots - \boldsymbol{\Phi}_p z^p) = 0$$

lie outside the unit circle.

Example: VAR(1), $n = 2$

For example assume that y_{1t} correspond to the IBM log returns and y_{2t} correspond to the S&P 500 log returns.

$$\begin{aligned} y_{1t} =& c_{10} + \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \epsilon_{1t} \\ y_{2t} =& c_{20} + \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \epsilon_{2t} \\ \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim& N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right) \end{aligned}$$

Because $\boldsymbol{\Omega}$ is positive definite, it can be decomposed $\boldsymbol{\Omega} = \mathbf{LGL}'$ where $\boldsymbol{L}$ is lower triangular with unit diagonal elements and $\boldsymbol{G}$ is a diagonal matrix.

Example: VAR(1), $n = 2$

- ϕ_{12} denotes the linear dependence of y_{1t} on $y_{2,t-1}$ in the presence of $y_{1,t-1}$
- ϕ_{12} is the conditional effect of $y_{2,t-1}$ on y_{1t} given $y_{1,t-1}$
- Q: What does it mean $\phi_{12} = 0$? and $\phi_{21} = 0$?
- If $\phi_{12} = \phi_{21} = 0$ then y_{1t} and y_{2t} are uncoupled
- If $\phi_{12} \neq 0$ and $\phi_{21} \neq 0$ then there is a feedback relationship between the two series
- $\sigma_{12} = \sigma_{21}$ and measures the *concurrent correlation* between y_{1t} and y_{2t} (Γ_0)
- If $\sigma_{12} = 0$ then there is no current linear relationship between the two component series

Example: VAR(1), $n = 2$

Structural equation: the two equation system is converted into one equation:
The reduced form:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

The structural form:

$$\begin{aligned} \mathbf{L^{-1}} \mathbf{y_t} =& \mathbf{L^{-1}} \mathbf{c} + \mathbf{L^{-1}} \mathbf{\Phi}_1 \mathbf{y_{t-1}} + \mathbf{L^{-1}} \boldsymbol{\epsilon}_t \\ =& \mathbf{c}^* + \mathbf{\Phi_1^*} \mathbf{y}_{t-1} + \boldsymbol{\eta}_t \end{aligned}$$

Example: VAR(1), $n = 2$

For example:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.4 \end{pmatrix} + \underbrace{\begin{pmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{pmatrix}}_{\Phi} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad \mathbf{\Omega} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

$$L = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix} \quad L^{-1} = \begin{pmatrix} 1 & 0 \\ -0.5 & 1 \end{pmatrix}$$

In fact,

$$\Omega = \begin{pmatrix} 1 & 0 \\ -0.5 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -0.5 & 1 \end{pmatrix}'$$

Example: VAR(1), $n = 2$

Multiplying the equation by L^{-1} we get the structural VAR which shows explicitly the concurrent linear dependence of y_{2t} on y_{1t}

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0.5y_{1t} \end{pmatrix} + \begin{pmatrix} 0.2 & 0.3 \\ -0.7 & 0.95 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

$$\mathbf{\Omega}_u = \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix}$$

Orthogonal error.

Structural VAR (SVAR)

```
> matrix(c(2,1,1,1), nrow=2, byrow=T)->Omega
> library(bdsmatrix)
> L=gchol(Omega)
> G= diag(c(diag(L)))
> L=as.matrix(L)
> L

     [,1] [,2]
[1,]  1.0    0
[2,]  0.5    1

> G

     [,1] [,2]
[1,]    2  0.0
[2,]    0  0.5
```

Structural VAR (SVAR)

```
> c= c(0.2, 0.4)
> Phi=matrix(c(0.2, 0.3, -0.6, 1.1), nrow=2, byrow=T)
> c.star= solve(L)%*%c
> Phi.star= solve(L)%*%Phi
> c.star
     [,1]
[1,]  0.2
[2,]  0.3
> Phi.star
     [,1] [,2]
[1,]  0.2 0.30
[2,] -0.7 0.95
```

Structural VAR (SVAR)

```
> Sigma.u = solve(L)%*%Omega%*%solve(t(L))
> Sigma.u
     [,1] [,2]
[1,]    2  0.0
[2,]    0  0.5
```

Example: VAR(1), $n = 2$

Stationarity:

$$det(\lambda \mathbf{I}_2 - \mathbf{\Phi}) = det \begin{pmatrix} \lambda - 0.2 & -0.3 \\ 0.6 & \lambda - 1.1 \end{pmatrix} = 0$$

$$\lambda^2 - 1.3\lambda + 0.4 = 0$$

$$\lambda_1 = 0.8 \qquad \lambda_2 = 0.5$$

Q: Is it stationary?

Exercise (5minutes)

Find roots z:

$$det(\mathbf{I}_2 - \mathbf{\Phi} z) = 0$$

Q: Is it stationary?

Important issues of a stationary VAR

- If the process is stationary, then the ϵ_{it} are uncorrelated with $\mathbf{y}_{i,t-1}, \dots, \mathbf{y}_{i,t-p}$
- Endogeneity is avoided by using lagged values of $y_{1t}, y_{2t}, \dots, y_{nt}$
- The VAR(p) model is just a semingly unrelated regression (SUR) model with lagged variables and deterministic terms as common regressors
- Then, parameters of a vector autoregression can be estimated consistently with n OLS regresssions, one for each equation.

Q: What estimator you think will be more efficient?

Conditional maximum likelihood for a VAR

- Let $\mathbf{y}_t$ denote a $n \times 1$ vector containing the n variables at time t.

$$\mathbf{y_t} = \mathbf{c} + \boldsymbol{\phi_1} \mathbf{y_{t-1}} + \ldots + \boldsymbol{\phi_p} \mathbf{y_{t-p}} + \boldsymbol{\epsilon_t} \qquad \boldsymbol{\epsilon}_t \sim IIDN(\mathbf{0}, \boldsymbol{\Omega})$$

- Suppose we have observed each of these n variables for $T + p$ time periods
- The initial values are $\mathbf{y_{-p+1}}, \mathbf{y_{-p+2}}, \ldots, \mathbf{y_0}$ and to base the estimation on the last T observations $\mathbf{y_1}, \mathbf{y_1}, \ldots, \mathbf{y_T}$
- Want to find the conditional likelihood function

$$f_{\mathbf{Y_T},\ldots\mathbf{Y_1}|\mathbf{Y_0},\ldots\mathbf{Y_{-p+1}}}(\mathbf{y_T}, \ldots, \mathbf{y_1} | \mathbf{y_0}, \mathbf{y_{-1}}, \ldots, \mathbf{y_{-p+1}}; \boldsymbol{\theta})$$

SYDDANSK UNIVERSITET

Conditional maximum likelihood for a VAR

- Then, we maximise the conditional likelihood function with respect to $\boldsymbol{\theta}$ which contains $\mathbf{c}, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p$ and $\boldsymbol{\Omega}$.
- Because $\epsilon_t \sim IIDN(\mathbf{0}, \boldsymbol{\Omega})$ then

$$\mathbf{y_t} | \mathbf{y_0}, \mathbf{y_{-1}}, \ldots, \mathbf{y_{-p+1}} \sim N(\mathbf{c} + \boldsymbol{\phi_1} \mathbf{y}_{t-1} + \ldots + \boldsymbol{\phi_p} \mathbf{y}_{t-p}, \boldsymbol{\Omega})$$

- If we define $\mathbf{x_t} = (\mathbf{1}, \mathbf{y_{t-1}}, \ldots, \mathbf{y_{t-p}})'$ and $\boldsymbol{\Pi}' = (\mathbf{c}, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$ then, the conditional distribution can be written as

$$\mathbf{y_t} | \mathbf{y_0}, \mathbf{y_{-1}}, \ldots, \mathbf{y_{-p+1}} \sim N(\boldsymbol{\Pi}' \mathbf{x}_t, \boldsymbol{\Omega})$$

Q: We know the conditional density of a normal, don't we?

Conditional maximum likelihood for a VAR

- The conditional likelihood function of our sample is:

$$f_{\mathbf{Y_T},\ldots,\mathbf{Y_1}|\mathbf{Y_0},\ldots\mathbf{Y_{-p+1}}}(\mathbf{y_T},\ldots,\mathbf{y_1}|\mathbf{y_0},\mathbf{y_{-1}},\ldots,\mathbf{y_{-p+1}};\boldsymbol{\theta}) = \prod_{t=1}^{T} f_{\mathbf{Y_t}|\mathbf{Y_{t-1}},\ldots,\mathbf{Y_{-p+1}}}(\mathbf{y_t}|\mathbf{y_{t-1}},\ldots,\mathbf{y_{-p+1}};\boldsymbol{\theta})$$

- Taking logs:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \log f_{\mathbf{Y_t}|\mathbf{Y_{t-1}},\ldots,\mathbf{Y_{-p+1}}}(\mathbf{y_t}|\mathbf{y_{t-1}},\ldots,\mathbf{y_{-p+1}};\boldsymbol{\theta}) \\ &= \frac{-Tn}{2}\log(2\pi) + \frac{T}{2}\log|\boldsymbol{\Omega}^{-1}| \\ &\quad - \frac{1}{2}\sum_{t=1}^{T}\left[(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)\right]\end{aligned}$$

MLE of $\mathbf{\Pi}$

Assuming normality of the innovations and uncorrelation, the MLE of $\mathbf{\Pi}$ which contains $\mathbf{c}$ and $\mathbf{\Phi}_j$ is:

$$\hat{\mathbf{\Pi}}' = \left[\sum_{t=1}^{T} \mathbf{y_t x_t'}\right] \left[\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x_t x_t'}\right]^{-1}$$

In particular the jth row, corresponding to the jth equation is:

$$\hat{\boldsymbol{\pi}}_j{}' = \left[\sum_{t=1}^{T} \mathbf{y_{jt} x_t'}\right] \left[\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x_t x_t'}\right]^{-1}$$

Q: Does this last equation reminds you of something?

MLE of $\mathbf{\Omega}$

Substituting $\hat{\mathbf{\Pi}}$ into the conditional log likelihood function, we obtain the following:

$$\mathcal{L}(\mathbf{\Omega}, \hat{\mathbf{\Pi}}) = \frac{-Tn}{2} \log(2\pi) + \frac{T}{2} \log |\mathbf{\Omega}^{-1}| - \frac{1}{2} \sum_{t=1}^{T} \hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_t'$$

The estimator:

$$\hat{\mathbf{\Omega}} = \frac{1}{T} \sum_{i=1}^{T} \hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_t'$$

The ij-element of $\hat{\mathbf{\Omega}}$:

$$\sigma_{ij} = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_{it} \hat{\epsilon}_{jt}$$

VAR in R

Let us simulate $t = 250$ observations of:

$$
\begin{aligned}
y_{1t} =& -0.7 + 0.7 y_{1,t-1} + 0.2 y_{2,t-1} + \epsilon_{1,t} \\
y_{2t} =& 1.3 + 0.2 y_{1,t-1} + 0.7 y_{2,t-1} + \epsilon_{2,t}
\end{aligned}
$$

Therefore:

$$
\mathbf{\Phi} = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} -0.7 \\ 1.3 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}, \mathbf{\Omega} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}
$$

```
> Phi1= matrix(c(0.7, 0.2, 0.2, 0.7), nrow=2, ncol=2)
> mu.vec= c(1,5)
> c.vec= as.vector((diag(2)- Phi1) %*% mu.vec)
> Omega= matrix(c(1, 0.5, 0.5, 1), 2,2)
> library(vars)
> eigen(Phi1)$values
[1] 0.9 0.5
```

VAR in R

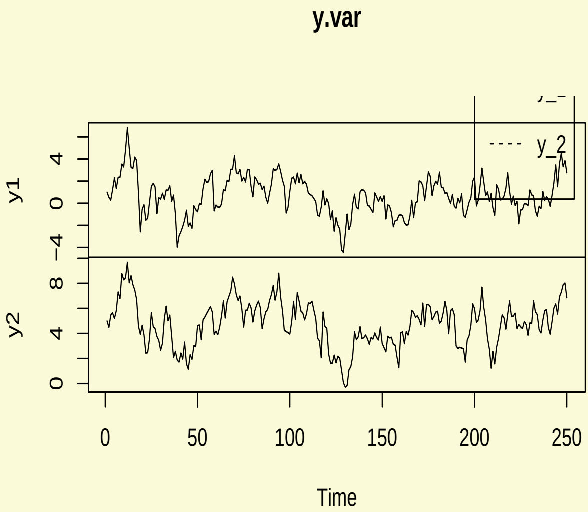
```
> library(tseries)
> library(MASS)
> set.seed(42)
> T = 250
> nvar= 2 # 2 variables
> y.var = matrix(0,nvar,T)
> y.var[,1] = mu.vec
> e.var =  mvrnorm(T,mu= rep(0, nvar), Sigma=Omega)
> e.var = t(e.var)
> for (i in 2:T) {
+       y.var[,i] = c.vec+Phi1%*%y.var[,i-1]+e.var[,i]
+ }
> y.var = t(y.var)
> dimnames(y.var) = list(NULL,c("y1","y2"))
> eigen(Phi1)$values

[1] 0.9 0.5

> colMeans(y.var)

       y1        y2
0.6757314 4.7459675
```

VAR in R



VAR in R

Let us estimate it:

```
> y1=VAR(y.var, lag.max= 4, ic="AIC")
> y1

VAR Estimation Results:
=======================

Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const

     y1.l1      y2.l1      const
 0.5932973  0.2486591 -0.8995433


Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + const

    y1.l1     y2.l1     const
0.1664443 0.7229253 1.2090046

>
```

VAR in R

Let us estimate it:

```
> y2=VAR(y.var, p=2)
> y2

VAR Estimation Results:
=======================

Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + y1.l2 + y2.l2 + const

     y1.l1      y2.l1      y1.l2      y2.l2      const
0.63052937 0.22582772 -0.07910579 0.05884478 -1.04080425


Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + y1.l2 + y2.l2 + const

     y1.l1      y2.l1      y1.l2      y2.l2      const
0.18483169 0.71743728 -0.03641378 0.02010926 1.15361456
```

VAR in R

Forecast:

```
> predict(y1, n.ahead= 10)

$y1
          fcst      lower    upper       CI
 [1,] 2.440713  0.4864977 4.394929 1.954215
 [2,] 2.195191 -0.2368253 4.627206 2.432016
 [3,] 1.994919 -0.7151820 4.705021 2.710101
 [4,] 1.826463 -1.0714308 4.724357 2.897894
 [5,] 1.682347 -1.3492984 4.713992 3.031645
 [6,] 1.557938 -1.5712537 4.687130 3.129192
 [7,] 1.450036 -1.7512194 4.651292 3.201256
 [8,] 1.356225 -1.8986983 4.611149 3.254924
 [9,] 1.274563 -2.0205449 4.569671 3.295108
[10,] 1.203431 -2.1218800 4.528742 3.325311

$y2
          fcst    lower    upper       CI
 [1,] 6.622180 4.766145 8.478216 1.856035
 [2,] 6.402589 3.996305 8.808874 2.406285
 [3,] 6.202975 3.460656 8.945295 2.742319
 [4,] 6.025335 3.054596 8.996075 2.970740
 [5,] 5.868876 2.735960 9.001793 3.132916
 [6,] 5.731781 2.481087 8.982475 3.250694
 [7,] 5.611964 2.274567 8.949361 3.337397
 [8,] 5.507386 2.105591 8.909180 3.401794
 [9,] 5.416169 1.966252 8.866086 3.449917
[10,] 5.336634 1.850600 8.822668 3.486034
```

Asymptotic distribution of $\mathbf{\Pi}$

If the following conditions are satisfied:

- Independent innovations,
- Bounded fourth-cumulant of the innovations, and
- Roots of the polynomial outside the unit circle

We name $\hat{\boldsymbol{\pi}}_{\boldsymbol{T}} = vec(\hat{\mathbf{\Pi}}_{\boldsymbol{T}})$ which are the estimator of the coefficients using a sample of size T. Then,

Asymptotic distribution of $\hat{\mathbf{\Pi}}$

1. $\frac{1}{T}\sum_{t=1}^{T} \mathbf{x_t x_t'} \rightarrow^P \mathbf{Q}$ where $Q = E(\mathbf{x_t x_t'})$
2. $\hat{\boldsymbol{\pi}}_{\boldsymbol{T}} \rightarrow^P \boldsymbol{\pi}$
3. $\hat{\boldsymbol{\Omega}}_{\boldsymbol{T}} \rightarrow^P \boldsymbol{\Omega}$
4. $\sqrt{T}(\hat{\boldsymbol{\pi}}_{\boldsymbol{T}} - \boldsymbol{\pi}) \rightarrow^L N(\mathbf{0}, (\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1}))$

For the coefficients of the ith regression

$$\sqrt{T}(\hat{\pi}_{iT} - \pi_i) \rightarrow^L N(\mathbf{0}, \sigma_i^2 \mathbf{Q}^{-1}))$$

for $\sigma_i^2 = E(\epsilon_{it}^2)$

Asymptotic distribution of , $\hat{\Omega}$

Notation:

$$vec \quad \Omega = vec \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{21} \\ \sigma_{22} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{32} \\ \sigma_{33} \end{bmatrix}$$

Asymptotic distribution of , $\hat{\Omega}$

Notation: Because the variance covariance matrix is symmetric $\sigma_{12} = \sigma_{21}$ and so on. We do not need all the elements. So we have the same information with the vech function

$$vech \quad \Omega = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{22} \\ \sigma_{31} \\ \sigma_{32} \\ \sigma_{33} \end{bmatrix}$$

Asymptotic distribution of , $\hat{\boldsymbol{\Omega}}$

Proposition:
If the innovations are $IIDN(\mathbf{0}, \boldsymbol{\Omega})$ and the roots of the polynomial are outside the inner circle:

$$\sqrt{T}(vech(\hat{\boldsymbol{\Omega}} - vech(\boldsymbol{\Omega}))) \rightarrow^{L} N(0, \boldsymbol{\Sigma})$$

- Elements of $\boldsymbol{\Omega}$ are σ_{ij} — variance and covariances amongst ϵ_{it} and ϵ_{jt}
- Elements of $\boldsymbol{\Sigma}$ are the covariances between $\hat{\sigma}_{ij}$ and $\hat{\sigma}_{lm}$ all $1 \leq i, j, l, m \leq n$.
- These elements are $\sigma_{il}\sigma_{jm} + \sigma_{im}\sigma_{jl}$

Important issues

- Choice of p
  - Suggested by theoretical methods
  - Rule of thumb
  - Statistical criteria (trade-off fit against number of parameters)
- Choice of variables included in the model

Choice of p

- Theoretical methods: Keynesian model...
- Rules of thumb: quarterly data (p=4), monthly data (p=6). Very large p will result in a lot of parameters. So it is unwise to fit more than $T/$ parameters in each equation. For example, $p = 4$ would mean $n < 7$.
- Statistical criteria. We can keep increasing the likelihood by increasing the number of parameters. We have to find a trade-off between fit and number of parameters
  - AIC
  - SBIC
  - Hannan-Quinn (HC)

Choice of variables

- By institutional knowledge
  - experience on previous projects,
  - context: if we have a small open economy problem, we will be using foreign output and real exchange rate as variables, however we wouldn't use liquidity.
- Theoretical models

Small number of variables means likely high p and high p indicates the need of more variables

So not only p is important also n.

Structural Analysis

- It is very difficult to get conclusions out of outputs of the VAR model
- Three main analyses:
  1. Granger causality test
  2. impulse response functions
  3. forecast error decomposition

Granger's Causality

Granger (1969) introduced the causality problem.

A variable y_{2t} is *Granger causal* for a time series variable y_{1t} if the former helps to improve the forecast of the latter.

Let $y^*_{1,t+h|\mathcal{F}_t}$ be the optimal h-step forecast of y_{1t} at origin t based on a set information of the universe $\mathcal{F}_t$. So y_{2t} is *Granger non-causal* if and only if

$$y^*_{1,t+h|\mathcal{F}_t} = y^*_{1,t+h|\mathcal{F}_t \setminus \underbrace{\{y_{2t,s|s\leq t}\}}_{\mathcal{F}_{2t}}}$$

In other words:

$$MSE(\hat{E}(y_{1,t+h}|y_{1t}, y_{1,t-1}\ldots)) = MSE(\hat{E}(y_{1,t+h}|y_{1t}, y_{1,t-1}\ldots, y_{2t}, y_{2,t-1}, \ldots))$$

Bivariate Granger's Causality

Equivalently, we say that y_{1t} *is exogenous in the time series sense with respect to* y_{2t} if the above equation holds

Another way to say it is: y_{2t} *is not linearly informative about trhe future* y_{1t}

Formally, y_2 fails to Granger-cause y_1 if for all $s > 0$ the MSE of a forecast of $y_{1,t+s}$ based on $(y_{1t}, y_{1,t-1}, \ldots)$ is the same as the MSE of a forecast of $y_{1,t+s}$ based on $(y_{1t}, y_{1,t-1}, \ldots)$ and $(y_{2t}, y_{2,t-1}, \ldots)$

The notion of Granger causality does not imply true causality. It only implies forecasting ability.

Bivariate Granger's Causality in VAR(p)

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \phi_{11}^1 & 0 \\ \phi_{12}^1 & \phi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}^2 & 0 \\ \phi_{12}^2 & \phi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \end{pmatrix}$$
$$+ \ldots + \begin{pmatrix} \phi_{11}^p & 0 \\ \phi_{12}^p & \phi_{22}^p \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

y_{2t} does not Granger-cause y_{1t} if the coefficients of matrices $\mathbf{\Phi}_j$ are lower triangular for all $j = 1, \ldots, p$

Q: What about y_{1t} does not Granger-cause y_{2t} instead ? How are the matrices $\mathbf{\Phi}_j$?

Econometric tests for Granger's causality

The p linear coefficient restrictions implied by Granger non-causality may be tested using the Wald statistic

$$Wald = (\mathbf{R} \cdot \text{vec}(\hat{\mathbf{\Pi}}) \left\{ \mathbf{R}[\widehat{AVar}(\text{vec}(\hat{\mathbf{\Omega}})]\mathbf{R}' \right\}^{-1} \times (\mathbf{R} \cdot \text{vec}(\hat{\mathbf{\Pi}})$$

$$H_0 : y_2 \text{ does not Granger-cause } y_1$$

$$H_0 : \phi_{12}^1 = \ldots = \phi_{12}^p = 0$$

- The null hypothesis is rejected if $Wald > F(p, T - 2p - 1)$
- $\mathbf{R}$ in the Wald statistics does not refer to the cross correlation function but to set of linear constraints.
- For example, to test $H_0 : c_1 = c_2$ then $\mathbf{R} = (1, 0, 0, \ldots, -1, 0, \ldots, 0)$ where -1 is in the $n + 1$ position. Then $\mathbf{R}\boldsymbol{\pi} = 0 \Rightarrow c_1 = c_2$

Vector MA(∞) representation of VAR

If the VAR(p) process is covariance-stationary, then we can represent it as a Vector MA(∞):

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon_t} + \boldsymbol{\Psi_1 \epsilon_{t-1}} + \boldsymbol{\Psi_2 \epsilon_{t-2}} + \boldsymbol{\ldots}$$

$\boldsymbol{\Psi}_s$ is the impact of the shock $\boldsymbol{\epsilon}_t$ upon y_{t+s} holding the rest constant:

$$\boldsymbol{\Psi}_s = \frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\epsilon}_t'}$$

This matrix is also called the Impulse Response (IR). Its ij element identifies the consequences of one unit increase in the jth variable's innovations at date t (ϵ_{jt}) for the value of the ith variable at time $t+s$ ($y_{i,t+s}$), holding all the innovations constant.

Impulse response functions

- We would like to know whether the IRF can tell us about the effect of a change in y_{jt} on $y_{i,t+s}$.

$$\frac{\partial y^*_{i,t+s}}{\partial y_{jt}} = \frac{\partial y_{i,t+s}}{\epsilon_{jt}}?$$

- The answer is no in general, unless $\mathbf{\Omega}$ is diagonal (uncorrelated errors).
- We can transform our VAR model into a model with uncorrelated errors.
- Sims (1980) show how to do this transformation and how to estimate the triangular *structural VAR* (SVAR).
- He got a Nobel prize last year for this kind of work.