Lecture 17
Multivariate Time Series
VAR & SVAR

 1

Vector Time Series Models

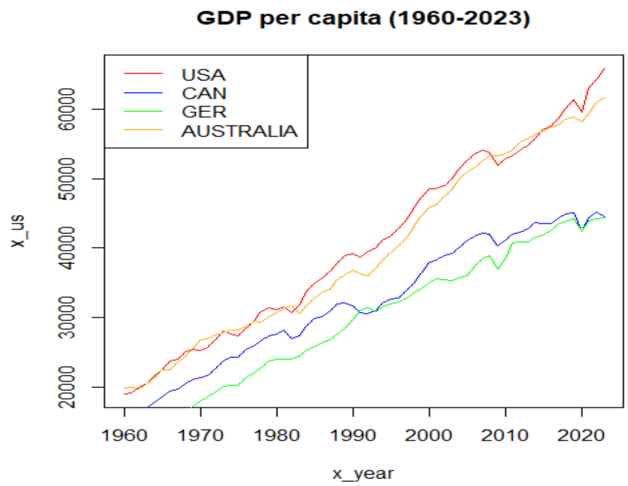
• A vector series consists of multiple single series.

• We motivated time series models by saying simple univariate ARMA models do forecasting very well. Then, why we need multiple series?

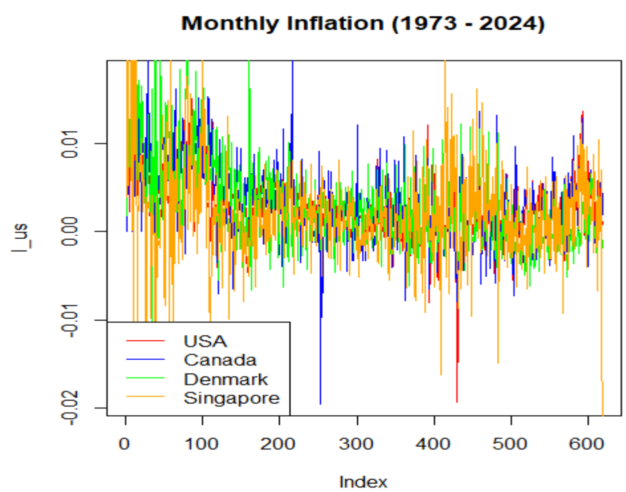
- To be able to understand the relationship between several variables, allowing for dynamics.
- To be able to get better forecasts

Example: Stock price surprises in one market (equity, NYSE) can spread easily to another market (options, Tokyo SE). Thus, a joint dynamic model may be needed to understand dynamic interrelations and may do a better forecasting job.

Vector Time Series Models



Vector Time Series Models



Vector Time Series Models

• Consider an m-dimensional time series $\boldsymbol{Y}_t = \{Y_1, Y_2, , \dots Y_m\}'$

• The series $\boldsymbol{Y}_t$is weakly stationary if its first two moments are time invariant and the cross covariance between Y_{it} and Y_{js} for all i and j are functions of the time difference $(s-t)$ only.

• The mean vector: $E[\boldsymbol{Y}_t] = \boldsymbol{\mu} = \{\mu_1, \mu_2, , \dots \mu_m\}'$

• The covariance matrix function

$$\Gamma(k) = Cov(\boldsymbol{Y}_{t-k}, \boldsymbol{Y}_t) = E\left[(\boldsymbol{Y}_{t-k} - \mu)(\boldsymbol{Y}_t - \mu)'\right]$$
$$= \begin{bmatrix} \gamma_{11}(k) & \gamma_{12}(k) & \cdots & \gamma_{1m}(k) \\ \gamma_{21}(k) & \gamma_{22}(k) & \cdots & \gamma_{2m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1}(k) & \gamma_{m2}(k) & \cdots & \gamma_{mm}(k) \end{bmatrix}$$

Vector Time Series Models

• The correlation matrix function:

$$\boldsymbol{\rho}(k) = \boldsymbol{D}^{-1/2}\Gamma(k)\boldsymbol{D}^{-1/2} = \left[\boldsymbol{\rho}_{ij}(k)\right]$$

where $\boldsymbol{D}$ is a diagonal matrix in which the i-th diagonal element is the variance of the i-th process, i.e.

• The covariance and correlation matrix functions are positive semi-definite.

$$\boldsymbol{D} = diag\left(\boldsymbol{\gamma}_{11}(0), \boldsymbol{\gamma}_{22}(0), \cdots, \boldsymbol{\gamma}_{mm}(0)\right).$$

• $\{\boldsymbol{Y}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$ if and only if $\{\boldsymbol{Y}_t\}$ is stationary with mean $\mathbf{0}$ vector and

$$\Gamma(\mathrm{k}) = \begin{cases} \Sigma, & \mathrm{k} = 0 \\ 0, & \text{otherwise} \end{cases}$$

Vector Time Series Models

• $\{\boldsymbol{Y}_t\}$ is a linear process if it can be expressed as

$$\boldsymbol{Y}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j\, \varepsilon_{t-j}, \qquad \{\boldsymbol{\varepsilon}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\{\boldsymbol{\Psi}_j\}$ is a sequence of mxT matrix whose entries are absolutely summable. That is,

$$\sum_{j=-\infty}^{\infty} |\boldsymbol{\Psi}_j(i,l)| < \infty, \qquad \text{for } i, l = 1, 2, \ldots, m$$

• For a linear process, $E[\boldsymbol{Y}_t] = \mathbf{0}$ and

$$\Gamma(k) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_{j+k} \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j{}', \quad k = 0, \pm 1, \pm 2, \ldots$$

Vector Time Series Models: MA Representation

• Let $\{\boldsymbol{Y}_t\}$ be a linear process:

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\, \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\Psi}(L) = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s\, L^s$

• For the process to be stationary, $\boldsymbol{\Psi}_s$ should be square summable in the sense that each of the mxm sequence $\Psi_{ij,s}$ is square summable.

• This is the Wold representation.

Example: VMA(2) with $m = 1$

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} + \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-1} \\ \varepsilon_{xt-1} \end{bmatrix} + \begin{bmatrix} \Psi_{31} & 0 \\ \Psi_{41} & \Psi_{42} \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-2} \\ \varepsilon_{xt-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

Vector Time Series Models: AR Representation

• Let $\{\boldsymbol{Y}_t\}$ be a linear process:

$$\boldsymbol{\Pi}(L)\ (\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\Pi}(L) = 1 - \sum_{s=0}^{\infty} \boldsymbol{\Pi}_s\, L^s$

• For the process to be invertible, $\boldsymbol{\Pi}_s$ should be absolute summable.

Example: VAR(1) with $m = 1$

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} - \begin{bmatrix} \Pi_{10} & 0 \\ 0 & \Pi_{20} \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

VARMA: Representation & Stationarity

• Let $\{\boldsymbol{Y}_t\}$ follow a VARMA$(p,\, q)$ linear process:

$$\boldsymbol{\Phi}_p(L)\ (\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}_q(L)\ \boldsymbol{\varepsilon}_t$$

where

$$\boldsymbol{\Phi}_p(L) = \boldsymbol{\Phi}_0 - \boldsymbol{\Phi}_1\, L - \boldsymbol{\Phi}_2\, L^2 - \boldsymbol{\Phi}_3\, L^3 - \dots - \boldsymbol{\Phi}_p\, L^p$$
$$\boldsymbol{\Theta}_q = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1\, L + \boldsymbol{\Theta}_2\, L^2 + \boldsymbol{\Theta}_2\, L^3 + \dots + \boldsymbol{\Theta}_q\, L^q$$

• Special cases:

$q = 0 \quad \Rightarrow \boldsymbol{\Phi}_p(L)\ (\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t$ -i.e., VAR(p)

$p = 0 \quad \Rightarrow (\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}_q(L)\ \boldsymbol{\varepsilon}_t$ -i.e., VAR(q)

• VARMA process is **stationary** if the zeros of $|\boldsymbol{\Phi}_p(L)|$ are outside the unit circle. That is, we can write:

$$(\boldsymbol{Y}_t - \boldsymbol{\mu})\ = \boldsymbol{\Psi}(L)\ \boldsymbol{\varepsilon}_t = \boldsymbol{\Phi}_p(L)^{-1}\boldsymbol{\Theta}_q(L)\ \boldsymbol{\varepsilon}_t$$

VARMA: Representation & Invertibility

- VARMA process is stationary if the zeros of $|\boldsymbol{\Phi}_p(L)|$ are outside the unit circle. That is, we can write:

$$(\boldsymbol{Y}_t - \boldsymbol{\mu}) \;= \boldsymbol{\Psi}(L)\; \boldsymbol{\varepsilon}_t = \boldsymbol{\Phi}_p(L)^{-1}\boldsymbol{\Theta}_q(L)\; \boldsymbol{\varepsilon}_t$$

- VARMA process is **invertible** if the zeros of $|\boldsymbol{\Theta}_q(L)|$ are outside the unit circle. That is, we can write:

$$\boldsymbol{\Pi}(L)\;(\boldsymbol{Y}_t - \boldsymbol{\mu}) \;= \boldsymbol{\varepsilon}_t$$
$$\boldsymbol{\Theta}_q(L)^{-1}\boldsymbol{\Phi}_p(L)(\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t$$

- Identification problem: Multiplying matrices by some arbitrary matrix polynomial may give us an identical covariance matrix. Then, the VARMA$(p,\, q)$ model is not identifiable (not unique p & q).

VARMA: Identification Problem

Example: VARMA(1,1) process

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} - \begin{bmatrix} 0 & \alpha + m \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} - \begin{bmatrix} 0 & -m \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

$$\begin{bmatrix} 1 & -(\alpha + m)L \\ 0 & 1 \end{bmatrix}\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & mL \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & -(\alpha+m)L \\ 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & mL \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & -\alpha L \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}$$

MA(∞)=VMA(1)

12

VARMA – Identification Problem

VECTOR ARMA MODELS - VARMA

Here, one of the issues is the *identifiability* problem.
Examples:
VMA(1) = VAR(1):

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

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

VARMA(1,1):

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

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

They are identical.

13

VARMA: Identification Problem

• To eliminate this problem, there are three methods suggested by Hannan (1969, 1970, 1976, 1979).

• From each of the equivalent models, choose the minimum MA order q and AR order p. The resulting representation will be unique if *Rank*$(\boldsymbol{\Phi}_p(L)) = m$.

• Represent $\boldsymbol{\Phi}_p(L)$ in lower triangular form. If the order of $\phi_{ij}(L)$ for $i, j = 1, 2, \ldots, m$, then the model is identifiable.

• Represent $\boldsymbol{\Phi}_p(L)$ in a form $\boldsymbol{\Phi}_p(L) = \phi_p(L)\, \boldsymbol{I}$, where $\phi_p(L)$ is a univariate AR(p). The model is identifiable if $\phi_p(L) \neq 0$.

14

VAR(1) Process: Stationarity & Eigenvalues

• In a VAR process, $Y_{i,t}$ depends not only the lagged values of $Y_{i,t}$ but also the lagged values of the other variables. For the VAR(1):

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 L)\,(\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t$$

- Always invertible.

- Stationary if $|\boldsymbol{I} - \boldsymbol{\Phi}_1 L|$ outside the unit circle. Let $\lambda = L^{-1}$.

$$|\boldsymbol{I} - \boldsymbol{\Phi}_1 L| = 0 \Rightarrow |\lambda - \boldsymbol{\Phi}_1 \boldsymbol{I}| = 0$$

$$\downarrow$$

The zeros of $[\boldsymbol{I} - \boldsymbol{\Phi}_1 L|$ is related to the eigenvalues of $\boldsymbol{\Phi}_1$.

• Hence, VAR(1) process is stationary if the eigenvalues of $\boldsymbol{\Phi}_1$, λ_i, $i = 1, 2, \ldots, m$, are all inside the unit circle.

15

VAR(1) Process

Example: Check stationarity of the following VAR(1) process:

$$\boldsymbol{Y}_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

$$= \begin{bmatrix} 1.1 & -0.3 \\ 0.6 & 0.2 \end{bmatrix} \boldsymbol{Y}_t + \boldsymbol{\varepsilon}_t$$

We check roots of $|\boldsymbol{I} - \boldsymbol{\Phi}_1 L| = 0$

Or equivalently, we check eigenvalues of $\boldsymbol{\Phi}_1$: $|\boldsymbol{\Phi}_1 - \lambda \boldsymbol{I}| = 0$

$$\Rightarrow \quad \begin{vmatrix} 1.1 - \lambda & -0.3 \\ 0.6 & 0.2 - \lambda \end{vmatrix} = 0 \Rightarrow \lambda_1 = 0.8;\ \lambda_2 = 0.5.$$

• The process is stationary.

16

VAR(1) Process: Autocovariance Matrix

• The autocovariance matrix:

$$\Gamma(\mathrm{k}) = \mathrm{E}\left[Y_{\mathrm{t-k}} Y_{\mathrm{t}}'\right] = \mathrm{E}\left[Y_{\mathrm{t-k}}\left(\Phi Y_{\mathrm{t-1}} + \varepsilon_{\mathrm{t}}\right)'\right]$$

$$= \mathrm{E}\left[Y_{\mathrm{t-k}} Y_{\mathrm{t-1}}' \Phi' + Y_{\mathrm{t-k}} \varepsilon_{\mathrm{t}}'\right]$$

$$\Gamma(k) = \begin{cases} \Gamma(-1)\Phi' + \Sigma, k = 0 \\ \Gamma(k-1)\Phi' = \Gamma(0)(\Phi')^k, |k| \geq 1 \end{cases}$$

• For k =1, $\Gamma(1) = \Gamma(0)(\Phi') \Rightarrow \Phi = \Gamma'(1)\Gamma^{-1}(0)$

$$\Sigma = \Gamma(0) - \underbrace{\Gamma(-1)}_{\Gamma'(1)}\Gamma^{-1}(0)\Gamma(1)$$

$$= \Gamma(0) - \underbrace{\Gamma'(1)\Gamma^{-1}(0)}_{\Phi}\Gamma(0)\underbrace{\Gamma^{-1}(0)\Gamma(1)}_{\Phi'}$$

$$= \Gamma(0) - \Phi\Gamma(0)\Phi'$$

17

VAR(1) Process: Autocovariance Matrix

• Then,

$$\Gamma(0) = \Sigma + \Phi\Gamma(0)\Phi'$$

$$vec(\Gamma(0)) = [I - \Phi \otimes \Phi]^{-1} vec(\Sigma)$$

where $\otimes$ = Kronecker product

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

$$e.g. \boldsymbol{X} = \begin{bmatrix} 3 & 2 \\ 4 & 6 \\ 1 & 7 \end{bmatrix} \Rightarrow vec(\boldsymbol{X}) = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 2 \\ 6 \\ 7 \end{bmatrix} \qquad e.g.\ \boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_{m1}B & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}$$

18

VAR(1) Process: Autocovariance Matrix – VMA

• In a VMA process, $Y_{i,t}$ depends not only the lagged values of $\varepsilon_{i,t}$ but also the lagged values of the errors of other variables. For the VMA(1):

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \varepsilon_t + \boldsymbol{\Theta}_1\, \boldsymbol{\varepsilon}_{t-1}, \qquad \{\boldsymbol{\varepsilon}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$$

- Always stationary.

- The autocovariance function:

$$\Gamma(0) = \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1 \boldsymbol{\Sigma}\, \boldsymbol{\Theta}_1{}'$$

$$\Gamma(k) = \begin{cases} -\boldsymbol{\Sigma}\, \boldsymbol{\Theta}_1' & k = 1 \\ -\boldsymbol{\Theta}_1 \boldsymbol{\Sigma} & k = -1 \\ \mathbf{0} & \text{otherwise} \end{cases}$$

- The autocovariance matrix function cuts of after lag 1.

• Thus, VMA(1) process is invertible if the eigenvalues of *Θ;* λ_i, $i = 1, 2, \ldots, m$, are all inside the unit circle.

19

VARMA – Identification

• Same idea as in univariate case. We define the **Sample Correlation Matrix Function** (**SCMF**): Given a vector m series of T observations, the sample correlation matrix function is

$$\hat{\rho}(k) = |\hat{\rho}_{ij}(k)|$$

where $\hat{\rho}_{ij}(k)$'s are the crosscorrelation for the i-th and j-th component series.

• It is useful to identify VMA(q).

• Tiao and Box (1981) proposed to use +,– and . signs to show the significance of the cross correlations:

+ (-) sign: the value is greater (less) than 2 times the estimated SE

. sign: the value is within the 2 times estimated SE

20

Partial Autoregression or Partial Lag Correlation Matrix Function

• They are useful to identify VAR order. The partial autoregression matrix function is proposed by Tiao and Box (1981), but it is not a proper correlation coefficient.

• Then, Heyse and Wei (1985) have proposed the partial lag correlation matrix function which is a proper correlation coefficient.

• Both of them can be used to identify the VARMA(p, q).

21

Granger Causality

• In principle, the concept is as follows:
If X causes Y, then, changes of X happened first then followed by changes of Y.

• Then, if X causes Y, there are two conditions to be satisfied:
1. X can help in predicting Y. (Regression of X on Y has a big R^2.)
2. Y can not help in predicting X.

• In most regressions, it is hard to discuss causality. For instance, the significance of the coefficient β in the regression

$$y_t = \beta x_t + \varepsilon_t \; \beta$$

only tells there is a relationship between x_t and y_t, not that x_t causes y_t.

22

Granger Causality

• Vector autoregression allows a test of '**causality**' in the previous sense. This test is first proposed by Granger (1969) and later by Sims (1972) therefore we called it **Granger (or Granger-Sims) causality.**

• We will restrict our discussion to a system of two variables, x_t and y_t: y_t is said to Granger-cause x_t if current or lagged values of y_t helps to predict future values of x_t.

-- On the other hand, y fails to Granger-cause x_t if for all $s > 0$, the MSE of a forecast of x_{t+s} based on $(x_t, x_{t-1}, \ldots)$ is the same as that is based on $(y_t, y_{t-1}, \ldots)$ and $(x_t, x_{t-1}, \ldots)$.

• For linear functions, y_t fails to Granger-cause x_t if

$$MSE\left[\hat{E}\left(x_{t+s} | x_t, x_{t-1}, \cdots\right)\right] = MSE\left[\hat{E}\left(x_{t+s} | x_t, x_{t-1}, \cdots, y_t, y_{t-1}, \cdots\right)\right]$$

23

Granger Causality

• Restricting ourselves to linear functions, y_t fails to Granger-cause x_t if

$$\text{MSE}[\text{E}[x_{t+s} | x_t, x_{t-1}, \ldots] = \text{MSE}[\text{E}[x_{t+s} | x_t, x_{t-1}, \ldots, y_t, y_{t-1}, \ldots]$$

• Equivalently, we can say that x_t is exogenous in the time series sense with respect to y_t, or y_t is not linearly informative about future x_t.

• A variable X is said to Granger cause another variable Y, if Y can be better predicted from the past of X and Y together than the past of Y alone, other relevant information being used in the prediction (Pierce, 1977).

24

Granger Causality: VAR Formulation

• In the VAR equation, the example we proposed above (x_t *Granger causes* y_t) implies a lower triangular coefficient matrix:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}^1 & 0 \\ \phi_{21}^1 & \phi_{22}^1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \phi_{11}^p & 0 \\ \phi_{21}^p & \phi_{22}^p \end{bmatrix} \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

• Or if we use MA representations,

$$\boldsymbol{Y}_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \Psi_{11}(L) & 0 \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

where

$$\Psi_{11}(L) = \phi_{ij}^0 + \phi_{ij}^1 L + \phi_{ij}^2 L^2 + \cdots,$$
$$\phi_{11}^0 = \phi_{22}^0 = 1, \phi_{21}^0 = 0$$

25

Granger Causality: Test

• Consider a linear projection of y_t on past, present and future x_t's,

$$y_t = \mathrm{c} + \sum_{j=0}^{\infty} b_j\, x_{t-j} + \sum_{j=0}^{\infty} d_j\, x_{t+j} + \varepsilon_t,$$

where $E[\varepsilon_t x_\tau] = 0$ for all t and τ. Then, y_t fails to Granger-cause x_t iff $d_j = 0$ for $j = 1, 2, \ldots.$

• Steps

1) Check that both series are stationary in mean, variance and covariance (if, not, transform data via differences, logs, etc.)

2) Estimate AR(p) models for each series. Make sure residuals are white noise. *F*-tests and/or AIC, BIC can be used to determine p.

3) Re-estimate both models, with all the lags of the other variable.

4) Use *F*-tests to determine whether, after controlling for past *Y*, past values of *X* can improve forecasts *Y* (and vice versa).

26

Granger Causality: Setting up the *F*-Test

• Causality Model:

$$x_t = c_1 + \sum_{j=0}^{p_x} \alpha_j\, x_{t-j} + \sum_{j=0}^{p_y} \beta_j\, y_{t-j} + \varepsilon_t,$$

H_0 (y_t does not Granger cause x_t): $\beta_1 = \beta_2 = ... = \beta_{p_y} = 0$.

• Steps in practice

1) Once the lag structures are determined, estimate the causality model. Keep RSS_U.

2) Estimate a restricted regression (without the y_t's). Keep RSS_R.

3) Construct F-test as usual:

$$F = [(T-k)/p_y] * [(RSS_R - RSS_U)/RSS_U]$$

where $k = (1 + p_x + p_y)$ is the number of parameters from model U, q is the number of parameters from model $R = (1 + p_x)$ and $p_y = (k - q)$.

27

Granger Causality: Possible Outcomes

• There are 4 possible conclusions from test:

1. X Granger causes Y, but Y does not Granger cause X
2. Y Granger causes X, but X does not Granger cause Y
3. X Granger causes Y and Y Granger causes X --i.e., there is a feedback system or bidirectional causality.
4. X does not Granger cause Y and Y does not Granger cause X

28

Granger Causality: Example

• From Chuang and Susmel (2010): Bivariate analysis of relation between stock returns and Volume in Taiwan.

$$V_{ij,t} = \alpha_{ij,1} + \sum_{a=0}^{A} \beta_{ij,11a} DAVR_{m,t-a} + \sum_{b=0}^{B} \beta_{ij,12b} DMAD_{ij,t-b} + \sum_{c=1}^{C} \beta_{ij,13c} R_{ij,t-c}$$

$$+ \sum_{d=1}^{D} \gamma_{ij,11d} V_{ij,t-d} + \sum_{d=1}^{D} \gamma_{ij,12d} R_{m,t-d} + \varepsilon_{ij,1t},$$

$$R_{m,t} = \alpha_{ij,2} + \sum_{d=1}^{D} \gamma_{ij,21d} V_{ij,t-d} + \sum_{d=1}^{D} \gamma_{ij,22d} R_{m,t-d} + \varepsilon_{ij,2t},$$

$V_{ij,t}$: Detrended trading volume of portfolio ij,

$R_{m,t}$: Return on a value-weighted Taiwanese market index,

$R_{ij,t}$: Return of portfolio ij,

$DAVR_{m,t}$: Detrended absolute value of market returns, and

$DMAD_{ij,t}$: Detrended mean absolute portfolio return deviation.

Portfolio ij: Portfolio of size i and institutional ownership j.

29

Granger Causality: Example

• Estimation SUR

• Granger causality tests (Wald tests) $\gamma_{ij,12d}$

- For any portfolio ij we test H$_0$: $\gamma_{ij,12d} = 0$ for all d.

⇒ Market returns do not Granger-cause portfolio volume.

- Sign of causality. If the sum of the $\gamma_{ij,12d}$ coefficients is significantly positive ⇒ Positive causality from market returns to trading volume

- For any portfolio ij we test H$_0$: $\gamma_{ij,21d} = 0$ for all d.

⇒ Portfolio volume do not Granger-cause market returns.

- *W-D* statistics: Granger causality test --it follows a χ^2_D.

- *W*-1: Sum of the lagged coefficients is equal to zero (identify the sign of the causality) --it follows a χ^2_1

30

Granger Causality: Example

Panel A: Size-institutional ownership portfolios

P_i $_j$	Hypothesis 1	Does causality exist? (*W*-*D* statistic)	Sum of lagged coefficients	Hypothesis 2	Sign of causality (*W*-1 statistic)
P	$\gamma_{1l,12d} = 0$ for all d	Yes (19.4919***)	0.0382	$\sum_{d=1}^{2} \gamma_{1l,12d} = 0$	Positive (19.4514***)
$1l$	$\gamma_{1l,21d} = 0$ for all d	No (0.1566)	0.0466		
P	$\gamma_{1h,12d} = 0$ for all d	Yes (21.2543***)	0.0285	$\sum_{d=1}^{2} \gamma_{1h,12d} = 0$	Positive (21.1123***)
$1h$	$\gamma_{1h,21d} = 0$ for all d	No (0.0658)	0.0559		
P	$\gamma_{2l,12d} = 0$ for all d	Yes (15.8748***)	0.0446	$\sum_{d=1}^{2} \gamma_{2l,12d} = 0$	Positive (15.7221***)
$2l$	$\gamma_{2l,21d} = 0$ for all d	No (0.7614)	00.1864		
P	$\gamma_{2h,12d} = 0$ for all d	Yes (11.2518***)	0.0150	$\sum_{d=1}^{2} \gamma_{2h,12d} = 0$	Positive (11.1957***)
$2h$	$\gamma_{2h,21d} = 0$ for all d	No (1.9206)	0.2649		
P	$\gamma_{3l,12d} = 0$ for all d	Yes (39.4826***)	0.0569	$\sum_{d=1}^{3} \gamma_{3l,12d} = 0$	Positive (35.7789***)
$3l$					

31

Granger Causality: Example

• H_0: $\gamma_{ij,12d} = 0$ for all d $\Rightarrow$ rejected for all size-institutional ownership portfolios (shown in previous Table) and all volume-institutional ownership portfolios (not shown), respectively.

- The cumulative effect of lagged market returns on portfolio volume is positive –i.e., $\Sigma_j\, \gamma_{ij,12d,t-j} > 0$- and significant.

• H_0 : $\gamma_{ij,21d}$= 0 for all d. $\Rightarrow$ cannot be rejected for any size-institutional ownership portfolios (shown) and any volume-institutional ownership portfolios (not shown), respectively.

• No feedback relation between portfolio volume and market returns (consistent with the sequential information arrival orthe positive feedback trading hypotheses).

32

Granger Causality: Chicken or Egg?

• This causality test is also can be used in explaining which comes first: chicken or egg. More specifically, the test can be used in testing whether the existence of egg causes the existence of chicken or vise versa.

• Thurman and Fisher (1988) did this study using yearly data of chicken and egg productions in the US from 1930 to1983.

• The results:

1. Egg causes the chicken.
2. There is no evidence that chicken causes egg.

33

Granger Causality: Remarks

• Granger causality does not equal to what we usually mean by causality.

• Even if x_1 does not cause x_2, it may still help to predict x_2, and thus Granger-causes x_2 if changes in x_1 precedes that of x_2 for some reason (usually because of a third variable, missing in the model).

Example: A dragonfly flies much lower before a rain storm, due to the lower air pressure. We know that dragonflies do not cause a rain storm, but it does help to predict a rain storm, thus Granger-causes a rain storm.

34

Granger Causality: Exogeneity

• When x_1 does not cause x_2, we say that x_2 is **strongly exogenous** and thus Granger-causes x_1 if changes in x_2 precedes that of x_1 for some reason (usually because of a third variable, missing in the model).

Example: The dragonfly is strongly exogenous with respect to rain.

35

Structural VAR (SVAR)

• It is a simultaneous equations model. It is used to described dynamic effects in a multivariate system. For example,

$$\boldsymbol{B}\boldsymbol{Y}_t = \boldsymbol{\Gamma}_{\mathbf{0}} + \boldsymbol{\Gamma}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\Gamma}_2 \boldsymbol{Y}_{t-2} + ... + \boldsymbol{\Gamma}_p \boldsymbol{Y}_{t-p} + \varepsilon_t$$

where

$$\varepsilon_t \sim iid\ D(\mathbf{0}, \boldsymbol{\Sigma})$$

• Note:

- ε_{1t}, ε_{2t},..., ε_{nt} are called structural errors. $\boldsymbol{\Sigma}$ is a diagonal matrix.
- In general, cov(y_{it}, ε_{jt}) $\neq 0$ for all i, j.
- All variables are endogenous - OLS is not appropriate

• From this model, we can move to a reduced form, say

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}_{\mathbf{0}} + \boldsymbol{\Phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{Y}_{t-2} + ... + \boldsymbol{\Phi}_p \boldsymbol{Y}_{t-p} + a_t$$

• The a_t's are called reduced form errors, a linear combination of ε_t's.

Structural VAR (SVAR)

• Like in SEM, we have identification issues. To recover the structural parameters ($\mathbf{B}$, Γ, $\boldsymbol{\Sigma}$) we need to impose restrictions.

• Many applications in finance:
- The effect of financial news on stock prices (or returns):
  - [Δr, Δy, Δπ, Δd, Δc → P (or ΔP)] - see Chen et al (JB, 1986)
- Analysis of policy effects (ΔM^S, ΔG) on stock market.
- Relative importance of markets (stock vs. options markets).

• Long history in economics. Formalized by Sims (*Econometrica,* 1980), as a generalization of univariate analysis to an array of RV. Sims analyzed a 3x1 vector $\boldsymbol{Y}_t$ with elements, Money supply (Z_t), interest rates (X_t) & income (V_t), in reduced form:

$$\boldsymbol{Y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{Y}_{t-2} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{Y}_{t-p} + a_t \Rightarrow \text{VAR}(p)$$

SVAR: Sims (1980) Formulation

• Sims analyzed: Money supply (Z_t), interest rates (X_t) & income (V_t) in reduced form:

$$\boldsymbol{Y}_t = \begin{bmatrix} Z_t \\ X_t \\ V_t \end{bmatrix} = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{Y}_{t-2} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{Y}_{t-p} + a_t \Rightarrow \text{VAR}(p)$$

with

$$E[a_t] = 0$$

$$E[a_t a_\tau{}'] = \begin{cases} \boldsymbol{\Omega} & t = \tau \\ 0 & t \neq \tau \end{cases}$$

$\boldsymbol{\Phi}_i$ are matrices. For the 1st lag matrix $\Rightarrow$ $\Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}^{(1)}$

• A typical equation of the system is:

$$Z_t = c_1 + \phi^{(1)}{}_{11} Z_{t-1} + \phi^{(1)}{}_{12} X_{t-1} + \phi^{(1)}{}_{13} V_{t-1} + \ldots + \phi_{11}{}^{(p)} Z_{t-p} + \phi_{12}{}^{(p)} X_{t-p} + \phi_{13}{}^{(p)} V_{t-p} + a_{1t}$$

Note: Each equation has the same regressors.

SVAR: Multivariate Models

• VARMAX Models, like a VAR, but allows exogenous variables, $\boldsymbol{X}_t$:

$$\Phi_P(L)\ \boldsymbol{Y}_t = G_m(L)\ \boldsymbol{X}_t + \Theta_{\boldsymbol{q}}(L)\ \varepsilon_t$$

• Structural VAR Models:

$$\boldsymbol{BY}_t = \boldsymbol{\Gamma_0} + \boldsymbol{\Gamma}_1\, \boldsymbol{Y}_{t-1} + \boldsymbol{\Gamma}_2\, \boldsymbol{Y}_{t-2} + \ldots + \boldsymbol{\Gamma}_p\, \boldsymbol{Y}_{t-p} + \varepsilon_t \qquad \text{SVAR}(p)$$

where $\varepsilon_t \sim iid\ D(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal matrix.

- Some theory to determine $\boldsymbol{Y}_t$, but all variables are endogenous.

• VAR Models (reduced form):

$$\boldsymbol{Y}_t = \boldsymbol{\Phi_0} + \boldsymbol{\Phi}_1\, \boldsymbol{Y}_{t-1} + \boldsymbol{\Phi}_2\, \boldsymbol{Y}_{t-2} + \ldots + \boldsymbol{\Phi}_p\, \boldsymbol{Y}_{t-p} + a_t$$

where the error term is a WN vector: $E[a_t a_\tau{}'] = \begin{cases} \boldsymbol{\Omega} & t = \tau \\ 0 & t \neq \tau \end{cases}$

SVAR: Multivariate Models

• VAR Models (reduced form):

$$\boldsymbol{Y}_t = \boldsymbol{\Phi_0} + \boldsymbol{\Phi}_1\, \boldsymbol{Y}_{t-1} + \boldsymbol{\Phi}_2\, \boldsymbol{Y}_{t-2} + \ldots + \boldsymbol{\Phi}_p\, \boldsymbol{Y}_{t-p} + a_t$$

where the error term is a WN vector: $E[a_t a_\tau{}'] = \begin{cases} \boldsymbol{\Omega} & t = \tau \\ 0 & t \neq \tau \end{cases}$

• $\boldsymbol{Y}_t$ is a function of predetermined variables ($\boldsymbol{Y}_{t-j}$'s) and errors are well behaved: OLS is possible.

SVAR: VAR(1)

• Consider a bivariate $\boldsymbol{Y}_t = (y_t, x_t)$, first-order VAR model:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} - \begin{bmatrix} b_{12} & 0 \\ 0 & b_{21} \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

• The error terms (structural shocks) ε_{yt} and ε_{xt} are uncorrelated WN innovations with standard deviations σ_y and σ_x (& *zero* covariance).

• Note:

- y_t & x_t are endogenous. ε_{yt} affects y_t directly and x_t indirectly.
- Many parameters to estimate: **10**.

• The structural VAR is not a reduced form. In a reduced form representation y_t and x_t are just functions of lagged y_t and x_t.

SVAR: VAR(1) – Reduced Form

• To get a reduced form write the structural VAR in matrix form as:

$$\begin{bmatrix} b_{12} & 1 \\ 1 & b_{21} \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

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

• Premultipication by $\mathbf{B}^{-1}$ allow us to obtain a standard VAR(1):

$$\boldsymbol{Y}_t = \boldsymbol{B}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{0}} + \boldsymbol{B}^{-1}\boldsymbol{\Gamma}_1\, \boldsymbol{Y}_{t-1} + \boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t$$

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1\, \boldsymbol{Y}_{t-1} + a_t$$

• This is the reduced form to estimate (by OLS equation by equation). Now, we have only 9 (6 mean, 3 variance) parameters.

• From the a reduced form: $(\boldsymbol{I} - \boldsymbol{\Phi}_1\, L)\, \boldsymbol{Y}_t = \boldsymbol{\Phi}(L)\, \boldsymbol{Y}_t = \boldsymbol{\Phi}_0 + a_t$, the stability depends on the roots of $(\boldsymbol{I} - \boldsymbol{\Phi}_1\, L)$.

SVAR: Stability Conditions

• From the a reduced form:

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2\ L^2 - \cdots - \boldsymbol{\Phi}_p\ L^p)\ \boldsymbol{Y}_t = \boldsymbol{\Phi}_0 +\ a_t$$

$$\boldsymbol{\Phi}(L)\ \boldsymbol{Y}_t = \boldsymbol{\Phi}_0 +\ a_t$$

$\boldsymbol{\Phi}(L)$ is a nxn matrix polynomial in L, with the ij element

$$(\delta_{ij} - \phi_{ij}{}^{(1)}L - \phi_{ij}{}^{(2)}\ L^2 - \cdots - \phi_{ij}{}^{(p)}\ L^p) \quad \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

• A VAR(p) for $\boldsymbol{Y}_t$ is **stable** if the pxn roots of the characteristic polynomial are outside the unit circle.
- The characteristic polynomial:

$$|\boldsymbol{I} - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2\ L^2 - \cdots - \boldsymbol{\Phi}_p\ L^p| = 0$$

• Then, we define the constant:

$$\boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \cdots - \boldsymbol{\Phi}_p)^{-1}\ \boldsymbol{\Phi}_0$$

SVAR: Stability Conditions

• If the VAR is stable then a MA(∞) representation exists.

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{a}_t + \boldsymbol{\Psi}_1\ \boldsymbol{a}_{t-1} + \boldsymbol{\Psi}_2\ \boldsymbol{a}_{t-2} + \cdots = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{a}_t$$

• This representation will be the "key" to study the **impulse response function** (**IRF**) of a given shock.

Example: For the VAR(1), multiply both sides of reduced form by $(\boldsymbol{I} - \boldsymbol{\Phi}_1 L)^{-1}$. Then,

$$\boldsymbol{\Psi}(L) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 L)^{-1} \qquad \Rightarrow \boldsymbol{\Psi}_0 = \boldsymbol{I}$$

$$\boldsymbol{\Psi}_k = \boldsymbol{\Phi}_1{}^{\boldsymbol{k}}$$

$$\boldsymbol{\mu} = \boldsymbol{\Phi}(1)^{-1}\ \boldsymbol{\Phi}_0$$

Structural MA (SMA) Representation

• The SMA of $\boldsymbol{Y}_t$ is based on an infinite moving average of the structural innovations, $\boldsymbol{\varepsilon}_t$. Using $\boldsymbol{a}_t = \boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t$ in the Wold representation gives

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{a}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t$$
$$= \boldsymbol{\mu} + \boldsymbol{\Theta}(L)\,\boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\Theta}(L) = \boldsymbol{\Psi}(L)\boldsymbol{B}^{-1} = \boldsymbol{B}^{-1} + \boldsymbol{\Psi}_1\boldsymbol{B}^{-1} + \boldsymbol{\Psi}_2\boldsymbol{B}^{-1} + \cdots$$

That is,

$$\boldsymbol{\Theta}_0 = \boldsymbol{B}^{-1} \neq \mathbf{I}$$
$$\boldsymbol{\Theta}_1 = \boldsymbol{\Psi}_1\boldsymbol{B}^{-1} = \boldsymbol{\Phi}_1$$
$$\ldots.$$
$$\boldsymbol{\Theta}_k = \boldsymbol{\Psi}_k\boldsymbol{B}^{-1} = \boldsymbol{\Phi}_1{}^{\boldsymbol{k}}$$

SVAR: VAR(p) to VAR(1)

• Re-writing the system in deviations from its mean:

$$(\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\Phi}_1(\boldsymbol{Y}_{t-1} - \boldsymbol{\mu}) + \ldots + \boldsymbol{\Phi}_p(\boldsymbol{Y}_{t-p} - \boldsymbol{\mu}) + a_t$$

- Stack the vectors as

$$\eta_t \equiv \begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{bmatrix} \qquad F \equiv \begin{bmatrix} \Phi_1 & \Phi_2 \ldots\ldots & \Phi_{p-1} & \Phi_p \\ I_n & 0 \ldots\ldots\ldots & \ldots.. & 0 \\ 0 & I_n \ldots\ldots\ldots & \ldots.. & 0 \\ \vdots & & & \\ 0 & 0 \ldots\ldots\ldots & I_n & 0 \end{bmatrix} \qquad v_t = \begin{bmatrix} a_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(nxp)x1 (nxp)x(nxp) (nxp)x1

- Write the VAR(1): $\eta_t = F\eta_{t-1} + v_t \quad E(v_t v_\tau') = \begin{cases} H & t = \tau \\ 0 & t \neq \tau \end{cases}$

$$\text{where} \quad H = \begin{bmatrix} \Omega\ 0\ldots..0 \\ 0\ 0\ldots\ldots0 \\ \vdots \\ 0\ 0\ldots\ldots0 \end{bmatrix}$$

(nxp)x(nxp)

SVAR: Estimation – MLE

• We assume normality for the errors. Then, we use the conditioning trick to write down the joint likelihood. That is,

$$f(Y_T, Y_{T-1}.....Y_1 \mid Y_0, Y_{-1}....Y_{-p+1}; \nu) = \prod_{t=1}^{T} f(Y_t \mid Y_{t-1}, Y_{t-2}....Y_{t-p+1}; \nu)$$

$$Y_t \mid Y_{t-1}, Y_{t-2}.... \rightarrow N(c + \Phi_1 Y_{t-1} + ...\Phi_p Y_{t-p}, \Omega)$$

$$\Pi' \equiv [c\ \Phi_1\ \Phi_2.....\Phi_p]$$

$$X_t \equiv [1\ Y_{t-1}\ Y_{t-2}......Y_{t-p}]' \qquad \text{n x (np+1)}$$

$$Y_t = \Pi' X_t + a_t \qquad \text{(np+1) x 1}$$

$$\ell(\nu) = \sum_{t=1}^{T} \log f(Y_t \mid past; \nu) =$$

$$= -\frac{Tn}{2}\log(2\pi) + \frac{T}{2}\log\left|\Omega^{-1}\right| - \frac{1}{2}\sum_{t=1}^{T}\left[\left(Y_t - \Pi' X_t\right)'\Omega^{-1}\left(Y_t - \Pi' X_t\right)\right]$$

SVAR: Estimation – OLS = MLE

• Under the previous assumptions, we get that OLS = MLE. That is,

$$\hat{\Pi}_{mle} = \hat{\Pi}_{ols} \qquad \hat{\Pi}'_{ols} = \left[\sum_{t=1}^{T} Y_t X'_t\right]\left[\sum_{t=1}^{T} X_t X'_t\right]^{-1}$$

• Proof:

$$\sum_{t=1}^{T}\left(Y_t - \Pi' X_t\right)'\Omega^{-1}\left(Y_t - \Pi' X_t\right) =$$

$$= \sum_{t=1}^{T}\left(Y_t - \hat{\Pi}_{ols}' X_t + \hat{\Pi}_{ols}' X_t - \Pi' X_t\right)\Omega^{-1}\left(Y_t - \hat{\Pi}_{ols}' X_t + \hat{\Pi}'_{ols} X_t - \Pi' X_t\right) =$$

$$= \sum_{t=1}^{T}\left(\hat{a}_t + (\hat{\Pi}_{ols} - \Pi)' X_t\right)\Omega^{-1}\left(\hat{a}_t + (\hat{\Pi}_{ols} - \Pi)' X_t\right) =$$

$$= \sum_t \hat{a}'_t \Omega^{-1} \hat{a}_t + \sum_t X'_t (\hat{\Pi}_{ols} - \Pi)\Omega^{-1}(\hat{\Pi}_{ols} - \Pi)' X_t + 2\sum_t \hat{a}'_t \Omega^{-1}(\hat{\Pi}_{ols} - \Pi)' X_t$$

$$(*) \quad \sum_t \hat{a}'_t \Omega^{-1}(\hat{\Pi}_{ols} - \Pi)' X_t = tr\left[\sum_t \hat{a}'_t \Omega^{-1}(\hat{\Pi}_{ols} - \Pi)' X_t\right] =$$

$$= tr\left[\sum_t \Omega^{-1}(\hat{\Pi}_{ols} - \Pi)' X_t \hat{a}'_t\right] = tr\left[\Omega^{-1}(\hat{\Pi}_{ols} - \Pi)\sum_t X_t \hat{a}'_t\right] = 0$$

SVAR: Estimation – OLS (univariate)

$$\min \sum_{t=1}^{T} (Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t) =$$

$$= \min \sum_t \hat{a}'_t \Omega^{-1} \hat{a}_t + \sum_t X'_t (\hat{\Pi}_{ols} - \Pi) \Omega^{-1} (\hat{\Pi}_{ols} - \Pi)' X_t$$

because Ω is p.d. matrix $\rightarrow \Omega^{-1}$ is p.d., the smallest v alue is achieved when $\hat{\Pi}_{ols} = \Pi$

• Then, the (reduced form) VAR can be estimated equation by equation by OLS.

SVAR: Testing

• Testing as usual. For example, the LR in a VAR. We need to estimate the restricted model (under H_0) and unrestricted (under H_1). The likelihood is given by:

$$\ell(\hat{\Omega}, \hat{\Pi}) = -\frac{Tn}{2} \log(2\pi) + \frac{T}{2} \log \left| \hat{\Omega}^{-1} \right| - \frac{1}{2} \sum_{t=1}^{T} \hat{a}'_t \hat{\Omega}^{-1} \hat{a}_t$$

$$\frac{1}{2} \sum_{t=1}^{T} \hat{a}'_t \hat{\Omega}^{-1} \hat{a}_t = \frac{1}{2} trace \left[\sum_{t=1}^{T} \hat{a}'_t \hat{\Omega}^{-1} \hat{a}_t \right] = \frac{1}{2} trace \left[\sum_{t=1}^{T} \hat{\Omega}^{-1} \hat{a}_t \hat{a}'_t \right] =$$

$$= \frac{1}{2} trace \left[\hat{\Omega}^{-1} T \hat{\Omega} \right] = \frac{1}{2} trace \left[T I_n \right] = \frac{Tn}{2}$$

$$\ell(\hat{\Omega}, \hat{\Pi}) = -\frac{Tn}{2} \log(2\pi) + \frac{T}{2} \log \left| \hat{\Omega}^{-1} \right| - \frac{Tn}{2}$$

• Suppose we want to test $p_1 > p_0$: $H_0 : VAR(p_0)$
$H_1 : VAR(p_1)$

SVAR: Testing

• First, we estimate the model under H_0, with p_0 parameters. The estimation consists of n OLS regression of each variable on a constant and p_0 lags:

$$\ell(\hat{\Omega}, \hat{\Pi}) = -\frac{Tn}{2}\log(2\pi) + \frac{T}{2}\log\left|\hat{\Omega}_0^{-1}\right| - \frac{Tn}{2}$$

• Second, we estimate the model under H_1, with p_1 parameters:

$$\ell(\hat{\Omega}, \hat{\Pi}) = -\frac{Tn}{2}\log(2\pi) + \frac{T}{2}\log\left|\hat{\Omega}_1^{-1}\right| - \frac{Tn}{2}$$

• Construct LR, as usual:

$$LR = 2(\ell^*_1 - \ell^*_0) = T\left\{\log\left|\hat{\Omega}_1^{-1}\right| - \log\left|\hat{\Omega}_0^{-1}\right|\right\} = T\left\{\log\left|\hat{\Omega}_0\right| - \log\left|\hat{\Omega}_1\right|\right\}$$

$$LR \to \chi_m^2 \qquad m \equiv \text{number of restrictions} = n^2(p_1 - p_0)$$

each equation has $p_1 - p_0$ restriction on each variable $\rightarrow$

$n(p_1 - p_0)$ in each equation

SVAR: Testing – Asymptotic Distribution

• In general, linear hypotheses can be tested directly as usual and their asymptotic distribution follows from the next asymptotic result:

Let $\hat{\pi}_T = vec(\hat{\Pi}_T)$ denote the (nk × 1) (with k=1+np number of parameters estimated per equation) vector of coef. resulting from OLS regressions of each of the elements of y_t on x_t for a sample of size T:

$$\hat{\pi}_T = \begin{bmatrix} \hat{\pi}_{1.T} \\ \cdot \\ \hat{\pi}_{n.T} \end{bmatrix}, \text{ where } \hat{\pi}_{iT} = \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1} \left[\sum_{t=1}^{T} x_t y_{it}\right]$$

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

$$\sqrt{T}(\hat{\pi}_T - \pi) \to N(0, (\Omega \otimes M^{-1})), \text{ and the coef of regression i}$$

$$\sqrt{T}(\hat{\pi}_{iT} - \pi_i) \to N(0, \sigma_i^2 M^{-1}) \quad \text{with } \hat{M} = p\lim(1/T)\sum_t X_t X'_t$$

SVAR: Identification with IC

• In the same way as in the univariate AR(p) models, Information Criteria (IC) can be used to choose the p in a VAR:

$$\text{AIC} = \ln|\boldsymbol{\Omega}| + \frac{2(n^2p + n)}{T}$$

$$\text{SBC} = \ln|\boldsymbol{\Omega}| + \frac{(n^2p + n)\ln(T)}{T}$$

• Similar consistency and efficiency results to the ones obtained in the univariate world apply here.

• The main difference is that as the number of variables gets bigger, it is more *unlikely* that the AIC ends up *overparametrizing* --see Gonzalo and Pitarakis (2002).

SVAR: Granger Causality – Testing

• After selecting the lag structure for the VAR(p) –i.e., assuming a "correct" lag length p–, test for Granger causality, as usual.. Then,

$$x_t = \mu + \alpha_1\, x_{t-1} + \alpha_2\, x_{t-2} + \ldots + \alpha_p\, x_{t-p} +$$
$$+ \beta_1\, y_{t-1} + \beta_2\, y_{t-2} + \ldots + \beta_p\, y_{t-p} + \varepsilon_t$$

• Estimate model by OLS and test for the following hypothesis

H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ (y_t does not Granger-cause x_t)

H_1: any $\beta_i \neq 0$.

• Get RSS for the restricted and unrestricted model. Then, calculate the F-test:

$F = [(T\text{-}2p\text{-}1)/p] * [(RSS_R - RSS_U)/RSS_U]$

SVAR: IRF

• Goal: We want to study the reaction of a VAR(p) system to a shock.

$$\boldsymbol{\Phi}(L)\, \boldsymbol{Y}_t = \boldsymbol{\Phi}_0 + a_t.$$

Assuming the system is stable, we move to an MA representation:,

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\, \boldsymbol{a}_t$$

where

$$\boldsymbol{\Psi}(L) = [\boldsymbol{\Phi}(L)]^{-1}$$

Writing the system at time $t+s$:

$$\boldsymbol{Y}_{t+s} = \boldsymbol{\mu} + \boldsymbol{a}_t + \boldsymbol{\Psi}_1\, \boldsymbol{a}_{t+s-1} + \boldsymbol{\Psi}_2\, \boldsymbol{a}_{t+s-2} + \cdots + \boldsymbol{\Psi}_s\, \boldsymbol{a}_t + \cdots$$

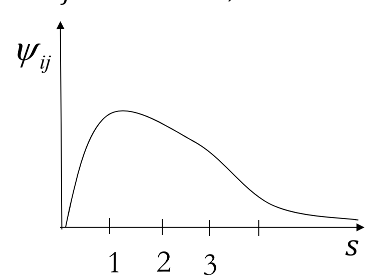
Then,

$$\frac{\partial Y_{t+s}}{\partial a'_t} = \Psi_s = \underset{m \text{ x } m}{\left[\psi_{ij}^{(s)}\right]} \qquad \text{(multipliers)}$$

$$\frac{\partial y_{i,t+s}}{\partial a_{jt}} = \psi_{ij}^{(s)} \longrightarrow$$ Reaction of the i-variable to a unit change in innovation j.

SVAR: IRF

• Impulse-response function: The response of $y_{i,t+s}$ to one-time impulse in $y_{j,t}$, given by $a_{j,t}$, with all other variables dated t or earlier held constant. (Usually, the size of the shock is in SD units, for example, $a_{j,t} = k\sigma_j$ with $k > 0$.)



where

$$\frac{\partial y_{i,t+s}}{\partial a_{jt}} = \psi_{ij}$$

SVAR: IRF – Example

Example: We have a stable VAR(1) model:

$$\boldsymbol{Y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

where

$$\boldsymbol{\Sigma}_a = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

• We start at a point of equilibrium ($t < 0$). Then, at $t = 0$ we shock one variable, y_{2t}, in the VAR(1) system, by creating a $a_{2t=0} = 1$.

Assume there is no additional shock after $t = 0$.

In our example,
$t < 0$ $\quad y_{1t} = y_{2t} = 0$
$t = 0$ $\quad a_{10} = 0,\ \& \ a_{20} = 1 \ (\Rightarrow y_{20} = 1)$
$t > 0$ $\quad a_{1t} = 0,\ \& \ a_{2t} = 0$

SVAR: IRF – Example

Example (continuation):
Shock dynamics,
$t < 0$ $\quad y_{1t} = y_{2t} = 0$
$t = 0$ $\quad a_{10} = 0,\ \& \ a_{20} = 1 \ (\Rightarrow y_{20} = 1)$
$t > 0$ $\quad a_{1t} = 0,\ \& \ a_{2t} = 0$

• Reaction of the system

$$\begin{bmatrix} y_{10} \\ y_{20} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (impulse)$$

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

$$\begin{bmatrix} y_{12} \\ y_{22} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{21} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\vdots$$

$$\begin{bmatrix} y_{1s} \\ y_{2s} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^s \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \Phi_1^{\,s} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

SVAR: IRF – Orthogonal shocks

• If we work with the MA representation:

$$\boldsymbol{\Psi}(L) = [\boldsymbol{\Phi}(L)]^{-1}$$

with
$$\boldsymbol{\Psi}_1 = \boldsymbol{\Phi}_1$$
$$\boldsymbol{\Psi}_2 = \boldsymbol{\Phi}_1{}^2$$
$$\vdots$$
$$\boldsymbol{\Psi}_s = \Phi_1{}^s$$

• In this example, the variance-covariance matrix of the innovations is not diagonal –i.e., $\sigma_{12} \neq 0$. There is contemporaneous correlation between shocks, then

$$\begin{bmatrix} y_{10} \\ y_{20} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightarrow \text{Not very realistic}$$

• To avoid this problem, the variance-covariance matrix has to be diagonalized (the shocks have to be **orthogonal**) and here is where the serious problems appear.

SVAR: IRF – Orthogonal shocks

• Reminder: **A** is a p.d. symmetric matrix. Then, **Q A**$^{-1}$**Q'** = **I**.

• Then, the MA representation:

$$Y_t = \mu + \sum_{i=0}^{\infty} \Psi_i a_{t-i} \qquad \Psi_0 = I_n$$

$$Y_t = \mu + \sum_{i=0}^{\infty} \Psi_i Q^{-1} Q a_{t-i}$$

Let us call $M_i = \Psi_i Q^{-1}; \quad w_t = Qa_t \rightarrow Y_t = \mu + \sum_{i=0}^{\infty} M_i w_{t-i}$

$$E[w_t w'_t] = E[Qa_t a'_t Q'] = QE[a_t a'_t]Q' = Q\Sigma Q' = I_n$$

w_t has components that are all uncorrelated and unit variance

• Orthogonalized IRF: $\dfrac{\partial Y_{t+s}}{\partial w_t} = M_s = \Psi_s Q^{-1}$

<u>Problem</u>: $\boldsymbol{Q}$ is not unique.

SVAR – Variance Decomposition

• Contribution of the j-*th* orthogonalized innovation to the MSE of the s-period ahead forecast

$$MSE(\hat{Y}_t(s)) = E(Y_{t+s} - \hat{Y}_t(s))(Y_{t+s} - \hat{Y}_t(s))'$$
$$e_t(s) = Y_{t+s} - \hat{Y}_t(s) = a_{t+s} + \Psi_1 a_{t+s-1} + \ldots..\Psi_{s-1} a_{t+1}$$
$$E[e_t(s)e_t(s)'] = \Omega_a + \Psi_1 \Omega_a \Psi_1' + \ldots. + \Psi_{s-1}\Omega_a \Psi_{s-1}'$$
$$MSE(s) = Q^{-1}Q\Omega_a Q' Q^{-1'} + \Psi_1 Q^{-1} Q\Omega_a Q' Q^{-1'}\Psi_1' + \ldots.$$
$$+\Psi_{s-1}Q^{-1}Q\Omega_a Q' Q^{-1'}\Psi_{s-1}' =$$
$$= Q^{-1}Q^{-1'} + \Psi_1 Q^{-1}Q^{-1'}\Psi_1' + \ldots\ldots\Psi_{s-1}Q^{-1}Q^{-1'}\Psi_{s-1}' =$$
$$= M_0 M_0' + M_1 M_1' + \ldots\ldots\ldots M_{s-1}M'_{s-1}$$

recall that $M_i = \Psi_i Q^{-1}$ and $M_0 = Q^{-1}, \Psi_0 = I$

$\downarrow$

• Contribution of the first orthogonalized innovation to the MSE. (Do this for a two variables VAR model!)

SVAR: Variance Decomposition

Example: Variance decomposition in a two variables (y_t, x_t) VAR

- The s-step ahead forecast error for variable y_t is:

$$y_{t+s} - E_t y_{t+s} = M_0(1,1)\varepsilon_{yt+s} + M_1(1,1)\varepsilon_{yt+s-1} + \ldots + M_{s-1}(1,1)\varepsilon_{yt+1} +$$
$$M_0(1,2)\varepsilon_{xt+s} + M_1(1,2)\varepsilon_{xt+s-1} + \ldots + M_{s-1}(1,2)\varepsilon_{xt+1}$$

- Denote the variance of the s-step ahead forecast error variance of y_{t+s} as for $\sigma_y(s)^2$:

$$\sigma_y(s)^2 = \sigma_y^2[M_0(1,1)^2 + M_1(1,1)^2 + \ldots + M_{s-1}(1,1)^2] +$$
$$\sigma_x^2[M_0(1,2)^2 + M_1(1,2)^2 + \ldots + M_{s-1}(1,2)^2]$$

- The forecast error variance decompositions are proportions of $\sigma_y(s)^2$.

SVAR: Relative Importance of Variables

- The forecast error variance decompositions are proportions of $\sigma_y(s)^2$.

$$\text{due to shocks to } y = \sigma_y^2[M_0(1,1)^2 + M_1(1,1)^2 + ... + M_{s-1}(1,1)^2]/\sigma_y(s)^2$$
$$\text{due to shocks to } x = \sigma_x^2[M_0(1,2)^2 + M_1(1,2)^2 + ... + M_{s-1}(1,2)^2]/\sigma_y(s)^2$$

Note: We can use these proportions to measure relative importance of a variable in the system.

Q: Which markets have more relevance (shocks are more relevant to the system)? Are stock or options markets more important?

SVAR: Identification in VAR(1)

• We started with a SVAR model, and transformed into the reduced form or standard VAR for estimation purposes.

• Q: Is it possible to recover the parameters in the SVAR from the estimated parameters in the standard VAR? No!!

• There are **10** parameters in the bivariate SVAR(1) and only **9** estimated parameters in the standard VAR(1). Thus, the VAR is *underidentified.* We need restrictions (exclusion, linear restrictions, etc.)

• If one parameter in the bivariate SVAR above is restricted, say excluded, (order condition met!), the VAR is exactly identified.

• Sims (1980) suggests a recursive system to identify the model letting $b_{21} = 0$.

$$\begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

SVAR: Identification in VAR(1)

• $b_{21} = 0$ implies

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

• The recursive model imposes the restriction that the value y_t does not have a contemporaneous effect on x_t.

• Now, the parameters of the SVAR are identified -we have 9 equations.

$$\phi_{10} = b_{10} - b_{12}b_{20} \quad \phi_{20} = b_{20} \quad \text{var}(e_1) = \sigma_y^2 + b_{12}^2\sigma_x^2$$
$$\phi_{11} = \gamma_{11} - b_{12}\gamma_{21} \quad \phi_{21} = \gamma_{21} \quad \text{var}(e_2) = \sigma_x^2$$
$$\phi_{12} = \gamma_{12} - b_{12}\gamma_{22} \quad \phi_{22} = \gamma_{22} \quad \text{cov}(e_1, e_2) = -b_{12}\sigma_x^2$$

SVAR: Identification in VAR(1)

• Note both structural shocks can now be identified from the residuals of the standard VAR.

• $b_{21} = 0$ implies $\boldsymbol{y_t}$ does not have a contemporaneous effect on $\boldsymbol{x_t}$.

• This restriction manifests itself such that both ε_{yt} & ε_{xt} affect $\boldsymbol{y_t}$ contemporaneously but only ε_{xt} affects $\boldsymbol{x_t}$ contemporaneously.

• The residuals of e_{xt} are due to pure shocks to $\boldsymbol{x_t}$.

• The SVAR representation based on a recursive causal ordering may be computed using the **Choleski factorization** of $\boldsymbol{\Omega}$.

• Other restriction could have been used. For example, $b_{12} + b_{21} = 1$.

• There are other methods used to identify models, like the Blanchard and Quah (1989) decomposition.

SVAR: Criticisms

• A VAR model can be a good forecasting model, but in a sense it is an atheoretical model (as all the reduced form models are).

• To calculate the IRF, the order matters: **Q** is not unique. The Cholesky decomposition (**Q** is lower triangular) imposes an order in the recursive causal structure of the VAR. It is not a trivial issue.

• Sensitive to the lag selection.

• Dimensionality problem.

SVAR: Sensitive Analysis

• Ordering of the variables in $\boldsymbol{Y_t}$ determines the recursive causal structure of the structural VAR.

• This identification assumption is not testable.

• Sensitivity analysis is performed to determine how the structural analysis based on the IRFs and FEVDs (forecast error variance decomposition) are influenced by the assumed causal ordering.

• This sensitivity analysis is based on estimating the structural VAR for different orderings of the variables.

• If the IRFs and FEVDs change considerably for different orderings of the variables in $\boldsymbol{Y_t}$, then it is clear that the assumed recursive causal structure heavily influences the structural inference.