

# Multivariate Time Series: Part 1

## Introduction

Gerald P. Dwyer

Clemson University

February 2021

# Outline

## 1 Multivariate Time Series Introduction

- Representation and Identification
- Normalization of Vector Moving Averages
- Normalization of Vector Autoregressions
- Example of Importance of “Ordering”
- Generalized impulse response functions
- Summary

## Vector autoregression

- First-order vector autoregression with  $n$  variables

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- I will indicate vectors and matrices by putting them in bold text
- Variables are in  $\mathbf{y}'_t = [y_{1,t}, y_{2,t}, \dots, y_{n,t}]$
- $\mathbf{y}_t$  is an  $n \times 1$  vector
- $\mathbf{A}_0$  is an  $n \times 1$  vector as well
- $\mathbf{A}_1$  must be  $n \times n$  since  $\mathbf{y}_{t-1}$  is an  $n \times 1$  vector and  $\mathbf{A}_1$  is post-multiplied by  $\mathbf{y}_{t-1}$
- Therefore  $\mathbf{A}_1 \mathbf{y}_{t-1}$  is  $n \times 1$
- And  $\boldsymbol{\varepsilon}_t$  is  $n \times 1$
- $\boldsymbol{\varepsilon}_t$  has zero mean, constant variance and is serially uncorrelated

## Vector autoregression written out

- Written out explicitly, this first-order autoregression is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \dots \\ y_{n,t} \end{bmatrix} = \begin{bmatrix} a_{1,0} \\ a_{2,0} \\ \dots \\ a_{n,0} \end{bmatrix} + \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ a_{2,1} & \dots & a_{2,n} \\ \dots & \dots & \dots \\ a_{n,1} & \dots & a_{n,n} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \dots \\ y_{n,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \dots \\ \varepsilon_{n,t} \end{bmatrix}$$

## Vector autoregression written out

- Written out in terms of equations, this first-order autoregression is

$$y_{1,t} = a_{1,0} + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \dots + a_{1,n}y_{n,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,0} + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \dots + a_{2,n}y_{n,t-1} + \varepsilon_{2,t}$$

...

$$y_{n,t} = a_{n,0} + a_{n,1}y_{1,t-1} + a_{n,2}y_{2,t-1} + \dots + a_{n,n}y_{n,t-1} + \varepsilon_{n,t}$$

## Vector autoregression written out

- Written out in terms of equations, this first-order autoregression is

$$y_{1,t} = a_{1,0} + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \dots + a_{1,n}y_{n,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,0} + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \dots + a_{2,n}y_{n,t-1} + \varepsilon_{2,t}$$

...

$$y_{n,t} = a_{n,0} + a_{n,1}y_{1,t-1} + a_{n,2}y_{2,t-1} + \dots + a_{n,n}y_{n,t-1} + \varepsilon_{n,t}$$

- Everything depends on lagged values of everything

## With more than one lag

- Vector autoregression with more lags

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

## With more than one lag

- Vector autoregression with more lags

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- where  $\mathbf{A}(L)$  is a matrix of lag polynomials of whatever length of lags are desired



## With more than one lag

- Vector autoregression with more lags

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- where  $\mathbf{A}(L)$  is a matrix of lag polynomials of whatever length of lags are desired
- Can also write it as

$$\mathbf{y}_t = \mathbf{A}_0 + \sum_{i=1}^k \mathbf{A}_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t$$

## With more than one lag

- Vector autoregression with more lags

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- where  $\mathbf{A}(L)$  is a matrix of lag polynomials of whatever length of lags are desired
- Can also write it as

$$\mathbf{y}_t = \mathbf{A}_0 + \sum_{i=1}^k \mathbf{A}_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t$$

- Suppose

$$E \boldsymbol{\varepsilon}_t = \mathbf{0}$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' = \boldsymbol{\Sigma},$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s' = \mathbf{0} \quad \forall t \neq s$$

## With more than one lag

- Vector autoregression with more lags

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- where  $\mathbf{A}(L)$  is a matrix of lag polynomials of whatever length of lags are desired
- Can also write it as

$$\mathbf{y}_t = \mathbf{A}_0 + \sum_{i=1}^k \mathbf{A}_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t$$

- Suppose

$$E \boldsymbol{\varepsilon}_t = \mathbf{0}$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' = \boldsymbol{\Sigma},$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s' = \mathbf{0} \quad \forall t \neq s$$

- This is a very general representation in terms of the relationships of series

## Fundamental representation

- As with the univariate case, there is a version of Wold's theorem for multivariate time series

## Fundamental representation

- As with the univariate case, there is a version of Wold's theorem for multivariate time series
- Suppose that  $\mathbf{y}_t$  is a set of series that is purely indeterministic and covariance stationary

## Fundamental representation

- As with the univariate case, there is a version of Wold's theorem for multivariate time series
- Suppose that  $\mathbf{y}_t$  is a set of series that is purely indeterministic and covariance stationary
- Then there exists a representation

$$\mathbf{y}_t = \mathbf{B}(L) \boldsymbol{\varepsilon}_t$$

where  $\mathbf{B}(L)$  is a matrix of lag polynomials of arbitrary length with

$$E \boldsymbol{\varepsilon}_t = \mathbf{0}$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' = \boldsymbol{\Sigma}$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s' = \mathbf{0} \quad \forall t \neq s$$

## Fundamental representation

- As with the univariate case, there is a version of Wold's theorem for multivariate time series
- Suppose that  $\mathbf{y}_t$  is a set of series that is purely indeterministic and covariance stationary
- Then there exists a representation

$$\mathbf{y}_t = \mathbf{B}(L) \boldsymbol{\varepsilon}_t$$

where  $\mathbf{B}(L)$  is a matrix of lag polynomials of arbitrary length with

$$E \boldsymbol{\varepsilon}_t = \mathbf{0}$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' = \boldsymbol{\Sigma}$$

$$E \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s' = \mathbf{0} \quad \forall t \neq s$$

- Can also write this as

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{B}_i \boldsymbol{\varepsilon}_{t-i}$$

## MA and AR representation

- How are the moving average and autoregressive representations related?



## MA and AR representation

- How are the moving average and autoregressive representations related?
- The moving-average representation from Wold's Theorem with a constant term is

$$\mathbf{y}_t = \mathbf{B}_0 + \mathbf{B}(L) \boldsymbol{\varepsilon}_t$$

## MA and AR representation

- How are the moving average and autoregressive representations related?
- The moving-average representation from Wold's Theorem with a constant term is

$$\mathbf{y}_t = \mathbf{B}_0 + \mathbf{B}(L) \boldsymbol{\varepsilon}_t$$

- The autoregressive representation is

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

## MA and AR representation

- How are the moving average and autoregressive representations related?
- The moving-average representation from Wold's Theorem with a constant term is

$$\mathbf{y}_t = \mathbf{B}_0 + \mathbf{B}(L) \boldsymbol{\varepsilon}_t$$

- The autoregressive representation is

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- If  $\mathbf{B}(L)$  is invertible, then

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}_0 + \mathbf{B}(L) \boldsymbol{\varepsilon}_t \\ \mathbf{B}(L)^{-1} \mathbf{y}_t &= \mathbf{B}(L)^{-1} \mathbf{B}_0 + \boldsymbol{\varepsilon}_t \end{aligned}$$

## MA and AR representation

- How are the moving average and autoregressive representations related?
- The moving-average representation from Wold's Theorem with a constant term is

$$\mathbf{y}_t = \mathbf{B}_0 + \mathbf{B}(L) \boldsymbol{\varepsilon}_t$$

- The autoregressive representation is

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- If  $\mathbf{B}(L)$  is invertible, then

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}_0 + \mathbf{B}(L) \boldsymbol{\varepsilon}_t \\ \mathbf{B}(L)^{-1} \mathbf{y}_t &= \mathbf{B}(L)^{-1} \mathbf{B}_0 + \boldsymbol{\varepsilon}_t \end{aligned}$$

- After some manipulation, this can be written as the autoregressive representation

## MA and AR representation

- The MA equation can be written

$$\mathbf{B}(L)^{-1} \mathbf{y}_t = \mathbf{B}(L)^{-1} \mathbf{B}_0 + \varepsilon_t$$

## MA and AR representation

- The MA equation can be written

$$\mathbf{B}(L)^{-1} \mathbf{y}_t = \mathbf{B}(L)^{-1} \mathbf{B}_0 + \varepsilon_t$$

- What is  $\mathbf{B}(L)^{-1}$ ?

## MA and AR representation

- The MA equation can be written

$$\mathbf{B}(L)^{-1} \mathbf{y}_t = \mathbf{B}(L)^{-1} \mathbf{B}_0 + \varepsilon_t$$

- What is  $\mathbf{B}(L)^{-1}$ ?
- It is the matrix inverse of  $\mathbf{B}(L)$

## MA and AR representation

- The MA equation can be written

$$\mathbf{B}(L)^{-1} \mathbf{y}_t = \mathbf{B}(L)^{-1} \mathbf{B}_0 + \varepsilon_t$$

- What is  $\mathbf{B}(L)^{-1}$ ?
- It is the matrix inverse of  $\mathbf{B}(L)$
- Simplify by looking at univariate case with one lag

$$B(L)^{-1} = \frac{1}{1 + b_1 L}$$



## MA and AR representation

- The MA equation can be written

$$\mathbf{B}(L)^{-1} \mathbf{y}_t = \mathbf{B}(L)^{-1} \mathbf{B}_0 + \varepsilon_t$$

- What is  $\mathbf{B}(L)^{-1}$ ?
- It is the matrix inverse of  $\mathbf{B}(L)$
- Simplify by looking at univariate case with one lag

$$B(L)^{-1} = \frac{1}{1 + b_1 L}$$

- What does this equation represent?

## Inverse of polynomial in lag operator

- What is  $\mathbf{B}(L)^{-1}$ ?

## Inverse of polynomial in lag operator

- What is  $\mathbf{B}(L)^{-1}$ ?
- For the univariate case with one lag

$$B(L)^{-1} = \frac{1}{1 + b_1 L}$$

## Inverse of polynomial in lag operator

- What is  $\mathbf{B}(L)^{-1}$ ?
- For the univariate case with one lag

$$B(L)^{-1} = \frac{1}{1 + b_1 L}$$

- Know for a geometric series with an initial value of one and a common ratio  $r$  that

$$\frac{1}{1 - r} = \sum_{i=0}^{\infty} r^i$$

## Inverse of polynomial in lag operator

- What is  $\mathbf{B}(L)^{-1}$ ?
- For the univariate case with one lag

$$B(L)^{-1} = \frac{1}{1 + b_1 L}$$

- Know for a geometric series with an initial value of one and a common ratio  $r$  that

$$\frac{1}{1 - r} = \sum_{i=0}^{\infty} r^i$$

- Therefore

$$B(L)^{-1} = \sum_{i=0}^{\infty} (-b_1 L)^i$$

## Inverse of polynomial in lag operator

- Therefore, for an MA(1), the autoregressive representation is given by

$$y_t = b_0 + \varepsilon_t + b_1 \varepsilon_{t-1}$$

$$\sum_{i=0}^{\infty} (-b_1 L)^i y_t = \sum_{i=0}^{\infty} (-b_1 L)^i b_0 + \varepsilon_t$$

$$y_t = b_0 \sum_{i=0}^{\infty} (-b_1)^i - \sum_{i=1}^{\infty} (-b_1)^i y_{t-i} + \varepsilon_t$$

## Inverse of polynomial in lag operator

- Therefore, for an MA(1), the autoregressive representation is given by

$$y_t = b_0 + \varepsilon_t + b_1 \varepsilon_{t-1}$$

$$\sum_{i=0}^{\infty} (-b_1 L)^i y_t = \sum_{i=0}^{\infty} (-b_1 L)^i b_0 + \varepsilon_t$$

$$y_t = b_0 \sum_{i=0}^{\infty} (-b_1)^i - \sum_{i=1}^{\infty} (-b_1)^i y_{t-i} + \varepsilon_t$$

- This is a simple illustration of a general point

## Inverse of polynomial in lag operator

- Therefore, for an MA(1), the autoregressive representation is given by

$$y_t = b_0 + \varepsilon_t + b_1 \varepsilon_{t-1}$$

$$\sum_{i=0}^{\infty} (-b_1 L)^i y_t = \sum_{i=0}^{\infty} (-b_1 L)^i b_0 + \varepsilon_t$$

$$y_t = b_0 \sum_{i=0}^{\infty} (-b_1)^i - \sum_{i=1}^{\infty} (-b_1)^i y_{t-i} + \varepsilon_t$$

- This is a simple illustration of a general point
- The general MA representation can generate an AR representation with serially uncorrelated errors



## Inverse of polynomial in lag operator

- Therefore, for an MA(1), the autoregressive representation is given by

$$y_t = b_0 + \varepsilon_t + b_1 \varepsilon_{t-1}$$

$$\sum_{i=0}^{\infty} (-b_1 L)^i y_t = \sum_{i=0}^{\infty} (-b_1 L)^i b_0 + \varepsilon_t$$

$$y_t = b_0 \sum_{i=0}^{\infty} (-b_1)^i - \sum_{i=1}^{\infty} (-b_1)^i y_{t-i} + \varepsilon_t$$

- This is a simple illustration of a general point
- The general MA representation can generate an AR representation with serially uncorrelated errors
- In general, requires an infinite number of lags in the autoregressive representation

## Normalization issues arise

- In the univariate case, we assume that the coefficient of the variable's current own innovation is unity
- Why? Suppose we didn't. Moving-average representation is

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

- Suppose that all  $\psi_i = 0 \forall i > 0$ . So

$$y_t = \psi_0 \varepsilon_t$$

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$
- $\text{Var} [y_t] = \psi_0^2 \text{Var} [\varepsilon_t]$  which is fine

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$
- $\text{Var} [y_t] = \psi_0^2 \text{Var} [\varepsilon_t]$  which is fine
  - ▶  $\text{Var} [y_t]$  is observable

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$
- $\text{Var} [y_t] = \psi_0^2 \text{Var} [\varepsilon_t]$  which is fine
  - ▶  $\text{Var} [y_t]$  is observable
- We can set  $\psi_0 = 1$  and then  $\text{Var} [\varepsilon_t] = \text{Var} [y_t]$



## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$
- $\text{Var} [y_t] = \psi_0^2 \text{Var} [\varepsilon_t]$  which is fine
  - ▶  $\text{Var} [y_t]$  is observable
- We can set  $\psi_0 = 1$  and then  $\text{Var} [\varepsilon_t] = \text{Var} [y_t]$
- We could alternatively set  $\text{Var} [\varepsilon_t] = 1$  and then  $\psi_0^2 = \text{Var} [y_t]$

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$
- $\text{Var} [y_t] = \psi_0^2 \text{Var} [\varepsilon_t]$  which is fine
  - ▶  $\text{Var} [y_t]$  is observable
- We can set  $\psi_0 = 1$  and then  $\text{Var} [\varepsilon_t] = \text{Var} [y_t]$
- We could alternatively set  $\text{Var} [\varepsilon_t] = 1$  and then  $\psi_0^2 = \text{Var} [y_t]$
- We could set  $\psi_0^2 = 0.5$  and then  $\text{Var} [\varepsilon_t] = 2 \text{Var} [y_t]$

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$
- $\text{Var}[y_t] = \psi_0^2 \text{Var}[\varepsilon_t]$  which is fine
  - ▶  $\text{Var}[y_t]$  is observable
- We can set  $\psi_0 = 1$  and then  $\text{Var}[\varepsilon_t] = \text{Var}[y_t]$
- We could alternatively set  $\text{Var}[\varepsilon_t] = 1$  and then  $\psi_0^2 = \text{Var}[y_t]$
- We could set  $\psi_0^2 = 0.5$  and then  $\text{Var}[\varepsilon_t] = 2 \text{Var}[y_t]$
- Either  $\psi_0$  or  $\text{Var}[\varepsilon_t]$  is not identified by the single number  $\text{Var}[y_t]$

## Normalization in moving average

- In the univariate case, we assume that the coefficient of the variable's own innovation is unity
- Suppose we have the general representation

$$y_t = \psi_0 \varepsilon_t$$

- Then  $E y_t = 0$  because  $E \varepsilon_t = 0$
- $\text{Var}[y_t] = \psi_0^2 \text{Var}[\varepsilon_t]$  which is fine
  - ▶  $\text{Var}[y_t]$  is observable
- We can set  $\psi_0 = 1$  and then  $\text{Var}[\varepsilon_t] = \text{Var}[y_t]$
- We could alternatively set  $\text{Var}[\varepsilon_t] = 1$  and then  $\psi_0^2 = \text{Var}[y_t]$
- We could set  $\psi_0^2 = 0.5$  and then  $\text{Var}[\varepsilon_t] = 2 \text{Var}[y_t]$
- Either  $\psi_0$  or  $\text{Var}[\varepsilon_t]$  is not identified by the single number  $\text{Var}[y_t]$
- Setting  $\psi_0 = 1$  generally most “natural” but it is arbitrary

## Normalization in vector moving average

- In the multivariate case, the representation is

$$\mathbf{y}_t = \mathbf{B}(L) \boldsymbol{\varepsilon}_t$$

- Take simpler two-variable case

$$y_{1,t} = \sum_{i=0}^{\infty} \psi_{11,i} \varepsilon_{1,t-i} + \sum_{i=0}^{\infty} \psi_{12,i} \varepsilon_{2,t-i}$$
$$y_{2,t} = \sum_{i=0}^{\infty} \psi_{21,i} \varepsilon_{1,t-i} + \sum_{i=0}^{\infty} \psi_{22,i} \varepsilon_{2,t-i}$$

- Nothing fundamental will be lost if we simplify to no lags of each innovation

$$y_{1,t} = \psi_{11,0} \varepsilon_{1,t} + \psi_{12,0} \varepsilon_{2,t}$$

$$y_{2,t} = \psi_{21,0} \varepsilon_{1,t} + \psi_{22,0} \varepsilon_{2,t}$$

## Normalization of zero-order vector moving average

- What are the observable summary statistics?

$$E y_1 \text{ and } E y_2$$

- These just identify constant terms and we have none so

$$E y_1 = E y_2 = 0$$

- In terms of second moments, we have

$$\text{Var} [y_1], \text{Var} [y_2] \text{ and } \text{Cov} [y_1, y_2]$$

- By analogy with the univariate case, we might expect that we can identify three parameters and others can be determined by normalizations

## Normalization of zero-order vector moving average

- It seems plausible that we may be able to identify three parameters from the data on three moments of variable
- What parameters do we have?

$$y_{1,t} = \psi_{11}\varepsilon_{1,t} + \psi_{12}\varepsilon_{2,t}$$

$$y_{2,t} = \psi_{21}\varepsilon_{1,t} + \psi_{22}\varepsilon_{2,t}$$

- We have

$$\psi_{11} \quad \psi_{12} \quad \psi_{21} \quad \psi_{22}$$

and

$$\text{Var}[\varepsilon_1], \text{Var}[\varepsilon_2] \text{ and } \text{Cov}[\varepsilon_1, \varepsilon_2]$$

- Seven parameters and three moments with which we might identify them

## Seven parameters and three moments

- The equations are

$$y_{1,t} = \psi_{11}\varepsilon_{1,t} + \psi_{12}\varepsilon_{2,t}$$

$$y_{2,t} = \psi_{21}\varepsilon_{1,t} + \psi_{22}\varepsilon_{2,t}$$

- The variances and covariances in terms of this moving-average representation are

$$\text{Var} [y_1] = \psi_{11}^2 \text{Var} [\varepsilon_1] + \psi_{12}^2 \text{Var} [\varepsilon_2] + 2\psi_{11}\psi_{12} \text{Cov} [\varepsilon_1, \varepsilon_2]$$

$$\text{Var} [y_2] = \psi_{21}^2 \text{Var} [\varepsilon_1] + \psi_{22}^2 \text{Var} [\varepsilon_2] + 2\psi_{21}\psi_{22} \text{Cov} [\varepsilon_1, \varepsilon_2]$$

$$\begin{aligned} \text{Cov} [y_1, y_2] &= \psi_{11}\psi_{21} \text{Var} [\varepsilon_1] + \psi_{12}\psi_{22} \text{Var} [\varepsilon_2] \\ &\quad + (\psi_{11}\psi_{22} + \psi_{12}\psi_{21}) \text{Cov} [\varepsilon_1, \varepsilon_2] \end{aligned}$$

- It seems plausible that we may be able to identify three parameters from the data on the three moments



## What normalizations and how can they be interpreted?

- The variances and covariances in terms of this moving-average representation are

$$\text{Var} [y_1] = \psi_{11}^2 \text{Var} [\varepsilon_1] + \psi_{12}^2 \text{Var} [\varepsilon_2] + 2\psi_{11}\psi_{12} \text{Cov} [\varepsilon_1, \varepsilon_2]$$

$$\text{Var} [y_2] = \psi_{21}^2 \text{Var} [\varepsilon_1] + \psi_{22}^2 \text{Var} [\varepsilon_2] + 2\psi_{21}\psi_{22} \text{Cov} [\varepsilon_1, \varepsilon_2]$$

$$\begin{aligned} \text{Cov} [y_1, y_2] &= \psi_{11}\psi_{21} \text{Var} [\varepsilon_1] + \psi_{12}\psi_{22} \text{Var} [\varepsilon_2] \\ &\quad + (\psi_{11}\psi_{22} + \psi_{12}\psi_{21}) \text{Cov} [\varepsilon_1, \varepsilon_2] \end{aligned}$$

- Set  $\psi_{11} = \psi_{22} = 1$ . This implies

$$\text{Var} [y_1] = \text{Var} [\varepsilon_1] + \psi_{12}^2 \text{Var} [\varepsilon_2] + 2\psi_{12} \text{Cov} [\varepsilon_1, \varepsilon_2]$$

$$\text{Var} [y_2] = \psi_{21}^2 \text{Var} [\varepsilon_1] + \text{Var} [\varepsilon_2] + 2\psi_{21} \text{Cov} [\varepsilon_1, \varepsilon_2]$$

$$\begin{aligned} \text{Cov} [y_1, y_2] &= \psi_{21} \text{Var} [\varepsilon_1] + \psi_{12} \text{Var} [\varepsilon_2] \\ &\quad + (1 + \psi_{12}\psi_{21}) \text{Cov} [\varepsilon_1, \varepsilon_2] \end{aligned}$$

- Still have three moments and five parameters

## One normalization

- Equations are

$$y_{1,t} = \psi_{11}\varepsilon_{1,t} + \psi_{12}\varepsilon_{2,t}$$

$$y_{2,t} = \psi_{21}\varepsilon_{1,t} + \psi_{22}\varepsilon_{2,t}$$

- One possible setup  $\psi_{11} = \psi_{22} = 1$  and  $\psi_{12} = \psi_{21} = 0$  in

$$\text{Var}[y_1] = \psi_{11}^2 \text{Var}[\varepsilon_1] + \psi_{12}^2 \text{Var}[\varepsilon_2] + 2\psi_{11}\psi_{12} \text{Cov}[\varepsilon_1, \varepsilon_2]$$

$$\text{Var}[y_2] = \psi_{21}^2 \text{Var}[\varepsilon_1] + \psi_{22}^2 \text{Var}[\varepsilon_2] + 2\psi_{21}\psi_{22} \text{Cov}[\varepsilon_1, \varepsilon_2]$$

$$\begin{aligned} \text{Cov}[y_1, y_2] &= \psi_{11}\psi_{21} \text{Var}[\varepsilon_1] + \psi_{12}\psi_{22} \text{Var}[\varepsilon_2] \\ &\quad + (\psi_{11}\psi_{22} + \psi_{12}\psi_{21}) \text{Cov}[\varepsilon_1, \varepsilon_2] \end{aligned}$$

which implies

$$\text{Var}[y_1] = \text{Var}[\varepsilon_1]$$

$$\text{Var}[y_2] = \text{Var}[\varepsilon_2]$$

$$\text{Cov}[y_1, y_2] = \text{Cov}[\varepsilon_1, \varepsilon_2]$$

## A second normalization

- A second normalization
- Equations are

$$y_{1,t} = \psi_{11}\eta_{1,t} + \psi_{12}\eta_{2,t}$$

$$y_{2,t} = \psi_{21}\eta_{1,t} + \psi_{22}\eta_{2,t}$$

- ▶ I use a different symbol for the innovations because these innovations may be different than the ones in the normalization above
- ▶ I use the same symbol for the coefficients in the moving-average representation but they will have different values
- Another possible setup  $\psi_{11} = \psi_{22} = 1$  and  $\psi_{21} = 0$  and  $\text{Cov}[\eta_1, \eta_2] = 0$
- This implies

$$y_{1,t} = \eta_{1,t} + \psi_{12}\eta_{2,t}$$

$$y_{2,t} = \eta_{2,t}$$

## A second normalization

- The restrictions  $\psi_{11} = \psi_{22} = 1$  and  $\psi_{21} = 0$  and  $\text{Cov}[\eta_1, \eta_2] = 0$  imply in

$$\text{Var}[y_1] = \psi_{11}^2 \text{Var}[\eta_1] + \psi_{12}^2 \text{Var}[\eta_2] + 2\psi_{11}\psi_{12} \text{Cov}[\eta_1, \eta_2]$$

$$\text{Var}[y_2] = \psi_{21}^2 \text{Var}[\eta_1] + \psi_{22}^2 \text{Var}[\eta_2] + 2\psi_{21}\psi_{22} \text{Cov}[\eta_1, \eta_2]$$

$$\begin{aligned} \text{Cov}[y_1, y_2] &= \psi_{11}\psi_{21} \text{Var}[\eta_1] + \psi_{12}\psi_{22} \text{Var}[\eta_2] \\ &\quad + (\psi_{11}\psi_{22} + \psi_{12}\psi_{21}) \text{Cov}[\eta_1, \eta_2] \end{aligned}$$

that

$$\text{Var}[y_1] = \text{Var}[\eta_1] + \psi_{12}^2 \text{Var}[\eta_2]$$

$$\text{Var}[y_2] = \text{Var}[\eta_2]$$

$$\text{Cov}[y_1, y_2] = \psi_{12} \text{Var}[\eta_2]$$

## A second normalization

- In

$$\text{Var} [y_1] = \text{Var} [\eta_1] + \psi_{12}^2 \text{Var} [\eta_2]$$

$$\text{Var} [y_2] = \text{Var} [\eta_2]$$

$$\text{Cov} [y_1, y_2] = \psi_{12} \text{Var} [\eta_2]$$

- We can substitute  $\text{Var} [\eta_2] = \text{Var} [y_2]$  into the other two equations to get

$$\text{Var} [\eta_1] = \text{Var} [y_1] - \psi_{12}^2 \text{Var} [y_2]$$

$$\psi_{12} = \frac{\text{Cov} [y_1, y_2]}{\text{Var} [y_2]}$$

## A second normalization

- In

$$\begin{aligned}\text{Var} [\eta_1] &= \text{Var} [y_1] - \psi_{12}^2 \text{Var} [y_2] \\ \psi_{12} &= \frac{\text{Cov} [y_1, y_2]}{\text{Var} [y_2]}\end{aligned}$$

- the coefficient  $\psi_{12}$  is just the regression coefficient in the regression of  $y_1$  on  $y_2$
- It might seem that  $\text{Var} [\eta_1]$  could be negative, in which case this normalization is nonsense but

$$\begin{aligned}\text{Var} [\eta_1] &= \text{Var} [y_1] - \left( \frac{\text{Cov} [y_1, y_2]}{\text{Var} [y_2]} \right)^2 \text{Var} [y_2] \\ &= \text{Var} [y_1] - \frac{(\text{Cov} [y_1, y_2])^2}{\text{Var} [y_2]}\end{aligned}$$

## A second normalization

- The variance is positive

$$\text{Var}[\eta_1] = \text{Var}[y_1] - \frac{(\text{Cov}[y_1, y_2])^2}{\text{Var}[y_2]} > 0$$

if

$$\text{Var}[y_1] - \frac{(\text{Cov}[y_1, y_2])^2}{\text{Var}[y_2]} > 0$$

$$\text{Var}[y_1] > \frac{(\text{Cov}[y_1, y_2])^2}{\text{Var}[y_2]}$$

$$1 > \frac{(\text{Cov}[y_1, y_2])^2}{\text{Var}[y_1] \text{Var}[y_2]}$$

- So the variance is positive as long as the squared correlation is less than unity

## A second normalization

- In sum, the equations

$$y_{1,t} = \eta_{1,t} + \psi_{12}\eta_{2,t} \quad \text{Cov} [\eta_{1,t}, \eta_{2,t}] = 0$$

$$y_{2,t} = \eta_{2,t}$$

- are related to the underlying moments of the data by

$$\text{Var} [\eta_1] = \text{Var} [y_1] - \psi_{12}^2 \text{Var} [y_2]$$

$$\text{Var} [\eta_2] = \text{Var} [y_2]$$

$$\psi_{12} = \frac{\text{Cov} [y_1, y_2]}{\text{Var} [y_2]}$$



## Comments on the two normalizations

- The innovations in the representation

$$y_{1,t} = \varepsilon_{1,t}$$

$$y_{2,t} = \varepsilon_{2,t}$$

$$\text{Cov} [\varepsilon_1, \varepsilon_2] = \text{Cov} [y_1, y_1]$$

are different than the innovations in the representation

$$y_{1,t} = \eta_{1,t} + \psi_{12}\eta_{2,t}$$

$$y_{2,t} = \eta_{2,t}$$

$$\text{Cov} [\eta_1, \eta_2] = 0$$

- They are different specializations of the same general set of equations

$$y_{1,t} = \psi_{11}\varepsilon_{1,t} + \psi_{12}\varepsilon_{2,t}$$

$$y_{2,t} = \psi_{21}\varepsilon_{1,t} + \psi_{22}\varepsilon_{2,t}$$

## Comments on the two normalizations

- The two normalizations are representations of the same data because

$$\text{Var} [\varepsilon_1] = \text{Var} [y_1]$$

$$\text{Var} [\varepsilon_2] = \text{Var} [y_2]$$

$$\text{Cov} [\varepsilon_1, \varepsilon_2] = \text{Cov} [y_1, y_2]$$

- and

$$\text{Var} [\eta_1] = \text{Var} [y_1] - \psi_{12}^2 \text{Var} [y_2]$$

$$\text{Var} [\eta_2] = \text{Var} [y_2]$$

$$\psi_{12} = \frac{\text{Cov} [y_1, y_2]}{\text{Var} [y_2]}$$

with

$$\text{Cov} [\eta_1, \eta_2] = 0$$

## A third normalization

- Another normalization of

$$y_{1,t} = \psi_{11}\varepsilon_{1,t} + \psi_{12}\varepsilon_{2,t}$$

$$y_{2,t} = \psi_{21}\varepsilon_{1,t} + \psi_{22}\varepsilon_{2,t}$$

is  $\psi_{11} = \psi_{22} = 1$  and  $\psi_{12} = 0$  and  $\text{Cov}[\varepsilon_1, \varepsilon_2] = 0$  which implies

$$y_{1,t} = v_{1,t}$$

$$y_{2,t} = \psi_{21}v_{1,t} + v_{2,t}$$

$$\text{Cov}[v_1, v_2] = 0$$

- Do this yourself

## Three different normalizations fit the data equally well

- The general equation has unidentified parameters

$$y_{1,t} = \psi_{11}\varepsilon_{1,t} + \psi_{12}\varepsilon_{2,t}$$

$$y_{2,t} = \psi_{21}\varepsilon_{1,t} + \psi_{22}\varepsilon_{2,t}$$

with unrestricted  $\text{Var}[\varepsilon_1]$ ,  $\text{Var}[\varepsilon_2]$  and  $\text{Cov}[\varepsilon_1, \varepsilon_2]$

## Three different normalizations can be computed and fit the data equally well

- One normalization

$$y_{1,t} = \varepsilon_{1,t}$$

$$y_{2,t} = \varepsilon_{2,t}$$

$$\text{Cov} [\varepsilon_1, \varepsilon_2] = \text{Cov} [y_1, y_2]$$

- A second normalization

$$y_{1,t} = \eta_{1,t} + \psi_{12}\eta_{2,t}$$

$$y_{2,t} = \eta_{2,t}$$

$$\text{Cov} [\eta_1, \eta_2] = 0$$

- A third normalization

$$y_{1,t} = \nu_{1,t}$$

$$y_{2,t} = \psi_{21}\nu_{1,t} + \nu_{2,t}$$

$$\text{Cov} [\nu_1, \nu_2] = 0$$

## A different way of looking at this issue

- The data are serially uncorrelated and it might seem most natural just to use

$$y_{1,t} = \varepsilon_{1,t}$$

$$y_{2,t} = \varepsilon_{2,t}$$

$$\text{Cov} [\varepsilon_1, \varepsilon_2] = \text{Cov} [y_1, y_1]$$

- But what if  $y_{2,t}$  helps to determine  $y_{1,t}$  but not vice versa?

$$y_{1,t} = \eta_{1,t} + \psi_{12}y_{2,t}$$

$$y_{2,t} = \eta_{2,t}$$

- ▶ Estimation by ordinary least squares implies  $\text{Cov} [\eta_1, \eta_2] = 0$
- Important to note there is the third normalization with  $y_{2,t} = \psi_{21}y_{1,t} + \nu_{2,t}$  which fits the data equally well

## Three different normalizations can be computed and fit the data equally well

- Only economic theory – if anything – can tell us which is the most informative way of looking at the data
  - ▶ All of them are **observationally equivalent** in the sense that they fit the data equally well
  - ▶ Equally well: the values of the likelihood functions are identical
  - ▶ The data will not tell us which of these three – and probably other representations – is inconsistent with the data
  - ▶ All three representations are consistent with the data
- There is no sense of the word in which we can infer from the data that one variable determines another or is exogenous to the other
- This problem of reducing the set of parameters to ones which are interesting and estimable goes under the name of **identification**

# Normalization in vector autoregressions

- The same issue arises in vector autoregressions

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}_1(L) \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- ▶ Suppress the constant term to simplify the algebra
- ▶ Look at only two variables
- ▶ Only one lag

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- ▶ No obvious reason why correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  need be zero
- ▶ And the correlation won't be zero in general



## Ordering of vector autoregressions

- Instead of

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Estimate

$$y_{1,t} = a_{1,1}^*y_{2,t} + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \eta_{2,t}$$

- Current value of  $y_{2,t}$  appears in the equation for  $y_{1,t}$
- What is correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$ ?
- What is correlation of  $\eta_{1,t}$  and  $\eta_{2,t}$ ?

## Correlation of error terms in vector autoregression

- What is the correlation of error terms in

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

## Correlation of error terms in vector autoregression

- What is the correlation of error terms in

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Covariance of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is given by  $E[\varepsilon_{1,t}\varepsilon_{2,t}]$

## Correlation of error terms in vector autoregression

- What is the correlation of error terms in

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Covariance of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is given by  $E[\varepsilon_{1,t}\varepsilon_{2,t}]$
- Called “regressions”  $\Leftrightarrow$  computed by least squares

## Correlation of error terms in vector autoregression

- What is the correlation of error terms in

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Covariance of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is given by  $E[\varepsilon_{1,t}\varepsilon_{2,t}]$
- Called “regressions”  $\Leftrightarrow$  computed by least squares
- Know  $E[\varepsilon_{1,t}y_{1,t-1}] = E[\varepsilon_{1,t}y_{2,t-1}] = 0$  by least squares orthogonality

## Correlation of error terms in vector autoregression

- What is the correlation of error terms in

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Covariance of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is given by  $E[\varepsilon_{1,t}\varepsilon_{2,t}]$
- Called “regressions”  $\Leftrightarrow$  computed by least squares
- Know  $E[\varepsilon_{1,t}y_{1,t-1}] = E[\varepsilon_{1,t}y_{2,t-1}] = 0$  by least squares orthogonality
- Know  $E[\varepsilon_{2,t}y_{1,t-1}] = E[\varepsilon_{2,t}y_{2,t-1}] = 0$  by least squares orthogonality

## Correlation of error terms in vector autoregression

- What is the correlation of error terms in

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Covariance of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is given by  $E[\varepsilon_{1,t}\varepsilon_{2,t}]$
- Called “regressions”  $\Leftrightarrow$  computed by least squares
- Know  $E[\varepsilon_{1,t}y_{1,t-1}] = E[\varepsilon_{1,t}y_{2,t-1}] = 0$  by least squares orthogonality
- Know  $E[\varepsilon_{2,t}y_{1,t-1}] = E[\varepsilon_{2,t}y_{2,t-1}] = 0$  by least squares orthogonality
- Also know

$$\begin{aligned} E[\varepsilon_{1,t}\varepsilon_{2,t}] &= E[\varepsilon_{1,t}(y_{2,t} - a_{2,1}y_{1,t-1} - a_{2,2}y_{2,t-1})] \\ &= E[\varepsilon_{1,t}y_{2,t}] \\ &= E[(y_{1,t} - a_{1,1}y_{1,t-1} - a_{1,2}y_{2,t-1})y_{2,t}] \\ &= E[y_{2,t}y_{1,t}] - a_{1,1}E[y_{2,t}y_{1,t-1}] - a_{1,2}E[y_{2,t}y_{2,t-1}] \\ &= \text{Cov}[y_{1,t}y_{2,t}] - a_{1,1}a_{2,1}\text{Var}[y_{1,t-1}] - a_{1,2}a_{2,2}\text{Var}[y_{2,t-1}] \end{aligned}$$

## Correlation of error terms with one variable in the other equation

- The correlation of error terms is zero in

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}$$



## Correlation of error terms with one variable in the other equation

- The correlation of error terms is zero in

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}$$

- Computed by least squares

## Correlation of error terms with one variable in the other equation

- The correlation of error terms is zero in

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}$$

- Computed by least squares
- Covariance of  $\eta_{1,t}$  and  $\eta_{2,t}$  is given by  $E[\eta_{1,t}\eta_{2,t}]$

## Correlation of error terms with one variable in the other equation

- The correlation of error terms is zero in

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}$$

- Computed by least squares
- Covariance of  $\eta_{1,t}$  and  $\eta_{2,t}$  is given by  $E[\eta_{1,t}\eta_{2,t}]$
- Know  $E[\eta_{1,t}y_{2,t}] = 0$  by least squares orthogonality

## Correlation of error terms with one variable in the other equation

- The correlation of error terms is zero in

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}$$

- Computed by least squares
- Covariance of  $\eta_{1,t}$  and  $\eta_{2,t}$  is given by  $E[\eta_{1,t}\eta_{2,t}]$
- Know  $E[\eta_{1,t}y_{2,t}] = 0$  by least squares orthogonality
- Also know

$$\begin{aligned} E[\eta_{1,t}y_{2,t}] &= E[\eta_{1,t}(a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \eta_{2,t})] \\ &= a_{2,1} E[\eta_{1,t}y_{1,t-1}] + a_{2,2} E[\eta_{1,t}y_{2,t-1}] + E[\eta_{1,t}\eta_{2,t}] \\ &= E[\eta_{1,t}\eta_{2,t}] \end{aligned}$$

## Correlation of error terms with one variable in the other equation

- In

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}$$

- Know

$$E[\eta_{1,t} y_{2,t}] = 0$$

- Also know

$$E[\eta_{1,t} y_{2,t}] = E[\eta_{1,t} \eta_{2,t}]$$

- Therefore

$$E[\eta_{1,t} \eta_{2,t}] = 0$$

## Correlation of error terms with one variable in the other equation

- What if we estimated instead

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + v_{1,t}$$

$$y_{2,t} = \hat{a}_2^* y_{1,t} + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + v_{2,t}$$

- Then

$$E[v_{1,t}v_{2,t}] = 0$$

- Do yourself

## Vector autoregressive representations

- All three of the systems

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- and

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \eta_{2,t}$$

- and

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \nu_{1,t}$$

$$y_{2,t} = a_2^* y_{1,t} + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \nu_{2,t}$$

fit the data equally well

- The correlation of  $y_{1,t}$  and  $y_{2,t}$  over and above dependence on lagged values just appears in different places

## Vector autoregressive representations

- In

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

correlation of current unpredictable changes appears in correlation of error term

- In

$$y_{1,t} = a_1^* y_{2,t} + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \eta_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \eta_{2,t}$$

correlation of current unpredictable changes appears in coefficient  $a_1^*$

- In

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \nu_{1,t}$$

$$y_{2,t} = a_2^* y_{1,t} + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \nu_{2,t}$$

correlation of current unpredictable changes appears in coefficient  $a_2^*$



## Government expenditures and real GDP

- Does this matter?
- Government expenditures and real GDP from Andrew Young “Why in the world are we all Keynesians now?”, page 12
  - ▶ Graphs show response of real GDP to a one-standard deviation “shock” to government spending
  - ▶ Top graph has current government expenditures appearing in the real GDP equation

$$\begin{aligned}y_{1,t} &= a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t} \\y_{2,t} &= a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}\end{aligned}$$

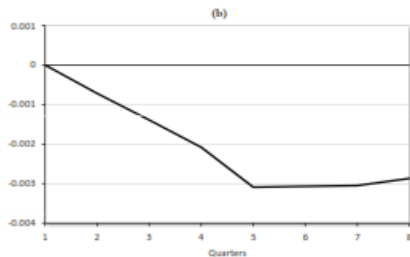
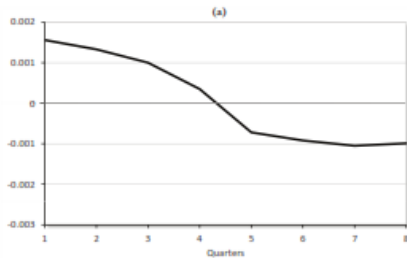
- ★ Effect of one-standard-deviation change of  $\eta_{2,t}$  on  $y_{1,t}, y_{1,t+1}, \dots$
- ▶ Bottom graph has current GDP appearing in the current government expenditures equation

$$\begin{aligned}y_{1,t} &= a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + v_{1,t} \\y_{2,t} &= a_2^* y_{1,t} + a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + v_{2,t}\end{aligned}$$

- ★ Effect of one-standard-deviation change of  $v_{2,t}$  on  $y_{1,t}, y_{1,t+1}, \dots$

# Government expenditures and real GDP

Figure 4  
Estimated Effects of Real GDP to a Federal Government Expenditures Shock



## Why not estimate simple VAR?

- Why not use

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

## Why not estimate simple VAR?

- Why not use

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Effect of one-standard-deviation change of  $\varepsilon_{2,t}$  on  $y_{1,t}$  doesn't represent change in government spending alone if the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is not zero

## Why not estimate simple VAR?

- Why not use

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Effect of one-standard-deviation change of  $\varepsilon_{2,t}$  on  $y_{1,t}$  doesn't represent change in government spending alone if the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is not zero
- There is a correlated change in  $y_{2,t}$

## Why not estimate simple VAR?

- Why not use

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Effect of one-standard-deviation change of  $\varepsilon_{2,t}$  on  $y_{1,t}$  doesn't represent change in government spending alone if the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is not zero
- There is a correlated change in  $y_{2,t}$
- If the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is zero *in this representation*, then can just use this representation

## Why not estimate simple VAR?

- Why not use

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Effect of one-standard-deviation change of  $\varepsilon_{2,t}$  on  $y_{1,t}$  doesn't represent change in government spending alone if the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is not zero
- There is a correlated change in  $y_{2,t}$
- If the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is zero *in this representation*, then can just use this representation
  - ▶ Why can one use this?

## Why not estimate simple VAR?

- Why not use

$$y_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \varepsilon_{2,t}$$

- Effect of one-standard-deviation change of  $\varepsilon_{2,t}$  on  $y_{1,t}$  doesn't represent change in government spending alone if the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is not zero
- There is a correlated change in  $y_{2,t}$
- If the correlation of  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is zero *in this representation*, then can just use this representation
  - ▶ Why can one use this?
  - ▶ What would be estimated coefficient if one estimated an equation in which the current value of  $y_{2,t}$  appeared in the  $y_{1,t}$  equation?



## Generalized impulse response functions

- “Generalized impulse response functions” attempt to get around the problem
- Rather than using one of the VARs above, use the impulse response when the variable has no current value of other variables appearing
- For current government spending  $y_{2,t}$  shock, use  $\eta_{2,t}$  in equation

$$\begin{aligned}y_{1,t} &= a_{1,1}^* y_{1,t-1} + a_{1,2} y_{2,t-1} + \eta_{1,t} \\y_{2,t} &= a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \eta_{2,t}\end{aligned}$$

- For real GDP shock  $y_{1,t}$ , use  $\nu_{1,t}$  in

$$\begin{aligned}y_{1,t} &= a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \nu_{1,t} \\y_{2,t} &= a_{2,1}^* y_{1,t-1} + a_{2,2} y_{2,t-1} + \nu_{2,t}\end{aligned}$$

- Problem: Not shocks to the same system, which makes them not even consistent with each other

## In short

- Identification in this context is choosing a normalization of the VAR
  - ▶ Cannot be determined based on the data
  - ▶ Makes it possible to estimate parameters in some specification
  - ▶ Can have important effects on empirical estimates

# Summary I

- A vector autoregression (VAR) is a very general representation of a time series
- Any set of covariance stationary time series has a VAR representation with serially uncorrelated innovations and potentially an infinite number of lags
- Normalizations are necessary to obtain a representation for a VAR
  - ▶ Problem is very obvious for moving average
  - ▶ The problem is there for VARs as well
- There are normalizations to identify a VAR
- Issue is how to identify the right (informative) relationship between contemporaneous values of the variables
- This relationship can appear in
  - ▶ Correlation of innovations
  - ▶ Coefficient of a variable in the other equation or equations
- Statistical criteria cannot determine the choice of which current values of variables appear in which equations

## Summary II

- More generally, Ordinary Least Squares may not even be the right way to summarize the relationship for some purposes