ECON 616: Lecture Four: VARs

Ed Herbst

Last Updated: March 13, 2025

Background

- *Overview:* Chapter 10 from cite:Hamilton.
- *Technical Details*: cite:Lutkepohl$_{1993}$
- *Other stuff*: There is a new book by cite:kilian$_{\text{lutkepohl2017}}$ – I haven't read it yet.
- *Some surveys*: cite:Stock2001, cite:Ramey$_{2016}$.

VARs

VARs have become an important tool for empirical macroeconomic research.

- Reduced Form representations of the data that summarize regular features and are suitable to conduct forecasts.
- Structural economic model can give some interpretation to a vector autoregression.

We'll talk about both today.

Some Theoretical Properties of VARs

A vector autoregression is a generalization of the AR(p) model to the multivariate case:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t \tag{1}$$

The random variable y_t is now a $n \times 1$ random vector that takes values in R^n.
For a theoretical analysis, it is often convenient to express the VAR(p) in the so-called companion form.

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \Phi_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Let $\xi_t = [y_t', y_{t-1}', \ldots, y_{t-p+1}']'$. The VAR can be rewritten as

$$\xi_t = F_0 + F_1 \xi_{t-1} + \nu_t \tag{2}$$

where the definitions of F_0, F_1, and ν_t can be deduced from the previous slide.
define the $n \times np$ matrix $M_n = [I, 0]$ where I is an $n \times n$ identity matrix.

It can be easily verified that $y_t = M_n \xi_t$.

The companion form is useful in two respects:

- to define stationarity in the context of a VAR
- to convince ourselves that without loss of much generality we can restrict econometric analyses to VAR(1) specifications.

Result For a vector autoregression to be covariance stationary it is necessary that all eigenvalues of the matrix F_1 are less than one in absolute value. $\square$

Example

Consider the univariate AR(2) process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

The AR(2) process can be written in companion form as a VAR(1) where $\xi_t = [y_t, y_{t-1}]'$ and

$$F_1 = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$$

The eigenvalues λ of the matrix F_1 satisfy the condition

$$det(F_1 - \lambda I) = 0 \iff (\phi_1 - \lambda)(-\lambda) - \phi_2 = 0$$

Provided that $\lambda \neq 0$ the equation can be rewritten as

$$0 = 1 - \phi_1 \frac{1}{\lambda} - \phi_2 \frac{1}{\lambda^2}$$

Thus, the condition $|\lambda| < 1$ is, at least in this example, equivalent to the condition that all the roots of the polynomial $\phi(z)$ are greater than one in absolute value. A generalization of this example can be found in Hamilton (1994, Chapter 1). $\square$

VAR(p)

Consider a VAR(p). The expected value of y_t has to satisfy the vector difference equation

$$\mathbb{E}[y_t] = \Phi_0 + \Phi_1 \mathbb{E}[y_{t-1}] + \ldots \Phi_p \mathbb{E}[y_{t-p}] \quad \text{for all } t \tag{3}$$

If the eigenvalues of F_1 are all less than one in absolute values and the VAR was initialized in the infinite past, then the expected value is given by

$$\mathbb{E}[y_t] = [I - \Phi_1 - \ldots \Phi_t]^{-1} \Phi_0 \tag{4}$$

To calculate the autocovariances we will assume that $\Phi_0 = 0$. Consider the companion form

$$\xi_t = F_1 \xi_{t-1} + \nu_t \tag{5}$$

If the eigenvalues of F_1 are all less than one in absolute value and the VAR was initialized in the infinite past, than the autocovariance matrix of order zero has to satisfy the equation

$$\Gamma_{\xi\xi,0} = \mathbb{E}[\xi_t \xi_t'] = F_1 \Gamma_{\xi\xi,0} F_1' + \mathbb{E}[\nu_t \nu_t'] \tag{6}$$

Obtaining a closed form solution for $\Gamma_{\xi\xi,0}$ is a bit more complicated than in the univariate AR(1) case.

Some Facts

Definition

Let A and B be 2×2 matrices with the elements

$$A = \left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right], \quad B = \left[\begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array} \right]$$

The *vec* operator is defined as the operator that stacks the columns of a matrix, that is,

$$vec(A) = [a_{11}, a_{21}, a_{12}, a_{22}]'$$

and the Kronecker product is defined as

$$A \otimes B = \left[\begin{array}{cc} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{array} \right] \quad \square$$

Lemma

Let A, B, C be matrices whose dimension are such that the product ABC exists. Then $vec(ABC) = (C' \otimes A)vec(B)$ $\quad \square$

VAR(p), continued

A closed form solution for the elements of the covariance matrix of ξ_t can be obtained as follows

$$\begin{aligned} vec(\Gamma_{\xi\xi,0}) &= (F_1 \otimes F_1) vec(\Gamma_{\xi\xi,0}) + vec(\mathbb{E}[\nu_t \nu_t']) \\ &= [I - (F_1 \otimes F_1)]^{-1} vec(\mathbb{E}[\nu_t \nu_t']) \end{aligned} \tag{7}$$

Since

$$\mathbb{E}[\xi_t \xi_{t-h}'] = F \mathbb{E}[\xi_{t-1} \xi_{t-h}'] + \mathbb{E}[\nu_t \xi_{t-h}'] \tag{8}$$

we can deduce that

$$\Gamma_{\xi\xi,h} = F_1^h \Gamma_{\xi\xi,0} \tag{9}$$

To obtain the autocovariance $\Gamma_{\xi\xi,-h}$ we have to keep track of a transpose in the general matrix case:

$$\Gamma_{\xi\xi,-h} = \mathbb{E}[\xi_{t-h} \xi_t'] = \left[\mathbb{E}[\xi_t \xi_{t-h}'] \right]' = \Gamma_{\xi\xi,h}' \tag{10}$$

VAR(p), continued

Once we have calculate that autocovariances for the companion form process ξ_t it is straightforward to obtain the autocovariances for the y_t process. Since $y_t = M_n \xi_t$ it follows that

$$\Gamma_{yy,h} = \mathbb{E}[y_t y'_{t-h}] = \mathbb{E}[M_n \xi_t \xi'_{t-h} M'_n] = M_n \Gamma_{\xi\xi,h} M'_n \tag{11}$$

Result: Consider the vector autoregression

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t$$

where $u_t \sim iid\mathcal{N}(0, \Sigma_u)$ with companion form

$$\xi_t = F_0 + F_1 \xi_{t-1} + \nu_t$$

Suppose that the eigenvalues of F_1 are all less than one in absolute values and that the vector autoregression was initialized in the infinite past. Under these assumptions the vector process y_t is covariance stationary with the moments

$$\begin{aligned} \mathbb{E}[y_t] &= [I - \Phi_1 - \ldots \Phi_t]^{-1} \Phi_0 && (12) \\ \Gamma_{yy,h} &= M_n \Gamma_{\xi\xi,h} M'_n \quad \forall h && (13) \end{aligned}$$

where

$$\begin{aligned} vec(\Gamma_{\xi\xi,0}) &= [I - (F_1 \otimes F_1)]^{-1} vec(\mathbb{E}[\nu_t \nu'_t]) && (14) \\ \Gamma &= F^h \Gamma \quad h > 0 \quad \square && (15) \end{aligned}$$

The Likelihood Function

We will now derive the likelihood function for a Gaussian VAR(p), conditional on initial observations $y_0, \dots, y_{-p+1}$. The density of y_t conditional on $y_{t-1}, y_{t-2}, \dots$ and the coefficient matrices $\Phi_0, \Phi_1, \dots, \Sigma$ is of the form

$$\begin{aligned} p(y_t | Y^{t-1}, \Phi_0, \dots, \Sigma) \quad \propto \quad & |\Sigma|^{-1/2} \exp \Big\{ -\frac{1}{2} (y_t - \Phi_0 - \Phi_1 y_{t-1} - \dots - \Phi_p y_{t-p})' \\ & \times \Sigma^{-1} (y_t - \Phi_0 - \Phi_1 y_{t-1} - \dots - \Phi_p y_{t-p}) \Big\} \end{aligned} \tag{16}$$

Define the $(np+1) \times 1$ vector x_t as

$$x_t = [1, y'_{t-1}, \dots, y'_{t-p}]'$$

Moreover, define the matrixes

$$Y = \begin{bmatrix} y'_1 \\ \vdots \\ y'_T \end{bmatrix}, \quad X = \begin{bmatrix} x'_1 \\ \vdots \\ x'_T \end{bmatrix}, \quad \Phi = [\Phi_0, \Phi_1, \dots, \Phi_p]'$$

The conditional density of y_t can be written in more compact notation as

$$p(y_t|Y^{t-1}, \Phi, \Sigma) \propto |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_t' - x_t'\Phi)\Sigma^{-1}(y_t' - x_t'\Phi)'\right\} \quad (17)$$

To manipulate the density we will use some matrix algebra facts.

Facts:

1. Let a be a $n \times 1$ vector, B be a symmetric positive definite $n \times n$ matrix, and tr the trace operator that sums the diagonal elements of a matrix. Then

$$a'Ba = tr[Baa']$$

2. Let A and B be two $n \times n$ matrices, then

$$tr[A + B] = tr[A] + tr[B]$$

In a first step, we will replace the inner product in the expression for the conditional density by the trace of the outer product

$$p(y_t|Y^{t-1}, \Phi, \Sigma) \propto |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(y_t' - x_t'\Phi)'(y_t' - x_t'\Phi)]\right\} \quad (18)$$

In the second step, we will take the product of the conditional densities of $y_1, \ldots, y_T$ to obtain the joint density. Let Y_0 be a vector with initial observations

$$\begin{aligned} p(Y|\Phi, \Sigma, Y_0) &= \prod_{t=1}^{T} p(y_t|Y^{t-1}, Y_0, \Phi, \Sigma) \\ &\propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} \sum_{t=1}^{T} tr[\Sigma^{-1}(y_t' - x_t'\Phi)'(y_t' - x_t'\Phi)]\right\} \\ &\propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1} \sum_{t=1}^{T} (y_t' - x_t'\Phi)'(y_t' - x_t'\Phi)\right]\right\} \\ &\propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(Y - X\Phi)'(Y - X\Phi)]\right\} \quad (19) \end{aligned}$$

Define the "OLS" estimator

$$\hat{\Phi} = (X'X)^{-1}X'Y \tag{20}$$

and the sum of squared OLS residual matrix

$$S = (Y - X\hat{\Phi})'(Y - X\hat{\Phi}) \tag{21}$$

It can be verified that

$$(Y - X\Phi)'(Y - X\Phi) = S + (\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi}) \tag{22}$$

This leads to the following representation of the likelihood function

$$\begin{aligned} p(Y|\Phi, \Sigma, Y_0) \quad &\propto \quad |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}S]\right\} \\ &\qquad \times \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}(\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi})]\right\} \end{aligned} \tag{23}$$

Alternative Representation

Let $\beta = vec(\Phi)$ and $\hat{\beta} = vec(\hat{\Phi})$. It can be verified that

$$tr[\Sigma^{-1}(\Phi - \hat{\Phi})' X' X (\Phi - \hat{\Phi})] = (\beta - \hat{\beta})' [\Sigma \otimes (X'X)^{-1}]^{-1} (\beta - \hat{\beta}) \quad (24)$$

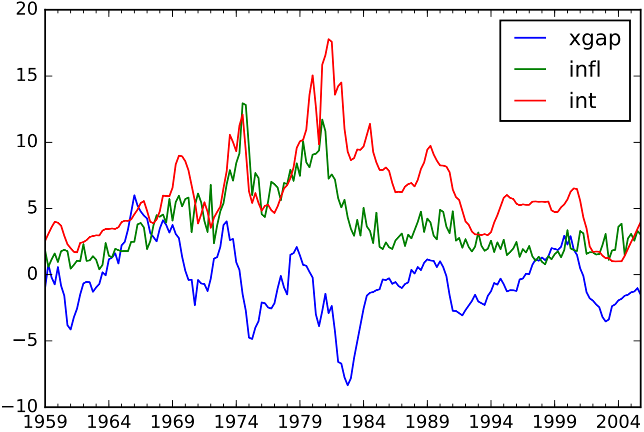
and the likelihood function has the alternative representation

$$\begin{aligned} p(Y|\Phi, \Sigma, Y_0) \quad \propto \quad & |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma^{-1} S] \right\} \\ & \times \exp\left\{ -\frac{1}{2} (\beta - \hat{\beta})' [\Sigma \otimes (X'X)^{-1}]^{-1} (\beta - \hat{\beta}) \right\} \end{aligned}$$

Inference

Above suggests we could estimate Φ and Σ via LS/MLE.

Consider a VAR(1) on the Output Gap, Inflation, and Interest Rate: [1959:Q1-2004:Q4]



MLE

$$\hat{\Phi}_0 = \begin{bmatrix} 0.44 \\ 0.24 \\ 0.24 \end{bmatrix}, \quad \hat{\Phi}_1 = \begin{bmatrix} 0.93 & -0.01 & -0.07 \\ 0.07 & 0.84 & 0.07 \\ 0.08 & 0.09 & 0.91 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} 0.62 & -0.04 & 0.24 \\ -0.04 & 1.27 & 0.16 \\ 0.24 & 0.16 & 0.87 \end{bmatrix}$$

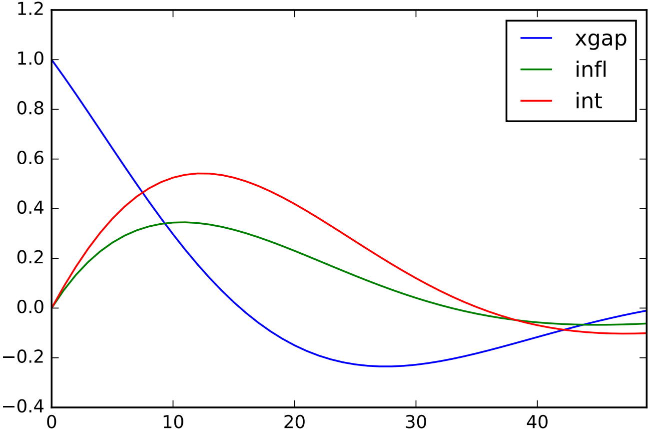
$|eig(\hat{\Phi}_1)| = [0.95, 0.95, 0.78] \implies$ stationary

Unconditional Mean

$$(I - \hat{\Phi}_1)^{-1}\hat{\Phi}_0 = [-0.54, 3.75, 6.09]$$

Impulse Response Function

Let's that $u_{1,t}$ equals 1 in some period t. What does that mean for $t+1, t+2, \ldots$?



Formally

IRF(h) $= \frac{dy t+h}{du_t}$

Can get via $MA(\infty)$ representation

$y_t = \Phi_0 + u_t + \Phi_1 u_{t-1} + \Phi^2 u_{t-2} + \dots$

Identifies consequences of a 1 unit increase in innovation $u_{1,t}$ on observables holding all innovations fixed.

Causality? Be careful. We're in reduced-form...

Still, get a sense of dynamics of system.

How to pick lags of VAR?

Could use *information criteria* Akaike, Bayesian, Schwarz, ... just like OLS.

OTOH, hand isn't it nice to use more lags = more complex dynamics, MP "long and variable"

	$p = 1$	$p = 6$
$std(\hat{u}_1)$	0.62	0.45
$std(\hat{u}_2)$	1.27	1.01
$std(\hat{u}_3)$	0.87	0.61

How to use more lags without overfitting?

Granger Causality

Economists often use regression results to make statements about causal relationships between variables.

- Suppose we would like to examine the monetarist hypothesis that a contraction of the money supply causes a decrease in aggregate output.
- It is tempting to regress output on a measure of lagged money supply and interpret a non-zero coefficient as "causal" relationship.
- Since this concept of causality is somewhat different from the usual notion of causality it gets a new name.

Bivariate Granger Causality The random variable $y_{2,t}$ fails to Granger cause the random variable $y_{1,t}$ if for all $s > 0$ the mean squared error of a forecast of $y_{1,t+s}$ based on $y_{1,t}, y_{1,t-1}, \ldots$ is the same as the mean squared error of a forecast that uses both $y_{1,t}, y_{1,t-1}, \ldots$ and $y_{2,t}, y_{2,t-1}, \ldots$. $\square$

Example

Consider the bivariate VAR(2)

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_1^{(0)} \\ \phi_2^{(0)} \end{bmatrix} + \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \ldots + \begin{bmatrix} \phi_{11}^{(p)} & \phi_{12}^{(p)} \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} \end{bmatrix} \Bigg[$$

If $y_{2,t}$ fails to Granger cause $y_{1,t}$, then it must be true that

$$\phi_{12}^{(1)} = \phi_{12}^{(2)} = \ldots = \phi_{12}^{(p)} = 0 \tag{26}$$

A discussion of Granger causality in the context of a VAR with more than two variables can be found in Hamilton (1994). We will now examine Granger causality in the context of forward looking behavior. Roughly speaking:

> *The weather forecast Granger causes the weather, but shooting the weatherman will not produce a sunny weekend. (Cochrane, 1994).*

Example

Consider a investor who has the choice between a riskless bond that yields a return r, and a risky asset that has a price p_t and will pay dividends d_{t+1} in the next period. In equilibrium under the absence of arbitrage

$$1 + r = \mathbb{E}_t \left[\frac{p_{t+1} + d_{t+1}}{p_t} \right] \tag{27}$$

The forward solution of this difference equation implies that the price of the risky asset is

$$p_t = \mathbb{E}_t \left[\sum_{\tau=1}^{\infty} \left(\frac{1}{1+r} \right)^{\tau} d_{t+\tau} \right] \tag{28}$$

Thus, according to the model, the stock price incorporates the market's best forecast of the present value of future dividends. If this forecast is based on more information than past dividends alone, then stock prices will Granger cause dividends, as investors try to anticipate movements in dividends.

Example

Suppose that

$$d_t = d + u_t + \delta u_{t-1} + \nu_t \tag{29}$$

where u_t and ν_t are independent Gaussian *iid* series. Suppose that the investor at time t knows values of current and past u_t and ν_t's. The forecast of $d_{t+\tau}$ based on this information is given by

$$\mathbb{E}_t[d_{t+\tau}] = \begin{cases} d + \delta u_t & \text{for } \tau = 1 \\ d & \text{for } \tau = 2, 3, \ldots \end{cases} \tag{30}$$

Thus, the stock price is given by

$$p_t = \frac{d}{r} + \frac{\delta u_t}{1 + r} \tag{31}$$

which implies that

$$\delta u_{t-1} = (1 + r)p_{t-1} - (1 + r)d/r \tag{32}$$

The system can be written as a bivariate VAR

$$\begin{bmatrix} p_t \\ d_t \end{bmatrix} = \begin{bmatrix} d/r \\ -d/r \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 + r & 0 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ d_{t-1} \end{bmatrix} + \begin{bmatrix} \delta u_t/(1 + r) \\ u_t + \nu_t \end{bmatrix} \tag{33}$$

Upshot

- In this model, Granger causation runs the opposite direction from the true causation.
- Dividends fail to “Granger-cause” prices, even though investors’ perceptions of dividends are the sole determinant of stock prices.
- On the other hand, prices do “Granger-cause” dividends, even though the market’s evaluation of the stock in reality has no effect on the dividend process. (Hamilton, 1994, Chapter 11).

How to think about causation?

A last word about cointegration

We will now analyze a simple bivariate system of cointegrated processes. Consider the model

$$y_{1,t} = \gamma y_{2,t} + u_{1,t} \tag{34}$$
$$y_{2,t} = y_{2,t-1} + u_{2,t} \tag{35}$$

where $[u_{1,t}, u_{2,t}]' \sim iid(0, \Omega)$.
Clearly, $y_{2,t}$ is a random walk. Moreover, it can be easily verified that $y_{1,t}$ follows a unit root process.

$$y_{1,t} - y_{1,t-1} = \gamma(y_{2,t} - y_{2,t-1}) + u_{1,t} - u_{1,t-1} \tag{36}$$

Therefore,

$$y_{1,t} = y_{1,t-1} + \gamma u_{2,t} + u_{1,t} - u_{1,t-1} \tag{37}$$

Thus, both $y_{1,t}$ and $y_{2,t}$ are integrated processes.

Model Continued

However, the linear combination

$$[1,\ -\gamma]\left[\begin{array}{c} y_{1,t} \\ y_{2,t} \end{array}\right] = y_{1,t} - \gamma y_{2,t} = u_{1,t} \tag{38}$$

is stationary. Therefore, $y_{1,t}$ and $y_{2,t}$ are cointegrated.

The vector $[1, -\gamma]'$ is called the cointegrating vector.

Note that the cointegrating vector is only unique up to normalization.

Rewriting the Model

The model can be rewritten as a VAR(1)

$$y_t = \Phi_1 y_{t-1} + \epsilon_t \tag{39}$$

The elements of the matrix Φ_1 and the definition of ϵ_t is given by

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} + \gamma u_{2,t} \\ u_{2,t} \end{bmatrix} \tag{40}$$

The matrix Φ_1 is of reduced rank in this example of cointegration. More generally cointegrated system can be casted in the form of a vector autoregression in levels of y_t.

Although both $y_{1,t}$ and $y_{2,t}$ follow univariate random walks, the cointegrated system cannot be expressed as a vector autoregression in differences $[\Delta y_{1,t}, \Delta y_{2,t}]'$. Consider

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 - L & \gamma L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} = \Theta(L) u_t \tag{41}$$

Since $|\Theta(1)| = 0$ the moving average polynomial is not invertible and no finite order VAR could describe Δy_t.

VECM

The cointegrated model can be written in the so-called vector error correction model (VECM) form:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \left(\begin{bmatrix} 1 & -\gamma \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \right) + \begin{bmatrix} u_{1,t} + \gamma u_{2,t} \\ u_{2,t} \end{bmatrix} \tag{42}$$

The term

$$\left(\begin{bmatrix} 1 & -\gamma \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \right) = y_{1,t-1} - \gamma y_{2,t-1} \tag{43}$$

is called error correction term. In economic models it often reflects a long-run equilibrium relationship such as a constant ratio of consumption and output. If the economy is out of equilibrium in period $t-1$, that is, $y_{1,t-1} - \gamma y_{2,t-1} \neq 0$, then the economy adjusts toward its long-run equilibrium and ${}_{t-1}[\Delta y_t] \neq 0$. If the "true" cointegrating vector is known, then both the left-hand-side variables and the error correction term are stationary.

Upshot

In practice, if one would like to model a bivariate vector process y_t, it has to be determined whether to fit

1. An unrestricted vector autoregression of the form
2. a vector autoregression in differences
3. or a vector error correction model (reduced rank regression)

How to pick:

- A likelihood ratio test or a Bayesian model selection criterion could be used
- if the processes $y_{1,t}$ and $y_{2,t}$ are integrated the analysis of the sampling distribution of the likelihood ratio test statistics is complicated
- Johansen (1995) provides a nice summary of the relevant asymptotic distribution theory.

SVARs

So far, we considered reduced form VARs, say,

$$y_t = \Phi_1 y_{t-1} + u_t, \quad \mathbb{E}[u_t u_t'] = \Sigma_u \tag{44}$$

in which the error terms u_t have the interpretation of one-step ahead forecast errors. If the eigenvalues of Φ_1 are inside the unit-circle then y_t has the following moving-average (MA) representation in terms of u_t:

$$y_t = (I - \Phi_1 L)^{-1} u_t = \sum_{j=0}^{\infty} \Phi_1^j u_{t-j} = \sum_{j=0}^{\infty} C_j u_{t-j} \tag{45}$$

Modern dynamic macro models suggest that the one-step ahead forecast errors are functions of some fundamental shocks, such as technology shocks, preference shocks, or monetary policy shocks.

Let ϵ_t a vector of such fundamental shocks and assume that $\mathbb{E}[\epsilon_t \epsilon_t'] = \mathcal{I}$. Moreover, assume that

$$u_t = \Phi_\epsilon \epsilon_t. \tag{46}$$

Then we can express the VAR in structural form as follows

$$\begin{aligned} y_t &= \Phi_1 y_{t-1} + \Phi_\epsilon \epsilon_t \\ \Phi_\epsilon^{-1} y_t &= \Phi_\epsilon^{-1} \Phi_1 y_{t-1} + \epsilon_t \end{aligned} \tag{47}$$

The moving-average representation of y_t in terms of the structural shocks is given by

$$y_t = \sum_{j=0}^{\infty} \Phi_1^j \Phi_\epsilon \epsilon_{t-j} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \epsilon_{t-j}. \tag{48}$$

The moving-average representation of y_t in terms of the structural shocks is given by

$$y_t = \sum_{j=0}^{\infty} \Phi_1^j \Phi_\epsilon \epsilon_{t-j} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \epsilon_{t-j}. \tag{49}$$

For~(44) and~(47) the matrix Φ_ϵ has to satisfy the restriction

$$\Phi_\epsilon \Phi_\epsilon' = \Sigma_u. \tag{50}$$

Notice that the matrix Φ_ϵ has n^2 elements. The covariance relationship, unfortunately, generates only $n(n+1)/2$ restrictions and does not uniquely determine Φ_ϵ. This creates an identification problem since all we can estimate from the data is Φ_1 and Σ_u.

In order to make statements about the propagation of structural shocks ϵ_t we have to make further assumptions. The papers by Cochrane (1994), Christiano and Eichenbaum (1999), and Stock and Watson (2001) survey such identifying assumptions. A cynical view of this literature is the following:

1. Propose an identification scheme, that determines all elements of Φ_ϵ.
2. Compute impulse response functions.
3. If impulse response functions are plausible, then stop; else, declare a "puzzle" and return to 1.

Here are some famous "puzzles:"

1. "Liquidity Puzzle:" When identifying monetary policy shocks as surprise changes in the stock of money one often finds that interest rates fall when the money stock is lowered.
2. "Price Puzzle:" When identifying monetary policy shocks as surprise changes in the Federal Funds Rate, one often finds that prices fall after a drop in interest rates.

These "puzzles" are typically resolved by considering more elaborate identification schemes.

Impulse Response Functions and Variance Decompositions

Impulse responses are defined as

$$\frac{\partial y_{t+h}}{\partial \epsilon_t'} = C_h \Phi_\epsilon \tag{51}$$

and correspond to the MA coefficient matrices in the moving average representation of y_t in terms of structural shocks.
The covariance matrix of y_t is given by

$$\Gamma_{yy,0} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \mathcal{I} \Phi_\epsilon' C_j' \tag{52}$$

Let $\mathcal{I}^i$ be matrix for which element i, i is equal to one and all other elements are equal to zero. Then we can define the contribution of the i'th structural shock to the variance of y_t as

$$\Gamma_{yy,0}^{(i)} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \mathcal{I}^{(i)} \Phi_\epsilon' C_j' \tag{53}$$

Thus the fraction of the variance of $y_{l,t}$ explained by shock i is

$$[\Gamma_{yy,0}^{(i)}]_{ll} / [\Gamma_{yy,0}]_{ll}.$$

We begin by decomposing the covariance matrix into the product of lower triangular matrices (Cholesky Decomposition):

$$\Sigma_u = AA', \tag{54}$$

where A is lower triangular. If Σ_u is non-singular the decomposition is unique. Let Ω be an orthonormal matrix, meaning that $\Omega\Omega' = \Omega'\Omega = \mathcal{I}$. We can characterize the relationship between the reduced form and the structural shocks as follows

$$u_t = A\Omega\epsilon_t \tag{55}$$

Notice that

$$\mathbb{E}[u_t u_t'] = \mathbb{E}[A\Omega\epsilon_t\epsilon_t'\Omega' A'] = A\Omega\mathbb{E}[\epsilon_t\epsilon_t']\Omega' A' = A\Omega\Omega' A' = AA' = \Sigma_u. \tag{56}$$

In general, it is quite tedious to characterize the space of orthonormal matrices. Let's try for $n = 2$:

$$\Omega(\varphi) = \left[\begin{array}{cc} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{array} \right] \tag{57}$$

where $\varphi \in (-\pi, \pi]$. Notice that, for instance,

$$\Omega(\pi/2) = -\Omega(-\pi/2) \tag{58}$$

which means that only the signs of the impulse responses change but not the shape.

Let's look at some famous identification schemes

Sims (1980)

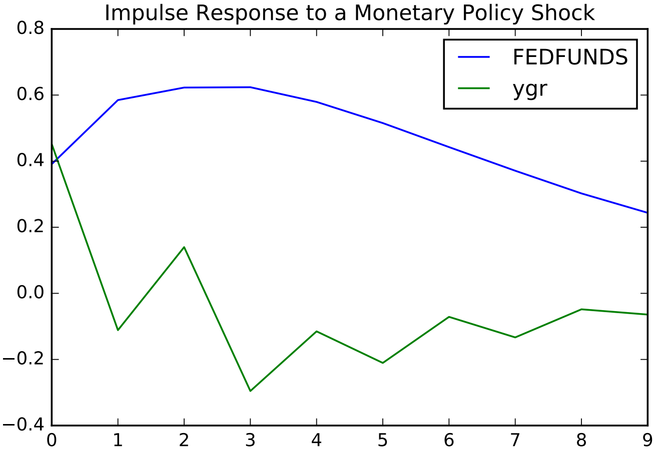
Suppose that

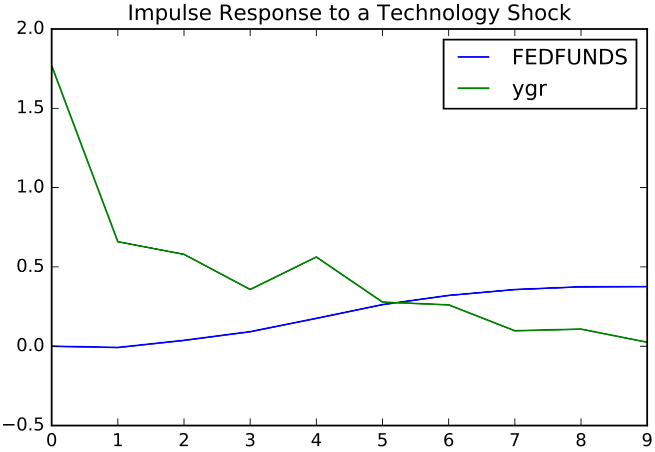
$$y_t = \begin{bmatrix} \text{Fed Funds Rate} \\ \text{Output Growth} \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} = \begin{bmatrix} \text{Monetary Policy Shock} \\ \text{Technology Shock} \end{bmatrix}.$$

Moreover, we assume that the central bank does not react contemporaneously to technology shocks because data on aggregate output only become available with a one-quarter lag. This assumption can be formalized through $\varphi = 0$. Then

$$u_t = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix}. \tag{59}$$

Further readings: cite:Sims1980.


Impulse Response to a Monetary Policy Shock
FEDFUNDS
ygr
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
1
2
3
4
5
6
7
8
9


Impulse Response to a Technology Shock
FEDFUNDS
ygr
2.0
1.5
1.0
0.5
0.0
−0.5
0
1
2
3
4
5
6
7
8
9

Blanchard and Quah (1989)

Now suppose that

$$y_t = \begin{bmatrix} \text{Inflation} \\ \text{Output Growth} \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} = \begin{bmatrix} \text{Monetary Policy Shock} \\ \text{Technology Shock} \end{bmatrix}$$

Moreover,

$$y_t = (\sum_{j=0}^{\infty} C_j L^j) u_t = C(L) u_t. \tag{60}$$

Consider the following assumption: monetary policy shocks do not raise output in the long-run. Let's examine the moving average representation of y_t in terms of the structural shocks

$$\begin{aligned} y_t &= \begin{bmatrix} c_{11}(L) & c_{12}(L) \\ c_{21}(L) & c_{22}(L) \end{bmatrix} \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} \\ &= \begin{bmatrix} \cdot & \cdot \\ a_{11}\cos\varphi c_{21}(L) + (a_{21}\cos\varphi + a_{22}\sin\varphi) c_{22}(L) & \cdot \end{bmatrix} \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} \\ &= \begin{bmatrix} d_{11}(L) & d_{12}(L) \\ d_{21}(L) & d_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} \end{aligned}$$

Suppose that in period $t = 0$ log output and log prices are equal to zero. Then the log-level of output and prices in period $t = T > 0$ is given by

$$y_T^c = \sum_{t=1}^{T} y_t = \sum_{t=1}^{T} \sum_{j=0}^{\infty} D_j \epsilon_{t-j} \tag{61}$$

Now consider the derivative

$$\frac{\partial y_T^c}{\partial \epsilon_1'} = \sum_{j=0}^{T-1} D_j \tag{62}$$

Letting $T \longrightarrow \infty$ gives us the long-run response of the level of prices and output to the shock ϵ_1:

$$\frac{\partial y_\infty^c}{\partial \epsilon_1'} = \sum_{j=0}^{\infty} D_j = D(1) \tag{63}$$

Here, we want to restrict the long-run effect of monetary policy shocks on output:

$$d_{21}(1) = 0 \tag{64}$$

This leads us to the equation

$$[a_{11}c_{21}(1) + a_{21}c_{22}(1)] \cos\varphi + a_{22}c_{22}(1) \sin\varphi = 0. \tag{65}$$

Notice that the equation has two solutions for $\varphi \in (-\pi, \pi]$. Under one solution a positive monetary policy shock is contractionary, under the other solution it is expansionary. The shape of the responses is, of course, the same.

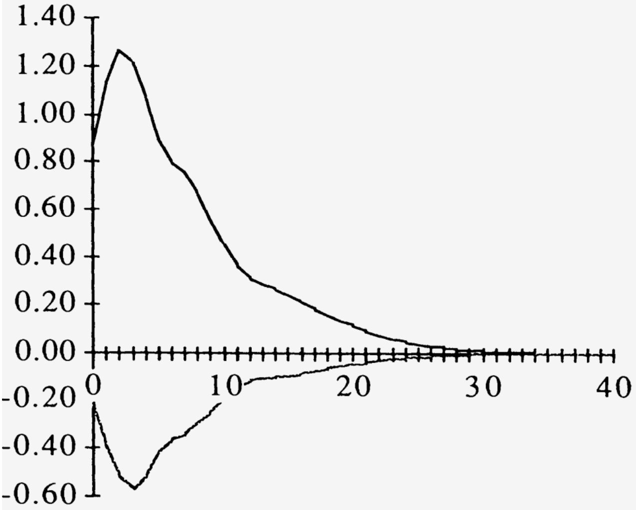


FIGURE 1. RESPONSE TO DEMAND, — = OUTPUT,
~~~ = UNEMPLOYMENT
~~~

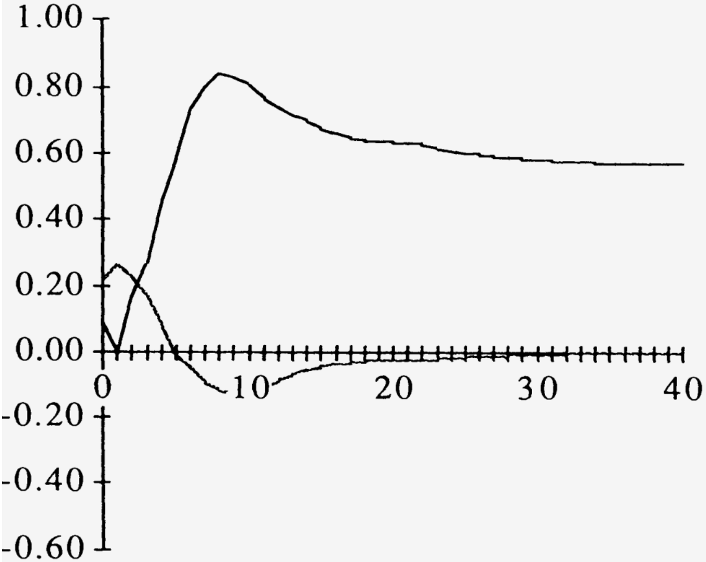


FIGURE 2. RESPONSE TO SUPPLY, — = OUTPUT, ~~~ = UNEMPLOYMENT

Sign Restrictions

Again consider

$$y_t = \begin{bmatrix} \text{Inflation} \\ \text{Output Growth} \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} = \begin{bmatrix} \text{Monetary Policy Shock} \\ \text{Technology Shock} \end{bmatrix}$$

and our identification assumption is: upon impact, a monetary policy shock raises both prices and output. It can be verified that

$$\frac{\partial y_t}{\partial \epsilon_{R,t}} = \begin{bmatrix} a_{11} \cos\varphi c_{11,1} + (a_{21}\cos\varphi + a_{22}\sin\varphi)c_{12,1} \\ a_{11} \cos\varphi c_{21,1} + (a_{21}\cos\varphi + a_{22}\sin\varphi)c_{22,1} \end{bmatrix}. \tag{66}$$

Thus, we obtain the sign restrictions

$$\begin{aligned} 0 &< a_{11} \cos\varphi c_{11,1} + (a_{21}\cos\varphi + a_{22}\sin\varphi)c_{12,1} \\ 0 &< a_{11} \cos\varphi c_{21,1} + (a_{21}\cos\varphi + a_{22}\sin\varphi)c_{22,1} \end{aligned}$$

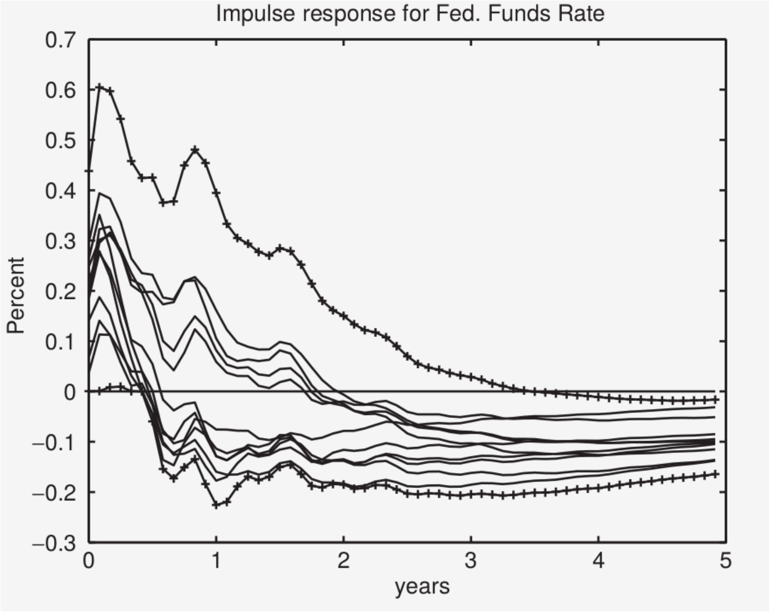
which restrict φ to be in a certain subset of $(-\pi, \pi]$ and will generate a range of responses.
Further readings: cite:Canova2002, cite:Faust1998, cite:Uhlig2005.

Uhlig, 2005

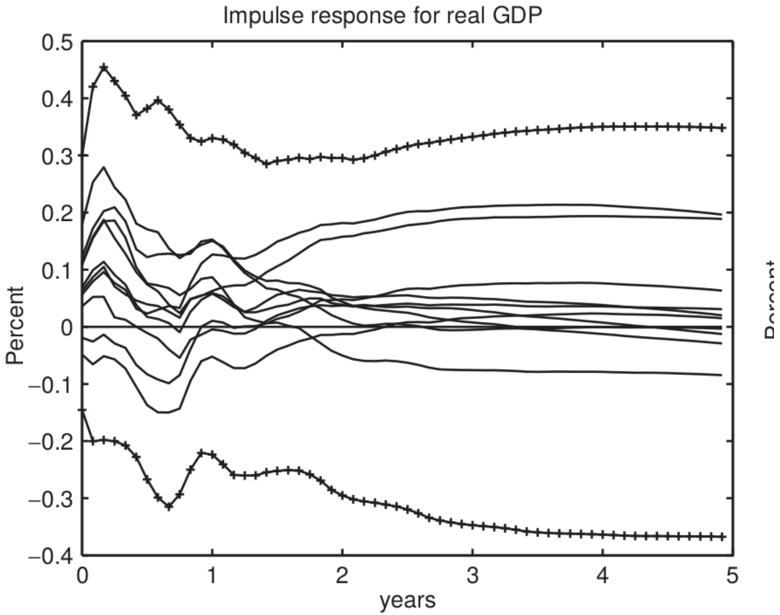
- What is the effect of MP on Output?
- Let's assume that after an MP shock R_{t+k} for $k = 1, \ldots K$.
- How does it compare to the standard ordering?

Result: Monetary Policy does not effect output!

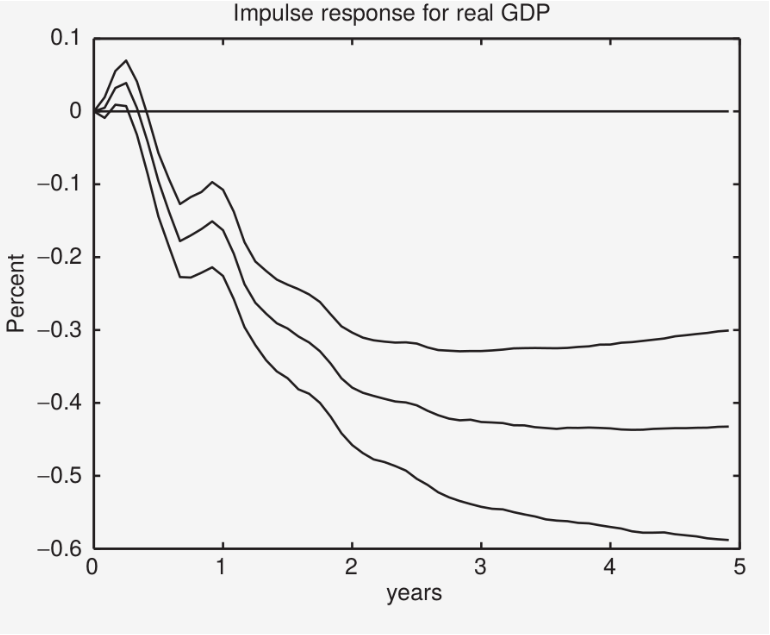
Sign Restrictions



Sign Restrictions



Cholesky



Impulse response for real GDP
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.6
Percent
0
1
2
3
4
5
years

Is that the last word?

NO

- Can you verify sign restrictions?
- It's hard to get (all) the right Omegas [cite:Arias$_{2014}$]
- Other "reasonable" sign restrictions give different results?
- Are we back to "puzzle"?

References I