Time Series Analysis:

Introduction to time series and forecasting

Handout 5: Multivariate processes I: Stationary VAR

Laura Mayoral

IAE and Barcelona GSE

Multivariate Time Series Analysis

- Multivariate time series models are useful for a number of analyses and applications:

  - Forecasting
  - Studying the dynamic interrelationships between a number of variables.
  - Studying the effects of economic shocks of interest.
    - What are the effects of monetary policy shocks?
    - What kind of shocks drive the business cycle?

- The empirical evidence obtained using these models can provide useful information to policymakers and practitioners.

- The theory developed for univariate case extends in a natural way to the multivariate case.

- For instance, consider the following model

$$\Phi_p\left(L\right)y_t = c + \Theta\left(L\right)\varepsilon_t \tag{1}$$

- If y_t is a scalar, then y_t is a univariate ARMA(p, q) process

- If y_t is a vector than contains n variables, i.e., $y_t = (y_{1t},\ \ldots,\ y_{nt})'$, then y_t is a Vector ARMA or VARMA process.

- This lecture reviews models for stationary vector processes.
- VAR (=vector autoregressive) models are very popular.

Roadmap

1. Preliminary Concepts

2. Covariance-stationarity. Wold Decomposition for vector processes. Invertibility

3. Structural MA representation.

4. VAR models. The structural VAR representation.

5. Reduced-form VAR models.

5.1. Stationarity

5.2 From the VAR to the MA representation

5.3 VAR(p) second moments

6. Forecasting

7. Granger Causality

8. IRFs

9. Variance decomposition

1. Preliminary Concepts

Random Vector: A random vector is a vector $X = (X_1, \ldots, X_n)'$ whose components are scalar-valued random variables on the same probability space.

Vector Random Process : A family of random vectors $\{X_t; t \in T\}$ indexed by t, with $t \in T$, where T is a set of time points. Typically T is the set of natural or integers numbers.

Matrix of polynomials in the lag operator: $\Phi(L)$ is a polynomial in the lag operator where Φ_i are matrices of coefficients.

Example:

$$\Phi(L) = \begin{bmatrix} 1 & -0.5L \\ L & 1+L \end{bmatrix} = \Phi_0 + \Phi_1 L$$

With

$$\Phi_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Phi_1 = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix}$$

When applied to the vector X_t

$$\Phi(L)X_t = \begin{bmatrix} 1 & -0.5L \\ L & 1+L \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} X_{1t} - 0.5X_{2t-1} \\ X_{1t-1} + X_{2t} + X_{2t-1} \end{bmatrix}$$

2. Covariance Stationarity

Definition.

It is similar to the univariate case. The vector y_t is covariance stationary if

i) $E(y_t) = \mu$ for all t

ii) $E\left((y_t - \mu)(y_{t-h} - \mu)'\right) = \underset{n \times n}{\Gamma(h)}$ for all t and $h = 0, 1, 2...$

Two remarks

- Stationarity of all the components of y_t does not imply stationarity of the vector process. Stationarity of the y_t requires that the components of y_t are stationary and costationary.

- Although $\Gamma(j) = \Gamma(-j)$ for a scalar process, the same is not true for vector processes. The correct statement is

$$\Gamma_j' = \Gamma_{-j}$$

Example: $n = 2$ and $\mu = 0$

$$\begin{aligned}
\Gamma_1 &= \begin{pmatrix} E(Y_{1t}Y_{1t-1}) & E(Y_{1t}Y_{2t-1}) \\ E(Y_{2t}Y_{1t-1}) & E(Y_{2t}Y_{2t-1}) \end{pmatrix} \\
&= \begin{pmatrix} E(Y_{1t+1}Y_{1t}) & E(Y_{1t+1}Y_{2t}) \\ E(Y_{2t+1}Y_{1t}) & E(Y_{2t+1}Y_{2t}) \end{pmatrix} \\
&= \begin{pmatrix} E(Y_{1t}Y_{1t+1}) & E(Y_{1t}Y_{2t+1}) \\ E(Y_{2t}Y_{1t+1}) & E(Y_{2t}Y_{2t+1}) \end{pmatrix}' = \Gamma'_{-1}
\end{aligned}$$

Simplest example.

Vector White noise process.

- ε_t (with dimension $n \times 1$) is a white noise process if

$$E(\varepsilon_t) = 0_{n\times 1}$$

$$E(\varepsilon_t \varepsilon'_{t-s}) = 0_{n\times n}$$

$$E(\varepsilon_t \varepsilon'_t) = \Omega_{n\times n}$$

Wold decomposition

Any zero-mean stationary vector process Y_t admits the following representation

$$Y_t = C(L)\varepsilon_t + \kappa_t$$

where $C(L)$ is the stochastic component with $C(L) = \sum_{i=0}^{\infty} C_i L^i$ and κ_t is a purely deterministic component, that is perfectly predictable using linear combinations of past values of Y_t.

If $\kappa_t = 0$, the process is regular. In the following we will consider only regular processes.

- The above-defined Wold representation verifies,

- ε_t is a white noise process

- $\varepsilon_t = Y_t - Proj(Y_t|Y_{t-1}, Y_{t-2}, \ldots)$

- C_i is a sequence of fixed n $\times$ n matrices that are square summable: $\sum_{i=0}^{\infty} ||C_i||^2 < \infty$.

- $C_0 = I_n$

- The Wold representation is unique.

Invertibility

Invertibility. A MA(q) process defined by the equation $Y_t = C(L)\epsilon_t$ is said to be invertible if there exists a sequence of absolutely summable matrices of constants $\{A_j\}_{j=0}^{\infty}$ such that $\sum_{j=0}^{\infty} A_j Y_{t-j} = \epsilon_t$.

Proposition. A MA(q) process defined by the equation $Y_t = C(L)\epsilon_t$ is invertible if and only if the determinant of $C(L)$ vanishes only outside the unit circle, i.e., if $det(C(z)) \neq 0$ for all $|z| \leq 1$.

If the process is invertible, it possesses a unique VAR representation (clear later on)

Example

Consider the process

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 1 & L \\ 0 & \theta - L \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

$\det(C(z)) = \theta - z$ which is zero for $z) = \theta$.

Thus, the process is invertible if and only if $|\theta| > 1$.

3. Structural and reduced-form models

Structural MA models

- Consider the multivariate model

$$y_t = C(L)\, u_t, \tag{2}$$

where y_t is a $\mathrm{n}_y \times 1$ vector of economic variables and u_t is an $\mathrm{n}_u \times 1$ vector of shocks. (Remark: n_y can be different from n_u).

- Equation (2) is called the **structural MA model** since the elements of u_t are given a structural economic interpretation.

- For example, one of the components of u_t can be interpreted as an exogenous shock to labor supply, another as a shock to the quantity of money, etc.

- These shocks are assumed to be unobservable and white noise

$$E(u_t u_s') = \begin{cases} \Sigma, \text{ if } t = s \\ 0, \text{ otherwise} \end{cases} .$$

(or sometimes, $i.i.d$).

- It is typically assumed that Σ is diagonal (that is, no contemporaneous correlations among the shocks exist).

- Notice that C_0 is not assumed to be the identity matrix.

4. VAR models

- Vector Autoregressive Models come in three varieties: structural form, reduced form and recursive.

- A reduced form VAR expresses each variable as a linear function of its own past values, the past values of all other variables being considered and a serially uncorrelated error term.

- A recursive VAR constructs the error terms in each regression equation to be uncorrelated with the error in the preceding equation. This is done by judiciously including some contemporaneous terms values as regressors (see below for an example).

- A structural VAR uses economic theory to sort out the contemporaneous links among the variables. To be able to do that, structural VARs require "identifying assumptions" that allow correlations to be interpreted causally.

Structural VAR models

From structural MA to Structural VAR models

- If $C(L)^{-1}$ exits, then premultiplying (2) by $C(L)^{-1}$ one obtains the **Structural VAR model**

$$B(L)\,y_t = u_t,$$

where $B(L) = B_0 - \sum_{k=1}^{\infty} B_i L^i$. Notice that in this representation the elements of u_t still are **the structural shocks** of the system.

- If the previous sum is finite, then we have a structural VAR(p) process.

$$B_0 y_t = B_1 y_{t-1} + ... + B_p y_{t-p} + u_t, \tag{3}$$

where u_t is assumed to be white noise (or sometimes $i.i.d$).

Remark: $n_y \geq n_u$ is a necessary condition for invertibility. If $n_y = n_u$, then C(L) is square and the invertibility condition is that the determinant of the polynomial $C(z)$, $|C(z)|$, has all its roots outside the unit circle.

Reduced-form VAR models

- The reduced-form VAR model is obtained by premultiplying (3) by B_0^{-1} :

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t,$$

where $\phi_i = B_0^{-1} B_i$ and $\varepsilon_t = B_0^{-1} u_t$.

- If u_t is white noise, so is ε_t. However, ε_t does not represent the structural shocks (the 'true' economic shocks) anymore, but a linear combination of them.

- Thus, although VAR models are useful devices in many instances, they do not have a structural interpretation (their connection with the theory is not inmediate).

- The variance covariance matrix of ε_t, $\Sigma_\varepsilon = B_0^{-1} \Sigma_u B_0^{-1\prime}$ will not be in general diagonal (even if Σ_u is), therefore, these shocks are contemporaneously correlated.

An example

- Consider a structrural VAR(1) model for $y_t = (y_{1t},\ y_{2t})'$,

$$B_0 y_t = \gamma_0 + \gamma_1 y_{t-1} + u_t,$$

that is

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_{01} \\ \gamma_{02} \end{pmatrix} + \tag{4}$$

$$+ \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-2} \end{pmatrix} + \begin{matrix} u_{1t} \\ u_{2t} \end{matrix} \tag{5}$$

where y_{1t} denotes the rate of growth of output and y_{2t} the rate of growth of money.

- Notice that both y_{1t} and y_{2t} are endogeneous regressors in (4).
- There are 13 parameters to estimate

Structural versus Reduced form VAR models, cont.

- In the (reduced-form) VAR representation, y_1 and y_2 are just functions of their past values.

- This solves the above mentioned endogeneity problem.

- To obtain this representation, premultiply by B_0^{-1} :

$$B_0^{-1} B_0 y_t = B_0^{-1} \gamma_0 + B_0^{-1} \gamma_1 y_{t-1} + B_0^{-1} u_t,$$

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t,$$

- where $B_0^{-1} \gamma_0 = c$, $B_0^{-1} \gamma_1 = \phi_1$ and $B_0^{-1} u_t = \varepsilon_t$.
- This representation is easy to estimate (OLS equation by equation).
- Notice that now we will only estimate 9 parameters!

Companion form

Rewriting a VAR(p) as a VAR(1).

A VAR(p) process can always be written as a VAR(1) one. To this end, define

$$\underset{np\times 1}{x_t} = \begin{pmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \dots \\ y_{t-p+1} - \mu \end{pmatrix}; \underset{np\times np}{F} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ I_n & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_n \end{pmatrix}; \underset{np\times 1}{\nu_t} = \begin{pmatrix} \varepsilon_t \\ 0 \\ \dots \\ 0 \end{pmatrix};$$

where $\mu = E\left(y_t\right)$. It follows that the VAR(p) process can be written as follows

$$x_t = F x_{t-1} + \nu_t, \tag{6}$$

where

$$E\left(\nu_t \nu_t'\right) = \begin{pmatrix} \Omega & \dots & 0 \\ 0 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 0 \end{pmatrix};$$

This representation is called the companion form of the VAR(p)

5.1. Stationarity of VAR(p) processes

Stationarity conditions for VAR(p) processes: Stability

A VAR(p) process y_t is said to be **stable** if all values of z satisfying

$$\left| I_n - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \right| = 0 \tag{7}$$

lie outside the unit circle.

Theorem

If the process y_t is stable, then it is covariance- stationary.

Remark: The reverse is not true. There exist stationary representations of processes that are unstable.

Another way of expressing the same result

- The process y_t is **stable** if the eigenvalues of the matrix F are all smaller than 1 in absolute value.

- Notice that an eigenvalue λ of F satisfies that

$$\left| I_n \lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - ... - \phi_p \lambda \right| = 0, \tag{8}$$

thus, if $|\lambda| < 1$, then y_t is stable.

- Why?

- Recall that a VAR(p) can be written as a VAR(1), $y_t = A y_{t-1} + \varepsilon_t$, then

$$Y_t = c + AY_{t-1} + \varepsilon_t$$

Substituting backward we obtain

$$\begin{aligned}
Y_t &= c + AY_{t-1} + \varepsilon_t \\
&= c + A(c + AY_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= (I + A)c + A^2 Y_{t-2} + A\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
Y_t &= (I + A + ... + A^{j-1})c + A^j Y_{t-j} + \sum_{i=0}^{j-1} A^i \varepsilon_{t-i}
\end{aligned}$$

If all the eigenvalues of A (the elements of the diagonal matrix Λ) are smaller than one in modulus then

1. $A^j = P\Lambda^j P^{-1} \to 0$.

2. the sequence A^i, $i = 0, 1, ...$ is absolutely summable.

3. the infinite sum $\sum_{i=0}^{j-1} A^i \varepsilon_{t-i}$ exists in mean square (by proposition C.10L);

4. $(I + A + ... + A^j)c \to (I - A)^{-1}c$ and $A^j \to 0$ as j goes to infinity.

Example A stationary VAR(1)

$$Y_t = AY_{t-1} + \epsilon_t,\ A = \begin{pmatrix} 0.5 & 0.3 \\ 0.02 & 0.8 \end{pmatrix},\ \Omega = E(\epsilon_t \epsilon_t') = \begin{pmatrix} 1 & 0.3 \\ 0.3 & .1 \end{pmatrix},\ \lambda = \begin{pmatrix} 0.81 \\ 0.48 \end{pmatrix}$$

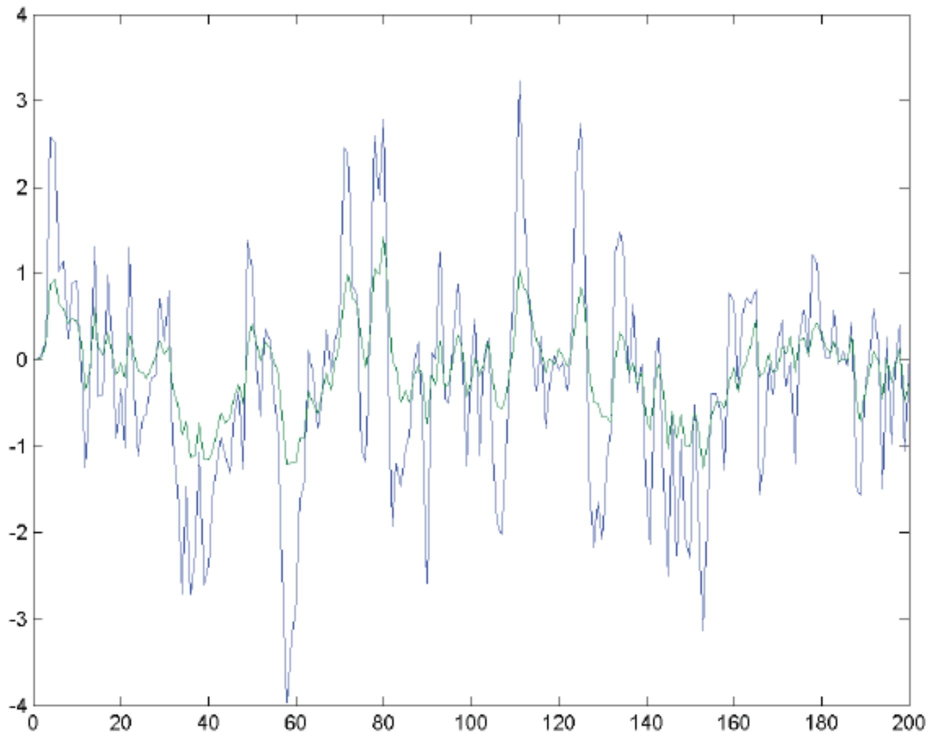


Figure 1: Blu: Y_1, green Y_2.

5.2. From the VAR to the Vector MA(∞) representation

Moving average representation of a stationary VAR process

- Consider first the VAR(1) case, $y_t = c + \phi_1 y_{t-1} + \varepsilon_t$. Iterating backwards:

$$y_t = (I_n + \phi_1 + ... + \phi_1^{t-1})c + \phi_1^t y_0 + \sum_{i=0}^{t-1} \phi_1^i \varepsilon_{t-i}.$$

If all the eigenvalues of ϕ_1 are smaller than 1 in abs. val., then the sequence ϕ_1^i is absolutely summable, implying that

$$\begin{aligned} & \phi_1^t y_0 \to 0 \\ & \sum_{i=0}^{t-1} \phi_1^i \varepsilon_{t-i} \to \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \\ & (I_n + \phi_1 + ... + \phi_1^{t-1})c \to c \sum_{i=0}^{\infty} \phi_1^i \\ = \ & \mu \end{aligned}$$

- This implies that

$$y_t = \mu + \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i},$$

where $\theta_i = \phi_1^i$.

- Notice that the first coefficient $\phi_1^0 = I_n$.

- For a general VAR(p) process

  - Write it first as a VAR(1)

  - Find the MA representation using the VAR(1) formulas.

The MA representation is not unique

- There are alternative MA(∞) representations based on vector white noise processes other than ε_t. To see that, consider a non-singular $n \times n$ matrix H and define

$$a_t = H\varepsilon_t$$

and

$$y_t = \mu + H^{-1}a_t + \phi_1 H^{-1}a_{t-1} + \phi_1^2 H^{-1}a_{t-2} + ...$$

5.3. Second Moments of the VAR representation

Let us consider the companion form of a stationary (zero mean for simplicity) VAR(p) process.

$$Y_t = AY_t - 1 + \varepsilon_t$$

The variance of Y_t is given by

$$\Sigma = E(Y_t Y_t') = A\Sigma A' + \Omega$$

where $E(\varepsilon\varepsilon') = \Omega$. A closed form solution for this formula can be obtained using the vec operator.

The vec operator is a linear transformation which converts the matrix into a column vector. Specifically, the vectorization of an m n matrix A, denoted vec(A), is the $mn \times 1$ column vector obtained by stacking the columns of the matrix A on top of one another:

You can read more about the vec operator here

https://en.wikipedia.org/wiki/Vectorization_(mathematics)

5.4. Estimation of VAR models

Let's consider now the estimation of model

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t.$$

The set of parameters to estimate is

- $\Pi = (c, \phi_1, ..., \phi_p)$,

- $\Omega.$, the variance-covariance matrix of ε_t

■ This model can be estimated by maximum likelihood (see Hamilton, chapter 11).

■ As in the univariate case, conditional MLE is equivalent to OLS estimation equation by equation.

Model selection

- Prior to estimating the VAR(p) process we need to select the order of the autoregressions.

- The same techniques as in the univariate case can be employed (AIC, BIC, HIC General-to-specific).

- Similar results hold in this case:

- AIC tends to choose larger models (it tends to overfit the model).

- BIC and HIC and consistent while AIC is not.

- Ivanov and Kilian (2005): study by simulation the properties of ICs. AIC seems to outperform for monthly data, HIC for quarterly data.

Reporting VAR results

- VAR models are particularly useful for describing the data and for constructing forecasts.

- Estimated VAR coefficients (they are difficult to interpret due to the complicated dynamic relationships) and R^2s typically go unreported.

- After VAR estimation, the following statistics are usually reported:

  - Granger-causality tests
  - IRFs
  - Forecast error variance decompositions

- We will use also the estimated coefficients to do forecasting

6. Forecasting: Introduction

1.1 Forecast based on conditional expectations

- Suppose we are interested in forecasting the value of Y_{t+1} based on a set of variables X_t.

- Let $Y_{t+1|t}$ denote such a forecast.

- To evaluate the usefulness of the forecast we need to specify a *loss* function. Here we specify a quadratic loss function. A quadratic loss function means that $Y_{t+1|t}$ is chose to minimize $E(Y_{t+1} - Y_{t+1|t})^2$.

- $E(Y_{t+1} - Y_{t+1|t})^2$ is known as the *mean squared error* associated with the forecast $Y_{t+1|t}$ denoted

$$MSE(Y_{t+1|t}) = E(Y_{t+1} - Y_{t+1|t})^2$$

- Fundamental result: the forecast with the smallest MSE is the expectation of $Y_{t+1|t}$ conditional on X_t that is

$$Y_{t+1|t} = E(Y_{t+1}|X_t)$$

We now verify the claim. Let $g(X_t)$ be any other function and let $Y_{t+1|t} = g(X_t)$. The associated MSE is

$$\begin{aligned} E\left[Y_{t+1} - g(X_t)\right]^2 &= E\left[Y_{t+1} - E(Y_{t+1}|X_t) + E(Y_{t+1}|X_t) - g(X_t)\right]^2 \\ &= E\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]^2 + \\ & +2E\left\{\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]\left[E(Y_{t+1}|X_t) - g(X_t)\right]\right\} + \\ & +E\{[E(Y_{t+1}|X_t) - g(X_t)\}^2 \end{aligned} \tag{2}$$

Define

$$\eta_{t+1} \equiv \left[Y_{t+1} - E(Y_{t+1}|X_t)\right]\left[E(Y_{t+1}|X_t) - g(X_t)\right] \tag{3}$$

The conditional expectation is

$$\begin{aligned} E(\eta_{t+1}|X_t) &= E\left\{\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]\left[E(Y_{t+1}|X_t) - g(X_t)\right]|X_t\right\} \\ &= \left[E(Y_{t+1}|X_t) - g(X_t)\right] E\left\{\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]|X_t\right\} \\ &= \left[E(Y_{t+1}|X_t) - g(X_t)\right]\left[E(Y_{t+1}|X_t) - E(Y_{t+1}|X_t)\right] \\ &= 0 \end{aligned}$$

Therefore by law of iterated expectations

$$E(\eta_{t+1}) = E(E(\eta_{t+1}|X_t)) = 0$$

This means that

$$E\left[Y_{t+1} - g(X_t)\right]^2 \;=\; E\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]^2 + E([E(Y_{t+1}|X_t) - g(X_t)])^2$$

Therefore the function that minimizes the MSE is

$$g(X_t) = E(Y_{t+1}|X_t)$$

$E(Y_{t+1}|X_t)$ is the optimal forecast of Y_{t+1} conditional of X_t under a quadratic loss function. The MSE of this optimal forecast is

$$E[Y_{t+1} - g(X_t)]^2 = E[Y_{t+1} - E(Y_{t+1}|X_t)]^2$$

1.2 Forecast based on linear projections

We now restrict the class of forecasts we consider to be linear function of X_t:

$$Y_{t+1|t} = \alpha' X_t$$

Suppose α' is such that the resulting forecast error is uncorrelated with X_t

$$E[(Y_{t+1} - \alpha' X_t) X_t'] = 0' \tag{4}$$

If (4) holds the we call $\alpha' X_t$ the linear projection of Y_{t+1} on X_t.

• The linear projection produces the smallest forecast error among the class of linear forecasting rules. To verify this let $g' X_t$ be any arbitrary forecasting rule.

$$\begin{aligned} E\left(Y_{t+1} - g' X_t\right)^2 &= E\left(Y_{t+1} - \alpha' X_t + \alpha' X_t - g' X_t\right)^2 \\ &= E\left(Y_{t+1} - \alpha' X_t\right)^2 + \\ &\quad +2E\left[\left(Y_{t+1} - \alpha' X_t\right)\left(\alpha' X_t - g' X_t\right)\right] + \\ &\quad +E\left(\alpha' X_t - g' X_t\right)^2 \end{aligned}$$

the middle term

$$\begin{aligned} E[(Y_{t+1} - \alpha' X_t)(\alpha' X_t - g' X_t)] &= E[(Y_{t+1} - \alpha' X_t) X_t' [\alpha - g]) \\ &= E[(Y_{t+1} - \alpha' X_t) X_t'] [\alpha - g] \\ &= 0' \end{aligned}$$

by definition of linear projection. Thus

$$E\left(Y_{t+1}-g'X_t\right)^2 \;=\; E\left(Y_{t+1}-\alpha'X_t\right)^2+E\left(\alpha'X_t-g'X_t\right)^2$$

The optimal linear forecast is the value $g'X_t = \alpha'X_t$. We use the notation

$$P(Y_{t+1}|X_t) = \alpha'X_t$$

to indicate the linear projection of Y_{t+1} on X_t.

Notice that

$$MSE[P(Y_{t+1}|X_t)] \geq MSE\left[E(Y_{t+1}|X_t)\right]$$

The projection coefficient α can be calculated in terms of moments of Y_{t+1} and X_t.

$$E(Y_{t+1}X_t') = \alpha'E(X_tX_t')$$

$$\alpha' = [E(X_tX_t')]^{-1}E(Y_{t+1}X_t')$$

Here we denote $\hat{Y}_{t+s|t} = P(Y_{t+s}|X_t, 1)$ the best linear forecast of Y_{t+s} conditional on X_t.

Forecasting with VAR models

Let us consider the VAR(p) process in companion form

$$Y_t = AY_{t-1} + \varepsilon_t$$

where ε_t is a white noise. The 1-step ahead linear forecast conditional on the information available up to time t is given by

$$Y_{t+1|t} = AY_t$$

where A is the linear projection of Y_{t+1} on Y_t. Iterating this expression, one can obtain the h-step ahead linear forecast of Y_t

$$Y_{t+h|t} = A^h Y_t$$

Forecasting with VAR models, II

This predictor delivers the minimum mean square error (MSE), [that is, the forecast with the smallest variance of the forecast error] among all the predictors constructed as linear combinations of Y_t

Using

$$Y_{t+h} = A^h Y_t + \sum_{i=0}^{h-1} A^i \varepsilon_{t+h-i}$$

we can obtain the forecast error:

$$Y_{t+h} - Y_{t+h|t} = \sum_{i=0}^{h-1} A^i \varepsilon_{t+h-i}$$

- From the forecast error we can obtain the MSE (the variance of the forecast error)

$$MSE(Y_{t+h|t}) = E(Y_{t+h} - Y_{t+h|t})(Y_{t+h} - Y_{t+h|t})' = \Sigma(h) = \sum_{i=0}^{h-1} A^i \Omega A^{i'}$$

- Notice that the MSE is non-decreasing in h.

$$MSE(Y_{t+h|t}) = \Sigma(h-1) + A^{h-1} \Omega A^{h-1'}$$

As $h \to \infty$, the MSE converges to the variance of Y_t.

Example, Stock and Watson 2001

Three-variable VAR: inflation rate, unemployment, interest rates.

Forecasting

Multistep-ahead forecasts, computed by iterating forward the reduced form VAR, are assessed in Table 2. Because the ultimate test of a forecasting model is its out-of-sample performance, Table 2 focuses on pseudo out-of-sample forecasts over the period from 1985:I to 2000:IV. It examines forecast horizons of two quarters, four quarters and eight quarters. The forecast *h* steps ahead is computed by estimating the VAR through a given quarter, making the forecast *h* steps ahead, reestimating the VAR through the next quarter, making the next forecast and so on through the forecast period.[6]

As a comparison, pseudo out-of-sample forecasts were also computed for a univariate autoregression with four lags—that is, a regression of the variable on lags of its own past values—and for a random walk (or "no change") forecast. Inflation rate forecasts were made for the average value of inflation over the forecast period, while forecasts for the unemployment rate and interest rate were made for the final quarter of the forecast period. Table 2 shows the root mean square forecast error for each of the forecasting methods. (The mean squared forecast error is computed as the average squared value of the forecast error over the 1985–2000 out-of-sample period, and the resulting square root is the root mean squared forecast error reported in the table.) Table 2 indicates that the VAR either does no worse than or improves upon the univariate autoregression and that both improve upon the random walk forecast.

Table 2

Root Mean Squared Errors of Simulated Out-Of-Sample Forecasts, 1985:1–2000:IV

Forecast Horizon	*Inflation Rate*			*Unemployment Rate*			*Interest Rate*		
	RW	*AR*	*VAR*	*RW*	*AR*	*VAR*	*RW*	*AR*	*VAR*
2 quarters	0.82	0.70	0.68	0.34	0.28	0.29	0.79	0.77	0.68
4 quarters	0.73	0.65	0.63	0.62	0.52	0.53	1.36	1.25	1.07
8 quarters	0.75	0.75	0.75	1.12	0.95	0.78	2.18	1.92	1.70

Notes: Entries are the root mean squared error of forecasts computed recursively for univariate and vector autoregressions (each with four lags) and a random walk ("no change") model. Results for the random walk and univariate autoregressions are shown in columns labeled RW and AR, respectively. Each model was estimated using data from 1960:I through the beginning of the forecast period. Forecasts for the inflation rate are for the average value of inflation over the period. Forecasts for the unemployment rate and interest rate are for the final quarter of the forecast period.

Granger Causality

7. Granger causality

- This concept was introduced by Granger (1969) ("Investigating Causal Relations by Econometric Models and Cross Spectral Methods", Econometrica, 37) and popularized by Sims (1972).

- 'Causality' here does not have the usual meaning of a 'causal effect'.

- Granger causality addresses the issue of how useful are some variables for **forecasting** others.

Simplest case: Bivariate Granger causality

Consider two economic variables y_t and x_t.

Intuition: If y_t helps to forecast x_t, then y_t Granger-causes x_t. If it cannot, y_t does not Granger-cause x_t.

More formally:

- Consider two forecasts of x_{t+h}, one using a linear function of past information on this variable (x_t, x_{t-1}, ...), denoted as $\hat{x}^{(1)}_{t+h} = \hat{E}\left(x_{t+h}|x_t, ...\right)$ and another that uses a linear function of (x_t, x_{t-1}, ...; y_t, $y_{t-1}, ...$), denoted as $\hat{x}^{(2)}_{t+h} = \hat{E}\left(x_{t+h}|x_t, ... , y_t, ...\right)$.

- Define the corresponding MSEs as $MSE^{(1)} = E\left(x_{t+h} - \hat{x}_{t+h}^{(1)}\right)^2$ and $MSE^{(2)} = E\left(x_{t+h} - \hat{x}_{t+h}^{(2)}\right)^2$.

- The variable y_t fails to Granger cause x if

$$MSE^{(1)} = MSE^{(2)}. \tag{9}$$

- Equivalently, if (9) holds, x is exogeneous in the time series sense with respect to y.

Tests for Granger Causality

- The are several ways to test for Granger Causality (see Hamilton, chapter 11).

- A simple approach would be to consider the regression

$$x_t = c_1 + \alpha_1 x_{t-1} + ... + \alpha_p x_{t-p} + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + u_t \tag{10}$$

for a particular lag length p, that can be estimated by OLS.

Then, conduct an F-test for the null hypothesis

$$H_0 : \beta_1 = ... = \beta_p = 0.$$

Computation.

- Estimate (10) and compute the sum of squared residuals:

$$RSS_1 = \sum_{t=1}^{T} \hat{u}_t^2.$$

- Estimate again (10) this time impossing H_0 and compute

$$RRS_0 = \sum_{t=1}^{T} \hat{e}_t^2$$

where $\hat{e}_t$ is obtained by estimating $x_t = \gamma_0 + \gamma_1 x_{t-1} + ... + \gamma_p x_{t-p} + e_t$ by OLS.

- Compute

$$F = \frac{T\left(RSS_0 - RSS_1\right)}{RSS_1}$$

- Under the assumption that x_t and y_t are stationary, reject H_0 if F is greater that the 5% values for a χ_p^2 variable.

Interpretation of Granger Causality

Relation between 'causality' and 'Granger causality'.

- Granger causality and causality are very different concepts. In fact they can run in the opposite direction.

- Example: consider a model of prices and dividends. One can set a model where stock prices reflect investor's perceptions of dividends. In spite of that, dividens fail to 'Granger cause' stock prices (=they are not useful for predicting stock prices).

- On the other hand, prices do 'Granger cause' dividents, even though the market's evaluation of the stock in reality has not effect on the dividend process (=the price of an stock reflects the beliefs on future dividends and therefore are useful for predicting future dividends).

- In general: time series that reflect forward-looking behavior, such as stock prices and interest rates, are often found to be excellent predictors of many key economic variables.

- But this does not mean that these series 'cause' GNP or inflation in the usual sense. Instead, these values reflect the market's best information about these variables.

- Granger causality tests should not be used to infer a direction of causation.

Application 1: Money, Income and Causality, Sims (1972)

"It has long been known that money stock and current dollar measures of economic activity are positively correlated.[...] A body of macro-economic theory, the "Quantity Theory," explains these empirical observations as reflecting a causal relation running from money to income. However, it is widely recognized that no degree of positive association between money and income can by itself prove that variation in money causes variation in income. Money might equally well react passively and very reliably to fluctuations in income." Sims (1972) "Money, Income, and Causality" AER.

Application 1: Money, Income and Causality, Sims (1972)

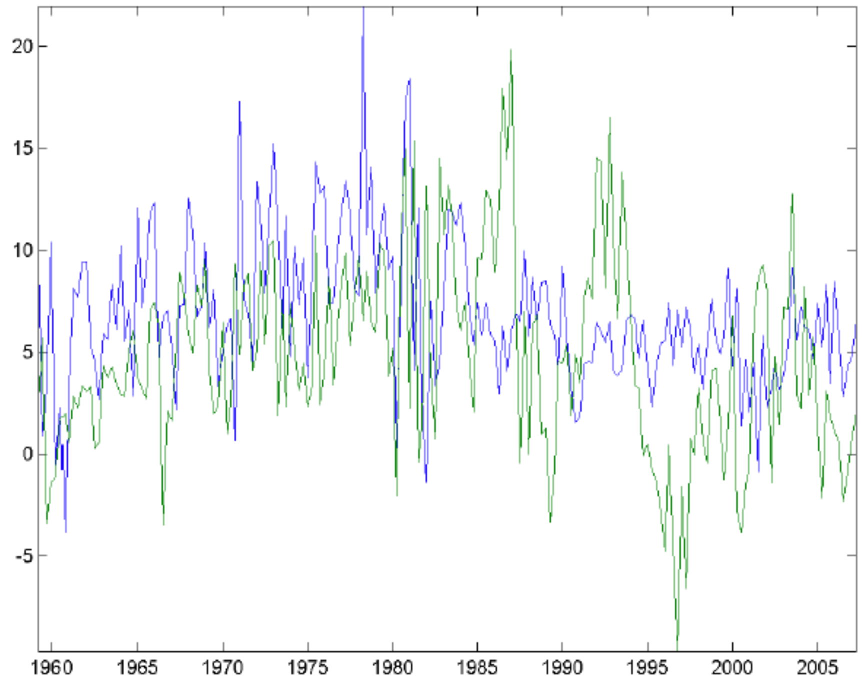


Figure 5: Blu: nominal gnp growth rates; green: M1 growth rates.

Application 1: Money, Income and Causality, Sims (1972)

Table 2: F-Tests of Granger Causality

	1959:II-1972:I	1959:II-2007:III
$M \to Y$	4.4440	2.2699
$Y \to M$	0.5695	3.5776
10%	2.0948	1.7071
5%	2.6123	1.9939
1%	3.8425	2.6187

In the first sample money Granger cause (at 5%) output but not the converse (Sims(72)'s result). In the second sample at the 5% both output Granger cause money and money Granger cause output.

Example 2

- It has been frequently found that interest rate spread (the difference between long and short yields) has been a good predictor for real GDP growth in the US (Estrella, 2000,2005).

- Recessions are often preceeded by a sharp fall in the spread (short increases compared to long rates).

- This implies that the spread should Granger cause output growth.

- However recent evidence suggests that its predictive power has decreased since the beginning of the 80s (see D'Agostino, Giannone and Surico, 2006).

- Thus, the spread should no longer cause output growth after mid 80's .

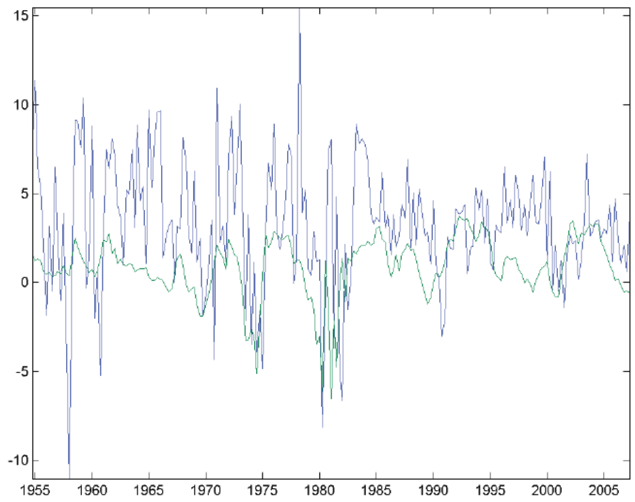


Figure 1: . Blue line is nominal GDP growth; Green line is M1 growth rates

We estimate a bivariate VAR for the growth rates of the real GDP and the difference between the 10-year rate and the federal funds rate. Data are from FRED StLouis Fed spanning from 1954:III-2007:III. The AIC criterion suggests $p = 6$.

Table 1: F-Tests of Granger Causality

	1954:IV-2007:III	1954:IV-1990:I	1990:I-2007:III
S_1	5.4233	6.0047	0.9687
10%	1.8050	1.8222	1.8954
5%	2.1460	2.1725	2.2864
1%	2.8971	2.9508	3.1864

We cannot reject the hypothesis that the spread does not Granger cause real output growth in the last period, while we reject the hypothesis for all the other sample. This can be explained by a change in the information content of private agents expectations, which is the information embedded in the yield curve.

Impulse Response Function

8. Impulse response functions for VAR processes

- Goal: to measure the dynamic reaction of a system of variables to a particular shock.

- More specifically: what's the response of one variable to an impulse in another variable in a system that involves a number of additional variables as well (that are assumed to remain constant)?

Responses to forecast errors (=the shocks of the system)

- Consider a stable VAR model that admits the following MA(∞) representation

$$y_t = c + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + ...$$

- we are now interested in the response of the variables in y_t to a change in the shock vector, ε_t, that is:

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t}.$$

- It is easy to check that in fact

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \psi_h.$$

- The matrix ψ_h is also called the matrix of dynamic multipliers.

- We can also consider the reaction of a particular variable in y_t, y_{it}, to a change of a shock ε_{jt}. This effect is measured by the corresponding element of the matrix ψ_h.

$$\frac{\partial y_{it+h}}{\partial \varepsilon_{jt}} = \psi_{ijh},$$

where ψ_{ijh} denotes the element (i,j) of ψ_h.

Computing the impulse response function from the VAR representation. An example

Consider the following VAR(1), where y1=investment, y2=income; y3=consumption.

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = \begin{pmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-2} \\ y_{3,t-3} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix}$$

Under these assumptions, at time $t = 0$ we will obtain:

$$y_0 = \begin{pmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \end{pmatrix} = \begin{pmatrix} u_{1,0} \\ u_{2,0} \\ u_{3,0} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

$$y_1 = \begin{pmatrix} y_{1,1} \\ y_{2,1} \\ y_{3,1} \end{pmatrix} = A_1 y_0 = \begin{pmatrix} .5 \\ .1 \\ 0 \end{pmatrix},$$

$$y_2 = \begin{pmatrix} y_{1,2} \\ y_{2,2} \\ y_{3,2} \end{pmatrix} = A_1 y_1 = A_1^2 y_0 = \begin{pmatrix} .25 \\ .06 \\ 0.2 \end{pmatrix},$$

The response of variable j to a unit shock in variable k is usually depicted graphically to get a visual impression of the dynamic interrelationships within the system.

Relation between Impulse responses and Granger causality

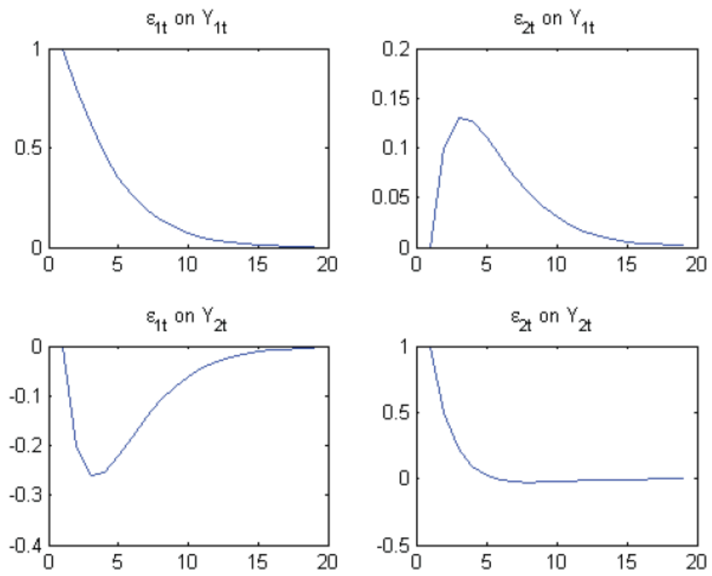
Remark: an innovation in variable k has no efect on the other variables if the former variable does not Granger-cause the set of the remaining variables.

Example 1 Let us assume that the estimated matrix of VAR coefficients is

$$A = \begin{pmatrix} 0.8 & 0.1 \\ -0.2 & 0.5 \end{pmatrix} \tag{33}$$

with eigenvalues 0.8562 and 0.4438. We generate impulse response functions of the Wold representation

$$C_j = A^j$$



Example 2 Let us now assume that the second variable does not Granger cause the first one so that

$$A = \begin{pmatrix} 0.8 & 0 \\ -0.2 & 0.5 \end{pmatrix} \tag{34}$$

with eigenvalues 0.8 and 0.5. Impulse response functions are plotted in the next figure

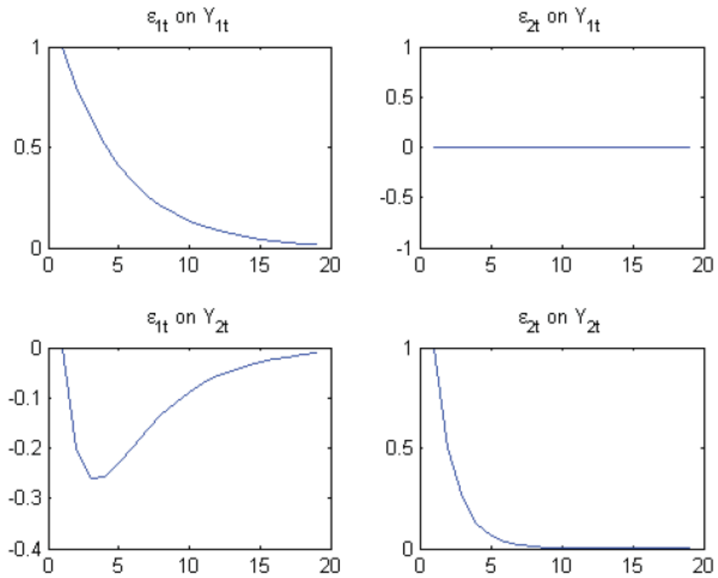


Figure 4: Impulse response functions when the second variable does not Granger cause the first one.

Responses to orthogonal impulses

- A problematic assumption in the previous framework is to consider that a shock occurs just in one variable.

- If shocks are correlated (as is usually the case in reduced-form VARs), a shock in one variable is likely to be accompanied by a shock in another variable.

- In this case, setting all other residuals to zero provide a misleading picture of the actual dynamic relationship.

- A possible solution would be to transform the system in such a way that the resulting shocks are **orthogonal.**

Recursive VARs

- Consider the following VAR

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t \tag{11}$$

where Σ_ε is the var-cov. matrix of ε_t. We will now look for a decomposition of Σ_ε such that $\Sigma_\varepsilon = W\Sigma^* W'$, where Σ^* is diagonal.

- To do that, consider the Cholesky decomposition of Σ_ε :

$$\Sigma_\varepsilon = PP'$$

where P is a lower triangular matrix.

- Define D as a diagonal matrix whose elements are those of the main diagonal of P and notice that

$$\Sigma_\varepsilon = \left(PD^{-1}\right) DD \left(D^{-1}P'\right)$$

Now, multiplying (11) by $A = \left(PD^{-1}\right)^{-1} = DP^{-1}$:

$$Ay_t = A_1 y_{t-1} + ... + A_p y_{t-p} + \omega_t$$

where $A_i = DP^{-1}\phi_i$, $\omega_t = DP^{-1}\varepsilon_t$ and $\Sigma_\omega = A\Sigma_\varepsilon A' = DD$, (a diagonal matrix).

Finally, adding $A_0 = I_k - A$ to both sides

$$y_t = A_0 y_t + A_1 y_{t-1} + ... + A_p y_{t-p} + \omega_t$$

Notice that A_0 is a lower triangular matrix with zero diagonal:- Thus the first equation of the system does not contain instantaneous $y's$ on the right hand side.

- The second equation may contain y_{1t} (but not y_{2t}, y_{3t}...)

- The third equation may contain y_{1t}, y_{2t} (but not $y_{3t}, ...$)

- This is the Recursive representation of a VAR process.

- Implication: the ordering of the variables matters, that is, it is important which of the variables is called y_1, y_2, etc

- Moreover, the ordering cannot be determined by using statistical methods but it should be determined by the analyst: thus, different analysts can reach to different conclusions.

- Determining the ordering in fact can be quite difficult.

- Alternative: relying on the 'structural' impulse response functions.

Example: Stock and Watson (2001)

- SW consider a three-variable VAR relating inflation, unemployment rate and the interest rate

- They order these variables as 1) inflation: 2) the inflation rate and 3) the interest rate.

- Then, in the first regression of the VAR, inflation is the dependent variable and the regressors are lagged values of all three variables.

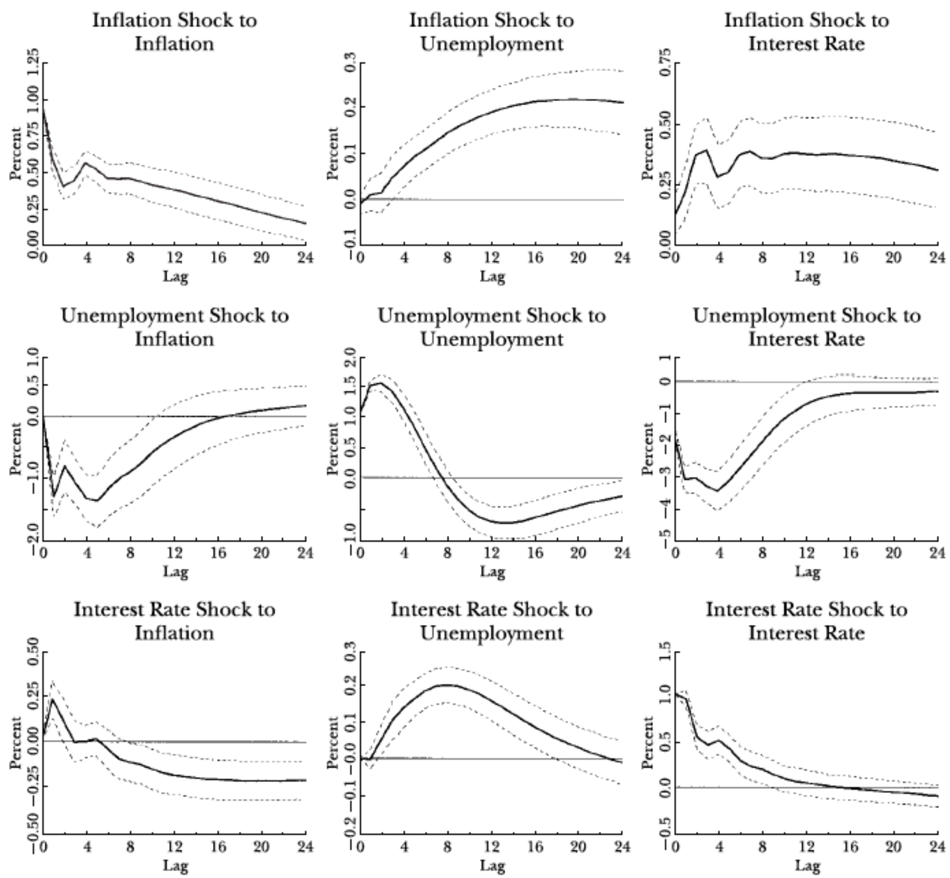
- In the second equation, the unemployment variable is the dependent variable and the regressors are the lags of all three variable plus the current value of the inflation rate.

- The third equation (interest rate) will contain lagged regressors and the contemporaneous values of inflation and the inflation rate.

A *recursive VAR* constructs the error terms in each regression equation to be uncorrelated with the error in the preceding equations. This is done by judiciously including some contemporaneous values as regressors. Consider a three-variable VAR, ordered as 1) inflation, 2) the unemployment rate, and 3) the interest rate. In the first equation of the corresponding recursive VAR, inflation is the dependent variable, and the regressors are lagged values of all three variables. In the second equation, the unemployment rate is the dependent variable, and the regressors are lags of all three variables *plus* the current value of the inflation rate. The interest rate is the dependent variable in the third equation, and the regressors are lags of all three variables, the current value of the inflation rate *plus* the current value of the unemployment rate. Estimation of each equation by ordinary least squares produces residuals that are uncorrelated across equations.[4] Evidently, the results depend on the order of the variables: changing the order changes the VAR equations, coefficients, and residuals, and there are $n!$ recursive VARs representing all possible orderings.

Figure 1

Impulse Responses in the Inflation-Unemployment-Interest Rate Recursive VAR



Impulse responses trace out the response of current and future values of each of the variables to a one-unit increase in the current value of one of the VAR errors, assuming that this error returns to zero in subsequent periods and that all other errors are equal to zero. The implied thought experiment of changing one error while holding the others constant makes most sense when the errors are uncorrelated across equations, so impulse responses are typically calculated for recursive and structural VARs.

The impulse responses for the recursive VAR, ordered π_t, u_t, R_t, are plotted in Figure 1. The first row shows the effect of an unexpected 1 percentage point increase in inflation on all three variables, as it works through the recursive VAR system with the coefficients estimated from actual data. The second row shows the effect of an unexpected increase of 1 percentage point in the unemployment rate, and the third row shows the corresponding effect for the interest rate. Also plotted are ± 1 standard error bands, which yield an approximate 66 percent confidence interval for each of the impulse responses. These estimated impulse responses show patterns of persistent common variation. For example, an unexpected rise in inflation slowly fades away over 24 quarters and is associated with a persistent increase in unemployment and interest rates.

The impulse response function: estimation

There are several ways to compute this function.

An easy way to do it is by simulation. The following steps could be followed:

1. Estimate a suitable VAR(p) model for y_t.

2. Set $y_{t-1} = y_{t-2} = y_{t-p} = 0$. Set $\varepsilon_{jt} = 1$ and all other elements of ε_t to zero.

3. Simulate the VAR(p) process that you have estimated.

4. The value of the vector y_{t+h} at date $t+h$ of this simulation corresponds to the jth column of the matrix ψ_h.

5. By repeting the same procedure for other $\varepsilon_{it} = 1$ we can obtain all the columns of ψ_h.

Error Bands for Impulse Response Functions

Method I: Asymptotic Hamilton (1994).

Method II: Bootstrap The idea behind bootstrapping (Runkle, 1987) is to obtain estimates of the small sample distribution for the impulse response functions without assuming that the shocks are Gaussian. Steps:

1. Estimate the VAR and save the $\widehat{\pi}$ and the fitted residuals $\{\widehat{u}_1, \widehat{u}_2, ..., \widehat{u}_T\}$.

2. Draw uniformly from $\{\widehat{u}_1, \widehat{u}_2, ..., \widehat{u}_T\}$ and set $\tilde{u}_1^{(1)}$ equal to the selected realization and use this to construct

$$Y_1^{(1)} = \widehat{A}_1 Y_0 + \widehat{A}_2 Y_{-1} + ... + \widehat{A}_p Y_{-p+1} + \tilde{u}_1^{(1)} \tag{20}$$

3. Taking a second draw (with replacement) $\tilde{u}_2^{(1)}$ generate

$$Y_2^{(1)} = \widehat{A}_1 Y_1^{(1)} + \widehat{A}_2 Y_0 + ... + \widehat{A}_p Y_{-p+2} + \tilde{u}_2^{(1)} \tag{21}$$

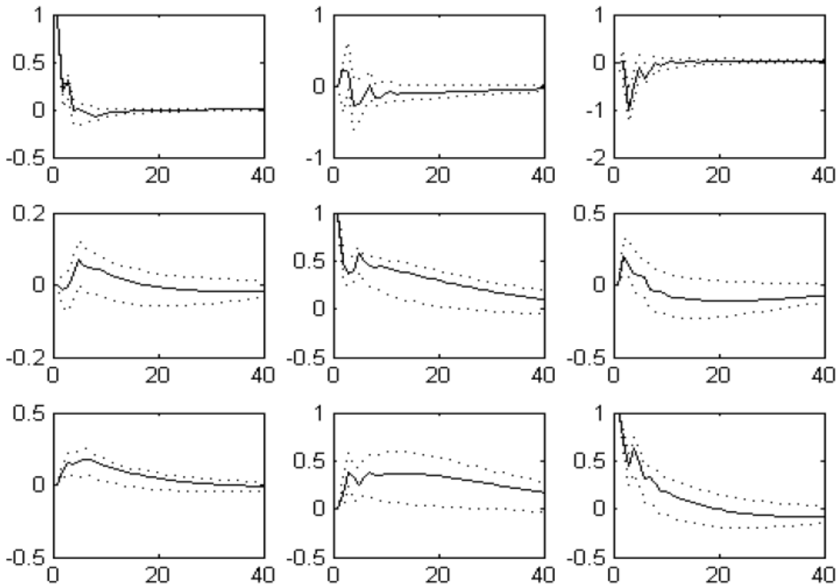
4. Proceeding in this fashion generate a sample of length T $\{Y_1^{(1)}, Y_2^{(1)}, ..., Y_T^{(1)}\}$ and use the sample to compute $\widehat{\pi}^{(1)}$ and the implied impulse response functions $C^{(1)}(L)$.

5. Repeat steps $(3)-(4)$ M times and collect M realizations of $C^{(l)}(L)$, $l = 1, ...M$ and takes for all the elements of the impulse response functions and for all the horizons the αth and $1-\alpha$th percentile to construct confidence bands.

Method III: Montecarlo As bootstrap but drawing the residuals from a normal distribution.

Example: A Monetary VAR with bootstrap bands

We estimate the standard monetary VAR which includes real output growth, the inflation rate and the federal funds rate. These three variables are the core variables for monetary policy analysis in VAR models. Data are taken from the St.Louis Fed FREDII database.



Variance Decomposition

- Idea: decompose the total variance of a time series into the percentages attributable to each structural shock.

- What type of questions can it answer:

  - What are the sources of the business cycle?

  - Is the shock "k" important for economic fluctuations?

Computation

- Consider the MA(∞) representation of y_t:

$$y_t = F(L)w_t$$

where $w_t = \Sigma_\varepsilon^{-1/2}\varepsilon_t,\ E(\varepsilon_t\varepsilon_t') = \Sigma_\varepsilon,\ E(w_t w_t') = I_n,\ F(L) = C(L)\Sigma_\varepsilon^{1/2}$.

- Then the variance of y_{it} is given by

$$var(y_{it}) = \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j^2} var(w_{kt})$$

$$= \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j^2}$$

Variance decomposition, cont.

- The variance of y_{it} generated by the kth shock

$$\sum_{j=0}^{\infty} {F_{ik}^{j}}^{2}$$

- Thus

$$\frac{\sum_{j=0}^{\infty} {F_{ik}^{j}}^{2}}{\sum_{k=1}^{n}\sum_{j=0}^{\infty} {F_{ik}^{j}}^{2}}$$

measures the percentage of the variance of y_{it} explained by the kth shock

Structural VAR model

- We have seen that impulse responses are an important tool to uncover the relations between the variables in a VAR.

- But there are some obstacles in their interpretation: the same underlying VAR can give rise to different IRFs.

- Thus, additional information (economic information) is needed to identify has to be used to decide on the proper set of IRFs for a particular model.

- Structural restrictions will be needed to identify the relevant innovations and IRFs. Selecting a particular set of structural restrictions will allow us to identify the **structural VAR model.**

Structural VAR models: identification strategy

- Assume you have estimated a reduced-form VAR for y_t. Would it be possible to recover the parameters of the original structural VAR model?

- Recall that the structural VAR is given by

$$B_0 y_t = B_1 y_{t-1} + ... + B_p y_{t-p} + u_t. \tag{12}$$

The number of parameters to be estimated is: $(p+1)\,n^2 + n(n+1)/2$

- The corresponding VAR is obtained pre-multiplying the previous model by B_0^{-1} :

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t.$$

The number of parameters to be estimated is: $pn^2 + n(n+1)/2$;

Identification of the structural VAR model

The structural VAR has n^2 more parameters than the reduced-form VAR model.

Order condition for identification: n^2 restrictions should be introduced.

- Sims (1980): After his influential paper, the additional restrictions are usually imposed on

- the covariance matrix of structural shocks Σ_u;
- the matrix of contemporaneous coefficients B_0
- and the matrix of long run multpliers, $A(1)^{-1}$.

- First n restrictions: set the elements of the diagonal of B_0 to be equal to 1.

- This leaves $n\,(n-1)$ restrictions to be imposed that must be deduced from economic considerations.

- Restrictions of Σ_u : It is usually assumed that this matrix is diagonal implying that contemporaneous shocks are not correlated. This imposes $n(n-1)/2$ additional restrictions.

- Thus, the conditions on B_0 (n) and on Σ_u (n$(n-1)\,/2)$ leave us with the need for looking for $n\,(n-1)\,/2$ additional restrictions.

- It is sometimes imposed that the matrix B_0 is lower triangular, yielding the required additional $n\,(n-1)\,/2$ conditions. (Sims, 1980).

- An alternative set of identification restrictions relies on long-run relationships, see Blanchard and Quah (1989) and King et al. (1991).

- These papers rely on restrictions on $B(1) = B_0 - \sum^p B_i$.

- Since $B(1) = C(1)^{-1}$, these restrictions can be viewed as restrictions of the sum of the impulse responses.

Last remarks on the interpretation of VAR models

Innovation accounting and impulse response analysis in the framework of VAR models have been pioneered by Sims (1980, 1981) and others as an alternative to classical macroeconomic analyses. Sims' main criticism of the latter type of analysis is that macroeconometric models are often not based on sound economic theories or the available theories are not capable of providing a completely specified model. If economic theories are not available to specify the model, statistical tools must be applied. In this approach, a fairly loose model is set up which does not impose rigid a priori restrictions on the data generation process. Statistical tools are then used to determine possible constraints. VAR models represent a class of loose models that may be used in such an approach. Of course, in order to interpret these models, some restrictive assumptions need to be made. In particular, the ordering of the variables may be essential for interpretations of the types discussed in the previous subsections. Sims (1981) suggests to try different orderings and investigate the sensitivity of the corresponding orthogonalized impulse responses and the related conclusions to the ordering of the variables.