Vector autoregressions

Based on the book 'New Introduction to Multiple Time Series Analysis' by Helmut Lütkepohl'

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

November 3, 2011

Outline

Objectives of analyzing multiple time series

Main objectives of time series analysis may be:

1. *Forecasting*: prediction of the unknown future by looking at the known past:

$$\hat{y}_{T+h} = f(y_T, y_{T-1}, \ldots)$$

   denotes the h-step prediction for the variable y;
2. Quantifying the dynamic response to an unexpected shock to a variable by the same variable h periods later and also by other related variable: *impulse-response analysis*;
3. *Control*: how to set a variable in order to achieve a given time path in another variable; description of *system dynamics* without further purpose.

Some basics: stochastic process

Assume a *probability space* $(\Omega, \mathcal{F}, Pr)$. A (discrete) *stochastic process* is a real-valued function

$$y : Z \times \Omega \to \mathbb{R},$$

such that, for each fixed $t \in Z$, $y(t, \omega)$ is a random variable. Z is a useful index set that represents time, for example $Z = \mathbb{Z}$ or $Z = \mathbb{N}$.

Some basics: multivariate stochastic process

A (discrete) *K–dimensional vector stochastic process* is a real-valued function

$$y : Z \times \Omega \to \mathbb{R}^K,$$

such that, for each fixed $t \in Z$, $y(t, \omega)$ is a K–dimensional random vector.

A *realization* is a sequence of vectors $y_t(\omega), t \in Z$, for a fixed ω. It is a function $Z \to \mathbb{R}^K$. A *multiple time series* is assumed to be a finite portion of a realization.

Given such a realization, the underlying stochastic process is called the *data generation process* (DGP).

Vector autoregressive processes

Let $y_t = (y_{1t}, \ldots, y_{Kt})'$, $\nu = (\nu_1, \ldots, \nu_K)'$, and

$$A_j = \begin{bmatrix} \alpha_{11,j} & \cdots & \alpha_{1K,j} \\ \vdots & \ddots & \vdots \\ \alpha_{K1,j} & \cdots & \alpha_{KK,j} \end{bmatrix}.$$

Then, a vector autoregressive process (VAR) satisfies the equation

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t,$$

with u_t a sequence of independently identically distributed random K–vectors with zero mean (conditions relaxed later).

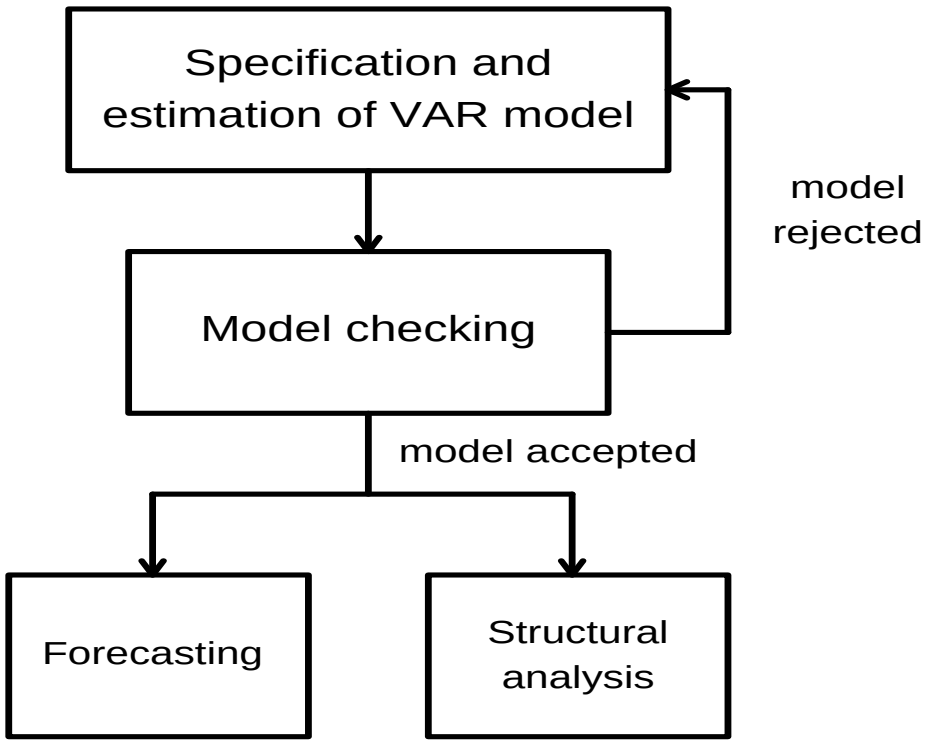
Forecasting using a VAR

Assume y_t follows a VAR(p). Then, the forecast $\hat{y}_{T+1}$ is given by

$$\hat{y}_{T+1} = \nu + A_1 y_T + \ldots + A_p y_{T-p+1},$$

i.e. the systematic part of the defining equation. Note that this also defines a forecast for each component of y_{T+1}.

A flowchart for VAR analysis



The VAR(p) model

The object of interest is the vector autoregressive process of order p that satisfies the equation

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \quad t = 0, \pm 1, \pm 2, \ldots$$

with u_t assumed as K–dimensional white noise, i.e. $\mathrm{E}u_t = 0$, $\mathrm{E}u_s u_t' = 0$ for $s \neq t$, and $\mathrm{E}u_t u_t' = \Sigma$ with nonsingular Σ (conditions relaxed).

First we concentrate on the VAR(1) model

$$y_t = \nu + A_1 y_{t-1} + u_t.$$

Substituting in the VAR(1)

Continuous substitution in the VAR(1) model yields

$$
\begin{aligned}
y_1 &= \nu + A_1 y_0 + u_1, \\
y_2 &= (I_K + A_1)\nu + A_1^2 y_0 + A_1 u_1 + u_2, \\
&\vdots \\
y_t &= (I_K + A_1 + \ldots + A_1^{t-1})\nu + A_1^t y_0 + \sum_{j=0}^{t-1} A_1^j u_{t-j},
\end{aligned}
$$

such that $y_1, \ldots, y_t$ can be represented as a function of $y_0, u_1, \ldots, u_t$. All $y_t, t \geq 0$, are a function of just one starting value and the errors.

The Wold representation of the VAR(1)

If all eigenvalues of A_1 have modulus less than one, substitution can be continued using the $y_j, j < 0$, and the limit exists:

$$y_t = (I_K - A_1)^{-1}\nu + \sum_{j=0}^{\infty} A_1^j u_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots,$$

and the constant portion can be denoted by μ.

The matrix sequence converges according to linear algebra results. The random vector converges in mean square due to an important statistical lemma.

Convergence of sums of stochastically bounded processes

Theorem

Suppose (A_j) *is an absolutely summable sequence of real* $(K \times K)$*–matrices and* (z_t) *is a sequence of* K*–dimensional random variables that are bounded by a common* $c \in \mathbb{R}$ *in the sense of*

$$\mathrm{E}(z_t' z_t) \leq c, \quad t = 0, \pm 1, \pm 2, \ldots.$$

Then there exists a sequence of random variables (y_t)*, such that*

$$\sum_{j=-n}^{n} A_j z_{t-j} \longrightarrow y_t,$$

as $n \longrightarrow \infty$*, in quadratic mean.* (y_t) *is uniquely defined except on a set of probability 0.*

Aspects of the convergent sum

The matrices converge geometrically and hence absolutely, and the theorem applies. The limit in the 'Wold' representation is well defined.

- This is called a Wold representation, as Wold's Theorem provides an infinite-order moving-average representation for all univariate covariance-stationary processes.
- Note that the white-noise property was not used. The sum would even converge for time-dependent u_t.

Expectation of the stationary VAR(1)

The Wold-type representation implies

$$\mathrm{E}(y_t) = (I_K - A_1)^{-1}\nu = \mu.$$

This is due to the fact that $\mathrm{E}u_t = 0$ for the white-noise terms and a statistical theorem that permits exchanging the limit and expectation operations under the conditions of the lemma. Note that the white-noise property (uncorrelated sequence) is not used.

Second moments of the stationary VAR(1)

LUETKEPOHL presents a derivation of the cross-covariance function

$$\begin{aligned}\Gamma_y(h) &= \mathrm{E}(y_t - \mu)(y_{t-h} - \mu)' \\ &= \lim_{n\to\infty} \sum_{i=0}^{n}\sum_{j=0}^{n} A_1^i \mathrm{E}(u_{t-i}u'_{t-j-h})(A_1^j)' \\ &= \lim \sum_{i=0}^{n} A_1^{h+i}\Sigma_u(A_1^i)' = \sum_{i=0}^{\infty} A_1^{h+i}\Sigma_u(A_1^i)',\end{aligned}$$

which uses $\mathrm{E}(u_t u'_s) = 0$ for $s \neq t$, $\mathrm{E}(u_t u'_t) = \Sigma_u$, and a corollary to the lemma that permits evaluation of second moments under the same conditions. Here, the white-noise property of u_t is used.

The definition of a stable VAR(1)

Definition

A VAR(1) is called *stable* iff all eigenvalues of A_1 have modulus less than one. By a mathematical lemma, this condition is equivalent to

$$\det(I_K - A_1 z) \neq 0 \quad \text{for} \quad |z| \leq 1.$$

No roots within or on the unit circle. Note that this definition differs from stability as defined by other authors. Stability is not equivalent to stationarity: a stable process started in $t = 1$ is not stationary; a backward-directed entirely unstable process is stationary.

Representation of VAR(p) as VAR(1)

All VAR(p) models of the form

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t$$

can be written as VAR(1) models

$$Y_t = \nu^\dagger + \mathbf{A} Y_{t-1} + U_t,$$

with

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & 0 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & I_K & 0 \end{bmatrix}.$$

More on the state-space VAR(1) form

In the VAR(1) representation of a VAR(p), the vectors Y_t, $\nu^\dagger$, and U_t have length Kp:

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \dots \\ y_{t-p+1} \end{bmatrix}, \quad \nu^\dagger = \begin{bmatrix} \nu \\ 0 \\ \dots \\ 0 \end{bmatrix}, \quad U_t = \begin{bmatrix} u_t \\ 0 \\ \dots \\ 0 \end{bmatrix}.$$

The big matrix $\mathbf{A}$ has dimension $Kp \times Kp$. This state-space form permits using all results from VAR(1) for the general VAR(p).

Stability of the VAR(p)

Definition

A VAR(p) is called *stable* iff all eigenvalues of $\mathbf{A}$ have modulus less than one. By a mathematical lemma, this condition is equivalent to

$$\det(I_{Kp} - \mathbf{A}z) \neq 0 \quad \text{for} \quad |z| \leq 1.$$

This condition is equivalent to the *stability condition*

$$\det(I_K - A_1 z - \ldots - A_p z^p) \neq 0 \quad \text{for} \quad |z| \leq 1,$$

which is usually more efficient to check. Equivalence follows from the determinant properties of partitioned matrices.

The infinite-order MA representation of the VAR(p)

The stationary stable VAR(p) can be represented in the convergent infinite-order MA form

$$Y_t = \mu^\dagger + \sum_{j=0}^{\infty} \mathbf{A}^j U_{t-j}.$$

This is, however, still an inconvenient process of dimension Kp. Formally, the first K entries of the vector Y_t are obtained via the $(K \times Kp)$–matrix

$$J = [I_K : 0 : \ldots : 0]$$

as $y_t = JY_t$.

The Wold representation of the VAR(p)

Using J, it follows that

$$\begin{aligned} y_t &= J\mu^\dagger + J\sum_{j=0}^{\infty} \mathbf{A}^j U_{t-j} \\ &= \mu + \sum_{j=0}^{\infty} J\mathbf{A}^j J' J U_{t-j} \\ &= \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-j} \end{aligned}$$

for the stable and stationary VAR(p), a Wold representation with $\Phi_j = J\mathbf{A}^j J'$. This is the *canonical* or *fundamental* or *prediction-error representation*.

First and second moments of the VAR(p)

Applying the lemma to the MA representation yields $E(y_t) = \mu$ and

$$\begin{aligned}
\Gamma_y(h) &= \mathrm{E}(y_t - \mu)(y_{t-h} - \mu)' \\
&= \mathrm{E}\left(\sum_{i=0}^{h-1} \Phi_i u_{t-i} + \sum_{i=0}^{\infty} \Phi_{h+i} u_{t-h-i}\right)\left(\sum_{j=0}^{\infty} \Phi_j u_{t-h-j}\right)' \\
&= \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_u \Phi_i'
\end{aligned}$$

The Wold-type representation with lag operators

Using the operator L defined by $Ly_t = y_{t-1}$ permits writing the AR(p) model as

$$y_t = \nu + (A_1 L + \ldots + A_p L^p) y_t + u_t$$

or, with $A(L) = 1 - A_1 L - \ldots - A_p L^p$,

$$A(L) y_t = \nu + u_t.$$

Then, one may write $\Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j$ and

$$y_t = \mu + \Phi(L) u_t = A^{-1}(L)(\nu + u_t),$$

thus formally $A(L)\Phi(L) = I$ or $\Phi(L) = A^{-1}(L)$. Note that $A(L)$ is a polynomial and $\Phi(L)$ is a power series.

Remarks on the lag operator representation

- The property $\Phi(L)A(L) = I$ allows to determine Φ_j iteratively by comparing coefficient matrices;
- Note that $\mu = A^{-1}(L)\nu = A^{-1}(1)\nu$ and that $A(1) = 1 - A_1 - \ldots - A_p$;
- It is possible that $A^{-1}(L)$ is a finite-order polynomial, while this is impossible for scalar processes;
- The MA representation exists iff the VAR(p) is stable, i.e. iff all zeros of $\det(A(z))$ are outside the unit circle: $A(L)$ is called *invertible*.

Remarks on stationarity

Formally, *covariance stationarity* of K–variate processes is defined by constancy of first moments $\mathrm{E}y_t = \mu \quad \forall t$ and of second moments

$$\mathrm{E}(y_t - \mu)(y_{t-h} - \mu)' = \Gamma_y(h) = \Gamma_y(-h)' \quad \forall t, h = 0, 1, 2, \ldots$$

Strict stationarity is defined by time invariance of all finite-dimensional joint distributions. Here, 'stationarity' refers to covariance stationarity, for example in the proposition:

Proposition

A stable VAR(p) process $y_t, t \in \mathbb{Z}$, is stationary.

Yule-Walker equations for VAR(1) processes

Assume the VAR(1) is stable and stationary. The equation

$$y_t - \mu = A_1(y_{t-1} - \mu) + u_t$$

can be multiplied by $(y_{t-h} - \mu)'$ from the right. Application of expectation yields

$$\mathrm{E}(y_t - \mu)(y_{t-h} - \mu)' = A_1 \mathrm{E}\{(y_{t-1} - \mu)(y_{t-h} - \mu)'\} + \mathrm{E}u_t(y_{t-h} - \mu)'$$

or

$$\Gamma_y(h) = A_1 \Gamma_y(h-1)$$

for $h \geq 1$.

The system of Yule-Walker equations for VAR(1)

For the case $h = 0$, the last term is not 0:

$$\mathrm{E}(y_t - \mu)(y_t - \mu)' = A_1 \mathrm{E}\{(y_{t-1} - \mu)(y_t - \mu)'\} + \mathrm{E}u_t(y_t - \mu)'$$

or

$$\Gamma_y(0) = A_1\Gamma_y(-1) + \Sigma_u = A_1\Gamma_y(1)' + \Sigma_u,$$

which by substitution from the equation for $h = 1$ yields

$$\Gamma_y(0) = A_1\Gamma_y(0)A_1' + \Sigma_u,$$

which can be transformed to

$$\mathrm{vec}\Gamma_y(0) = (I_{K^2} - A_1 \otimes A_1)^{-1}\mathrm{vec}\Sigma_u,$$

an explicit formula to obtain the process variance from given coefficient matrix and error variance.

How to use the Yule-Walker equations for VAR(1)

- For synthetic purposes, first evaluate $\Gamma_y(0)$ from given A_1 and Σ_u;
- Then, the entire ACF is obtained from $\Gamma_y(h) = A_1^h \Gamma_y(0)$;
- The big matrix in the $h = 0$ equation must be invertible, as the eigenvalues of $A_1 \otimes A_1$ are the squares of the eigenvalues of A_1, which have modulus less than one;
- Sure, the same trick works for VAR(p), as they have a VAR(1) representation, but you have to invert $((Kp)^2 \times (Kp)^2)$–matrices;
- For analytic purposes, $A_1 = \Gamma_0 \Gamma_1^{-1}$ can be used to estimate A_1 from the correlogram.

Autocorrelations of stable VAR processes

Autocorrelations are often preferred to autocovariances. Formally, they are defined via

$$\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)}\sqrt{\gamma_{jj}(0)}}$$

from the autocovariances for $i, j = 1, \ldots, K$ and $h \in \mathbb{Z}$. The matrix formula

$$R_y(h) = D^{-1}\Gamma_y(h)D^{-1}$$

with $D = \operatorname{diag}(\gamma_{11}(0)^{1/2}, \ldots, \gamma_{KK}(0)^{1/2})$ is given for completeness.

The forecasting problem

Based on an information set $\Omega_t \supseteq \{y_s, s \leq t\}$ available at t, the forecaster searches an approximation $\tilde{y}_t(h)$ to the 'unknown' y_{t+h} that minimizes some expected loss or cost

$$\mathrm{E}\{g(y_{t+h} - \tilde{y}_t(h))|\Omega_t\}.$$

The most common loss function $g(x) = x^2$ minimizes the forecast mean squared errors (MSE). t is the forecast *origin*, h is the forecast *horizon*, $\tilde{y}_t(h)$ is an *h–step predictor*.

Conditional expectation

Proposition

The h–step predictor that minimizes the forecast MSE is the conditional expectation

$$\tilde{y}_t(h) = \mathrm{E}(y_{t+h}|y_s, s \leq t).$$

Often, the casual notation $\mathrm{E}_t(y_{t+h})$ *is used.*

This property (proof constructive) also applies to vector processes and to VARs, where the MSE is defined by

$$MSE(y_t(h)) = \mathrm{E}\{y_{t+h} - \tilde{y}_t(h)\}\{y_{t+h} - \tilde{y}_t(h)\}'.$$

Conditional expectation in a VAR

Assume u_t is *independent* white noise (martingale difference sequence with $\mathrm{E}(u_{t+1}|u_s, s \leq t) = 0$ suffices), then for a VAR(p)

$$\mathrm{E}_t(y_{t+1}) = \nu + A_1 y_t + A_2 y_{t-1} + \ldots + A_p y_{t-p+1},$$

and, recursively,

$$\mathrm{E}_t(y_{t+2}) = \nu + A_1 \mathrm{E}_t(y_{t+1}) + A_2 y_t + \ldots + A_p y_{t-p+2},$$

etc., which allows the iterative evaluation for all horizons.

Larger horizons for a VAR(1)

By repeated insertion, the following formula is easily obtained:

$$\mathrm{E}_t(y_{t+h}) = (I_K + A_1 + \ldots + A_1^{h-1})\nu + A_1^h y_t,$$

which implies that the forecast tends to become trivial as h increases, given the geometric convergence in the last term.

Forecast MSE for VAR(1)

The MA representation $y_t = \mu + \sum_{j=0}^{\infty} A_1^j u_{t-j}$ clearly decomposes y_{t+h} into the predictor known in t and the remaining error, such that

$$y_{t+h} - y_t(h) = \sum_{j=0}^{h-1} A_1^j u_{t+h-j},$$

and

$$\begin{aligned} \Sigma_y(h) &= \mathrm{MSE}(y_t(h)) = \mathrm{E}\left(\sum_{j=0}^{h-1} A_1^j u_{t+h-j}\right)\left(\sum_{j=0}^{h-1} A_1^j u_{t+h-j}\right)' \\ &= \sum_{j=0}^{h-1} A_1^j \Sigma_u (A_1^j)' = \mathrm{MSE}(y_t(h-1)) + A_1^{h-1}\Sigma_u (A_1^{h-1})', \end{aligned}$$

such that MSE increases in h.

Forecast MSE for general VAR(p)

Using the Wold-type MA representation $y_t = \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-j}$, a scheme analogous to $p = 1$ works for VAR(p) with $p > 1$, using J. The forecast error variance is

$$\Sigma_y(h) = \mathrm{MSE}(y_t(h)) = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j',$$

which converges to $\Sigma_y = \Gamma_y(0)$ for $h \to \infty$.

These MSE formulae can also be used to determine interval forecasts (confidence intervals).

Structural VAR analysis

There are three (interdependent) approaches to the interpretation of VAR models:

1. Granger causality
2. Impulse response analysis
3. Forecast error variance decomposition (FEVD)

Granger causality

Assume two M- and N–dimensional sub-processes x and z of a K–dimensional process y, such that $y = (z', x')'$.

Definition

The process x_t is said to *cause* z_t in Granger's sense iff

$$\Sigma_z(h|\Omega_t) < \Sigma_z(h|\Omega_t \setminus \{x_s, s \leq t\})$$

for some t and h.

The set Ω_t is an information set containing $y_s, s \leq t$; the matrix $<$ is defined via positive definiteness of the difference; the correct interpretation of the $\setminus$ operator is doubtful.

The property is not antisymmetric: x may cause z and z may also cause x: *feedback*.

Instantaneous Granger causality

Again, assume two M- and N–dimensional sub-processes x and z of a K–dimensional process y.

Definition

There is *instantaneous causality* between process x_t and z_t in Granger's sense iff

$$\Sigma_z(1|\Omega_t \cup \{x_{t+1}\}) < \Sigma_z(1|\Omega_t).$$

The property is symmetric: x and z can be exchanged in the definition: instantaneous causality knows no direction.

Granger causality in a MA model

Assume the representation

$$y_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

It is easily motivated that x does not cause z iff $\Phi_{12,j} = 0$ for all j.

Granger causality in a VAR

A stationary stable VAR has an MA representation, so Granger causality can be checked on that one. Alternatively, consider the partitioned VAR

$$y_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} A_{11,j} & A_{12,j} \\ A_{21,j} & A_{22,j} \end{bmatrix} \begin{bmatrix} z_{t-j} \\ x_{t-j} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

It is easily shown that x does not cause z iff $A_{12,j} = 0, j = 1, \ldots, p$ (block inverse of matrix).

Remarks on testing for Granger causality in a VAR

- The property that $\Phi_{12,j} = 0$ characterizes non-causality is not restricted to VARs: it works for any process with a Wold-type MA representation;
- The property may also be generalized to x and z being two sub-vectors of y with $M + N < K$. Some extensions, however, do not work properly for the VAR representation, just for the MA representation;
- The definition is sometimes modified to *h–step causality*, meaning x_t does not improve $z_t(j), j < h$ but does improve $z_t(h)$: complications in naive testing for the VAR form, though not for the MA form.

Characterization of instantaneous causality

Proposition

Let y_t be a VAR with nonsingular innovation variance matrix Σ_u. There is no instantaneous causality between x_t and z_t iff

$$\mathrm{E}(u_{1t}u_{2t}') = 0.$$

This condition is certainly symmetric.

Instantaneous causality and the non-unique MA representation

Consider the Cholesky factorization of $\Sigma_u = PP'$, with P lower triangular. Then, it holds that

$$y_t = \mu + \sum_{j=0}^{\infty} \Phi_j P P^{-1} u_{t-j} = \mu + \sum_{j=0}^{\infty} \Theta_j w_{t-j},$$

with $\Theta_j = \Phi_j P$ and $w = P^{-1}u$ and

$$\Sigma_w = P^{-1}\Sigma_u (P^{-1})' = I_K.$$

In this form, instantaneous causality corresponds to $\Theta_{21,0} \neq 0$, which looks asymmetric. An analogous form and condition is achieved by exchanging x and z.

Impulse response analysis: the idea

The researcher wishes to add detail to the Granger-causality analysis and to quantify the effect of an *impulse* in a component variable $y_{j,t}$ on another component variable $y_{k,t}$.

The 'derivative' $\partial y_{k,t+h}/\partial y_{j,t}$ cannot be determined from the VAR model. The 'derivative'

$$\frac{\partial y_{k,t+h}}{\partial u_{j,t}}$$

corresponds to the (k, j) entry in the matrix Φ_h of the MA representation. It is not uniquely determined. The matrix of graphs of $\Phi_{kj,h}$ versus h is called the *impulse response function* (IRF).

Impulse response analysis: general properties

- If y_j does not Granger-cause y_k, the corresponding impulse response in (k, j) is constant zero;
- If the first $p(K-1)$ values of an impulse response are 0, then all values are 0;
- If the VAR is stable, all impulse response functions must converge to 0 as $h \to \infty$;
- It is customary to scale the impulse responses by the standard deviation of the response variable $\sqrt{\sigma_{kk}}$;
- The impulse response based on the canonical MA representation $\Phi_h, h \in \mathbb{N}$, ignores the correlation across the components of u in Σ_u and may not correspond to the true reaction.

Orthogonal impulse response

Re-consider the alternative MA representation based on

$$\Sigma_u = PP', \quad \Theta_j = \Phi_j P, \quad w = P^{-1}u,$$

that is,

$$y_t = \mu + \sum_{j=0}^{\infty} \Theta_j w_{t-j}.$$

Because of $\Sigma_w = I_K$, 'shocks' are orthogonal. Note that w_j is a linear function of $u_k, k \leq j$. The resulting matrix of graphs $\Theta_{kj,h}$ versus h is *an orthogonal impulse response* function (OIRF).

Orthogonal impulse response: properties

- Because of the multiplication by the matrix P, diagonal entries in Θ_0 will not be ones. This problem can be remedied simply via a diagonal matrix, such that $\tilde{\Theta}_0$ has diagonal ones and $\Sigma_{\tilde{w}}$ is diagonal;
- The OIRF can be quite different from the IRF based on Φ_h. If there is no instantaneous causality, both will coincide;
- The orthogonal IRF based on a re-ordering of variable components will differ from the correspondingly re-ordered OIRF. Additional to the permutations, a continuum of OIRF versions may be considered.

Ways out of the arbitrariness dilemma

- Some researchers suggest to arrange the vector of components such that the *a priori* 'most exogenous' variable appears first etc.;
- The *generalized impulse response function* (GIRF) according to Pesaran summarizes the OIRF for each response variable suffering the maximum response (coming last in the vector). It is not an internally consistent IRF;
- So-called *structural* VARs attempt to identify the 'shocks' from economic theory. They often use an additional matrix A_0 that permits an immediate reaction of a component y_k to another y_j and various identification restrictions. They may also be 'over-identified' and restrict the basic VAR(p) model.

Decomposition of the h–step error

Starting from an orthogonal MA representation with $\Sigma_w = I_K$,

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

the error of an h–step forecast is

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i},$$

and for the j–th component

$$y_{j,t+h} - y_{j,t}(h) = \sum_{i=0}^{h-1} \sum_{k=1}^{K} \theta_{jk,i} w_{k,t+i}.$$

All hK terms are orthogonal, and this error can be decomposed into the K contributions from the component errors.

Forecast error variance decomposition

Consider the variance of the j–th forecast component

$$\mathrm{MSE}(y_{j,t}(h)) = \sum_{i=0}^{h-1} \sum_{k=1}^{K} \theta_{jk,i}^2.$$

The share that is due to the k–th component error,

$$\omega_{jk,h} = \frac{\sum_{i=0}^{h-1} \theta_{jk,i}^2}{\mathrm{MSE}(y_{j,t}(h))},$$

defines the *forecast error variance decomposition* (FEVD) and is often tabulated or plotted versus h for $j, k = 1, \ldots, K$.

Invariants and others in structural analysis

1. *Granger causality* is independent of the choice of Wold-type MA representation. It is there or it is not;
2. *Impulse response functions* depend on the chosen representation. OIRF may differ for distinct orderings of the component variables;
3. *Forecast error variance decomposition* inherits the problems of IRF analysis: unique only in the absence of instantaneous causality.