Vector Autoregressions (VAR)

Ercan Karadas

New York University
Department of Economics

Spring, 2018

Content

- **Cross-correlations**
- **VAR model in standard/reduced form**
- **Properties of VAR(1), VAR(p)**
- **Structural VAR, identification**
- **Estimation, model specification, forecasting**
- **Impulse responses**

Resources

- Tsay, Chapter 8.1-3
- Christopher A. Sims(1980), Macroeconomics and Reality, Econometrica 48, 1980.

Notation

Here we use the multiple equations notation. Vectors are of the form:
$\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt})'$ a cloumn vector of length k.
$\boldsymbol{y}_t$ contains the values of k variables at time t.
k ... the number of variables (i.g. number of equations)
T ... the number of observations

We do not distinguish in notation between data and random variables.

Motivation

Multivariate data

From an empirical/atheoretical point of view observed time series movements are often related with each another. E.g.

- GDP growth and unemployment rate show an inverse pattern,
- oil prices might be a leading indicator for other energy prices, which on the other hand have an effect on oil.

Weak stationarity: multivariate

We assume our series $\boldsymbol{y}_t$ are **weakly stationary**. I.e. for $k = 2$

$$\mathrm{E}(\boldsymbol{y}_t) = \begin{bmatrix} \mathrm{E}(y_{1t}) \\ \mathrm{E}(y_{2t}) \end{bmatrix} = \boldsymbol{\mu}$$

and $\quad \mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t-\ell}) = \mathrm{E}(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_{t-\ell} - \boldsymbol{\mu})'$

$$\mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t-\ell}) = \begin{bmatrix} \mathrm{Cov}(y_{1t}, y_{1,t-\ell}) & \mathrm{Cov}(y_{1t}, y_{2,t-\ell}) \\ \mathrm{Cov}(y_{2t}, y_{1,t-\ell}) & \mathrm{Cov}(y_{2t}, y_{2,t-\ell}) \end{bmatrix} = \boldsymbol{\Gamma}_\ell$$

are *independent of time t*. $\quad \ell = \ldots, -1, 0, 1, \ldots$.

- $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$.
- $\boldsymbol{\Gamma}_0$ the contemporaneous/concurrent covariance matrix
- $\boldsymbol{\Gamma}_0 = [\gamma_{ij}(0)]$ is a $(k \times k)$ matrix.
- $\boldsymbol{\Gamma}_\ell$ the **cross-covariance matrix** of order ℓ, $\boldsymbol{\Gamma}_\ell = [\gamma_{ij}(\ell)]$

Cross-correlations

Cross-correlation matrices, CCMs

Let $\boldsymbol{D} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$ with $\sigma_i = \sqrt{\gamma_{ii}(0)}$, the diagonal matrix of the standard deviations of y_i's. Then

$$\boldsymbol{\rho}_\ell = \boldsymbol{D}^{-1}\Gamma_\ell \boldsymbol{D}^{-1}$$

with $\boldsymbol{\rho}_\ell = [\rho_{ij}(\ell)]$, is the **cross-correlation matrix** of order ℓ.

$$\rho_{ij}(\ell) = \frac{\gamma_{ij}(\ell)}{\sqrt{\gamma_{ii}(0)}\sqrt{\gamma_{jj}(0)}} = \frac{\mathrm{Cov}(y_{it}, y_{j,t-\ell})}{\sigma_i \, \sigma_j}$$

$\rho_{ij}(\ell)$ is the correlation coefficient between y_{it} and $y_{j,t-\ell}$.

Cross-correlation matrices: properties

We say with $\ell > 0$, if with $\gamma_{ij}(\ell) = \mathrm{Cov}(y_{it}, y_{j,t-\ell})$

$\rho_{ij}(\ell) \neq 0$: y_j *leads* y_i.

$\rho_{ji}(\ell) \neq 0$: y_i *leads* y_j.

$\rho_{ii}(\ell)$ is the autocorrelation coefficient of order ℓ of variable y_i.

Properties of $\boldsymbol{\Gamma}_\ell$ and $\boldsymbol{\rho}_\ell$:

- $\boldsymbol{\Gamma}_0$, the covariance matrix, is symmetric (and pos def).
- $\boldsymbol{\Gamma}_\ell$ with $\ell \neq 0$, is i.g. *not* symmetric.
  $\rho_{ij}(\ell) \neq \rho_{ji}(\ell)$ in general
- $\gamma_{ij}(\ell) = \gamma_{ji}(-\ell)$ and so $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Gamma}'_{(-\ell)}$:
  $\mathrm{Cov}(y_{it}, y_{j,t-\ell}) = \mathrm{Cov}(y_{j,t-\ell}, y_{it}) = \mathrm{Cov}(y_{jt}, y_{i,t+\ell}) = \mathrm{Cov}(y_{jt}, y_{i,t-(-\ell)})$

Therefore we consider only CCMs with $\ell \geq 0$.

Types of relationships

We distinguish several types of relationships between 2 series i and j ($i \neq j$):

- no linear relationship: $\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0 \quad$ for all $\ell \geq 0$
- concurrently correlated: $\rho_{ij}(0) \neq 0$
  Some cross-correlations of higher order may be $\neq 0$.
- **uncoupled**: $\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0 \quad$ for all $\ell > 0$.
  There is no lead-lag but possibly a concurrent relationship.
- **unidirectional** relationship (wrt time) from i to j:
  $\rho_{ij}(\ell) = 0$ for all $\ell > 0$ and $\rho_{ji}(\nu) \neq 0$ for some $\nu > 0$
- **feedback** relationship (wrt time):
  $\rho_{ij}(\ell) \neq 0$ for some $\ell > 0$ and $\rho_{ji}(\nu) \neq 0$ for some $\nu > 0$

Sample cross-correlations, properties

The sample values are straight forward as in the univariate case.

$$\hat{\gamma}_{ij}(\ell) = \frac{1}{T}\sum_{t=\ell+1}^{T}(y_{it} - \bar{y}_i)(y_{j,t-\ell} - \bar{y}_j)$$

- The estimates of γ's and ρ's are consistent (but biased in small samples).
- Under the hypothesis of multivariate white noise, the cross-correlations may be tested individually with the $(\pm 1.96/\sqrt{T})$-rule.
- The distribution in small samples may deviate considerably from the expected, e.g. for stock returns due to heteroscedasticity and fat tails. Then *bootstrap resampling methods* are recommended.

Cross-correlations and autocorrelated y_i's

- Cross-correlations can be interpreted in a nice way only, if *at least one* of both series is *white noise*.
- Otherwise the autocorrelation structure of one series may interact with that of the other and 'spurious' cross effects may be observed.
- One way to resolve this problem is to use a **prewhitening** technique.
  - Find a univariate model for one of the two series.
  - Apply this *estimated* model also to the other.
  - Interpret the cross-correlations between the residuals of the 1st model and the 'residuals' of the 2nd model, instead.

Multivariate portmanteau test

The univariate **Ljung-Box statistic** $Q(m)$ has been generalized to $(k \times k)$-dimensional CCMs. The null hypothesis

$$\mathrm{H}_0: \quad \boldsymbol{\rho}_1 = \ldots = \boldsymbol{\rho}_m = \mathbf{0}$$

is tested against the alternative

$$\mathrm{H}_A: \quad \boldsymbol{\rho}_i \neq \mathbf{0} \quad \text{for some } \; 1 \leq i \leq m$$

with

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} \operatorname{tr}(\hat{\boldsymbol{\Gamma}}_\ell' \hat{\boldsymbol{\Gamma}}_0^{-1} \hat{\boldsymbol{\Gamma}}_\ell \hat{\boldsymbol{\Gamma}}_0^{-1}) \;\sim\; \chi^2(k^2 m)$$

which is χ^2 distributed with $(k^2 m)$ degrees of freedom.

Multivariate portmanteau test

This test is suitable for the case where the interdependence of essentially (univariate) white noise returns is in question.

If only a single series is not white noise, i.e. $\rho_{ii}(\ell) \neq 0$ for some $\ell > 0$, the test will reject as the univariate tests will do.

The VAR model

The VAR in standard form

A model taking into account/approximating multivariate dynamic relationships is the VAR(p), **vector autoregression** of order p.

$$\boldsymbol{y}_t = \phi_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \epsilon_t$$

- $\boldsymbol{y}_t$ is a vector of length k. There are k equations.
- p is the order of the VAR.
- $\{\epsilon_t\}$ is a sequence of serially uncorrelated random vectors with concurrent full rank covariance matrix $\boldsymbol{\Sigma}$ (not diagonal i.g.). They are coupled. ($\boldsymbol{\Sigma} = \boldsymbol{\Gamma}_0^{(\epsilon)}$, $\boldsymbol{\Gamma}_\ell^{(\epsilon)} = \mathbf{0}$, $\ell \neq 0$)
- ϕ_0 is a $(k \times 1)$ vector of constants.
- $\boldsymbol{\Phi}$'s are $(k \times k)$ coefficient matrices.

The model is called **VARX**, if additional explanatories are included.

Example of VAR(1), $k = 2$

$$\boldsymbol{y}_t = \phi_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t$$

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

or equation by equation

$$\begin{aligned} y_{1t} &= \phi_1^{(0)} + \phi_{11}^{(1)} y_{1,t-1} + \phi_{12}^{(1)} y_{2,t-1} + \epsilon_{1t} \\ y_{2t} &= \phi_2^{(0)} + \phi_{21}^{(1)} y_{1,t-1} + \phi_{22}^{(1)} y_{2,t-1} + \epsilon_{2t} \end{aligned}$$

The *concurrent relationship* between y_1 and y_2 is measured by the off-diagonal elements of $\boldsymbol{\Sigma}$.

If $\sigma_{12} = 0$ there is no concurrent relationship.

$$\sigma_{12} = \mathrm{Cov}(\epsilon_{1t}, \epsilon_{2t}) = \mathrm{Cov}(y_{1t}(\epsilon_{1t}), y_{2t}(\epsilon_{2t}) | \boldsymbol{y}_{t-1})$$

Example of VAR(1), $k = 2$

$\mathbf{\Phi}_1$ measures the dynamic dependence in $\boldsymbol{y}$.

ϕ_{12}:
ϕ_{12} measures the linear dependence of y_{1t} on $y_{2,t-1}$ in the presence of $y_{1,t-1}$.
If $\phi_{12} = 0$, y_{1t} does not depend on $y_{2,t-1}$. Then, y_{1t} depends only on its own past.

Analogous for equation 2 and ϕ_{21}.

- If $\phi_{12} = 0$ and $\phi_{21} \neq 0$:
  There is a unidirectional relationship from y_1 to y_2.
- If $\phi_{12} = 0$ and $\phi_{21} = 0$:
  y_1 and y_2 are coupled (in the sense that possibly $\sigma_{12} \neq 0$).
- If $\phi_{12} \neq 0$ and $\phi_{21} \neq 0$:
  There is a feedback relationship between both series.

Properties of the VAR

Properties of the VAR(1)

We investigate the statistical properties of a VAR(1).

$$\boldsymbol{y}_t = \phi_0 + \boldsymbol{\Phi} \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t$$

Taking expectations we get $\quad \boldsymbol{\mu} = \mathrm{E}(\boldsymbol{y}_t) = \phi_0 + \boldsymbol{\Phi}\mathrm{E}(\boldsymbol{y}_{t-1}) \quad$ or

$$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{y}_t) = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1} \phi_0$$

if $(\boldsymbol{I} - \boldsymbol{\Phi})^{-1}$ exists.

We demean the series as in the univariate case and denote $\quad \tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{\mu}$

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{\Phi} \tilde{\boldsymbol{y}}_{t-1} + \boldsymbol{\epsilon}_t$$

Repeated substitutions gives the MA(∞) representation

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{\epsilon}_t + \boldsymbol{\Phi}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Phi}^2\boldsymbol{\epsilon}_{t-2} + \boldsymbol{\Phi}^3\boldsymbol{\epsilon}_{t-3} + \ldots$$

Properties of the VAR(1)

We see that $\boldsymbol{y}_{t-\ell}$ is predetermined wrt ϵ_t.

- $\mathrm{Cov}(\boldsymbol{\epsilon}_t, \boldsymbol{y}_{t-\ell}) = \mathbf{0}$, $\ell > 0$. $\boldsymbol{\epsilon}_t$ is uncorrelated with all past y's.
- ϵ_t may be interpreted as a shock or innovation at time t.
  It has possibly an effect on the future but not on the past.
- $\mathrm{Cov}(\boldsymbol{y}_t, \epsilon_t) = \boldsymbol{\Sigma}$. (Multiply by ϵ_t' and take expectations.)
- $\boldsymbol{y}_t$ depends on the innovation ϵ_{t-j} with the coeff matrix $\boldsymbol{\Phi}^j$.
  This implies that $\boldsymbol{\Phi}^j$ has to vanish for stationarity when $j \to \infty$.
- Thus, the eigenvalues of $\boldsymbol{\Phi}$ have to be smaller 1 in modulus. This is necessary and sufficient for weak stationarity.
  Then also $(\boldsymbol{I} - \boldsymbol{\Phi})^{-1}$ exists.
  (Cp. AR(1) process: $y_t = \alpha y_{t-1} + \epsilon_t$ with $|\alpha| < 1$.)
- $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}, \quad \ell > 0$.

Many properties generalize from the univariate AR to the VAR(p).

Properties, weakly stationary processes

- We define the (matrix) polynomial (for a VAR(p))

$$\mathbf{\Phi}(\mathrm{L}) = \boldsymbol{I} - \mathbf{\Phi}_1\mathrm{L} - \ldots - \mathbf{\Phi}_p\mathrm{L}^p$$

For stability it is required that the roots of the **characteristic equation**

$$|\mathbf{\Phi}(z)| = 0, \qquad |z| > 1$$

are outside the unit circle. E.g. VAR(1): $\mathbf{\Phi}(\mathrm{L}) = \boldsymbol{I} - \mathbf{\Phi}\,\mathrm{L}$.

The VAR in standard form is well defined and can be used to approximate any weakly stationary process arbitrarily well by choosing a suitable order p.

Representation of a VAR(p) as a VAR(1)

$$\text{VAR}(p): \quad \boldsymbol{y}_t = \phi_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \epsilon_t$$

Every VAR(p) can be written as a ($k\,p$)-dimensional VAR(1),

$$\boldsymbol{x}_t = \boldsymbol{\Phi}^* \boldsymbol{x}_{t-1} + \xi_t$$

$\boldsymbol{x}_t = (\tilde{\boldsymbol{y}}'_{t-p+1}, \ldots, \tilde{\boldsymbol{y}}'_t)'$ and $\xi_t = (0, \ldots, 0, \epsilon'_t)'$ both ($k\,p \times 1$).

$$\boldsymbol{\Phi}^* = \begin{bmatrix} \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{I} \\ \boldsymbol{\Phi}_p & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_{p-2} & \ldots & \boldsymbol{\Phi}_1 \end{bmatrix}$$

The ($k\,p \times k\,p$) matrix $\boldsymbol{\Phi}^*$ is called **companion matrix** to the matrix polynomial $\boldsymbol{\Phi}(\mathrm{L}) = \boldsymbol{I} - \boldsymbol{\Phi}_1 \mathrm{L} - \ldots - \boldsymbol{\Phi}_p \mathrm{L}^p$.

Representation of a VAR(p) as a VAR(1)

The last component of $\boldsymbol{x}_t$ is the mean corrected $\boldsymbol{y}_t$, $\tilde{\boldsymbol{y}}_t$.

The last row of Φ^* is essentially the VAR(p) recursion in reverse order.

Structural VAR and identification

Structural VAR

The VAR in standard form is also called VAR in **reduced form**, as it does not contain the concurrent relationships in $\boldsymbol{y}$ explicitly. A VAR in **structural form** is

$$\boldsymbol{\Theta} \boldsymbol{y}_t = \theta_0 + \boldsymbol{\Theta}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Theta}_p \boldsymbol{y}_{t-p} + \boldsymbol{\eta}_t$$

$\boldsymbol{\Theta}$ is the coefficient matrix of the y_t's. Its diagonal elements are all 1.
$\boldsymbol{\eta}_t$ is serially uncorrelated as $\boldsymbol{\epsilon}_t$, but its concurrent covariance matrix, $\boldsymbol{\Omega}$, is diagonal. (The concurrent eff are captured in a possibly non diagonal $\boldsymbol{\Theta}$.)

If $\boldsymbol{\Theta}$ is invertible, multiplication with $\boldsymbol{\Theta}^{-1}$ yields the VAR in standard form with

$$\boldsymbol{\Phi}_j = \boldsymbol{\Theta}^{-1} \boldsymbol{\Theta}_j$$

$$\boldsymbol{\epsilon}_t = \boldsymbol{\Theta}^{-1} \boldsymbol{\eta}_t \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1} \boldsymbol{\Omega} (\boldsymbol{\Theta}^{-1})'$$

Identification

Comparing the number of coefficients including the error covariances, we find

- Standard form: $0 + k + p\,k^2 + [(k^2 - k)/2 + k]$
- Structural form: $(k^2 - k) + k + p\,k^2 + k$

The number of parameters of a generally specified structural model is always (except $k = 1$) larger than that of the reduced form.

Identification

- There is always (at least) one structural form corresponding to a standard form (e.g. via the Cholesky decomposition of $\mathbf{\Sigma}$, see e.g. Tsay p.350)
- However, the representation in a structural form is not unique without putting the required number of restrictions on the parameters in the $\mathbf{\Theta}$ matrices. From the point of view of the structural form this is called the **identification problem**.
- A 'disadvantage' of the structural framework is that the concurrent relationships of the *y*'s can be interpreted only together with some economic theory.
  Sims'(1980) conjecture was that the dynamic relationships can be interpreted well, even without economic theory. Unfortunately, this does not hold.
  (See the discussion of orthogonalized shocks η_t below.)

Estimation of a VAR in standard form

We assume the true relationship is a VAR(p), the model is correctly specified, and the error is normal or close to normal.

- LS: If all variables are included in each equation (no restrictions on the Φ_j-parameters, $j > 0$) simple single equation LS is consistent and asy normal. Clearly, single equation LS cannot be efficient, since the cross-covariances in Σ are ignored.
- GLS: Using the estimated covariances of ϵ_t improves efficiency.

Interpretation of the t-statistics of the single equation estimates is straight forward.

Curse of dimensionality

A VAR(p) in standard form has

$$p\,k^2$$

parameters not counting the constant terms and the error variance-covariances.

The number of the parameters is

- linear in the order of the VAR, and
- increases quadratically with the dimension k.

The number of observations ($k\,T$) increases only linearly with k.

Remark: Emperically, it is difficult to interpret e.g. a significant $\hat{\Phi}^{(9)}_{7,23}$.

Model selection

Like in the univariate ARMA modeling, single coefficients in Φ_j, $j > 0$, are not set to zero a priori.

The model selection procedure starts with a maximal plausible order p_{max}. All models with $p = 0, 1, \ldots, p_{max}$ are estimated. The models can be ranked according to an information criterion.

The (conditional) ML estimate of the concurrent error covariance matrix of the model with order p is

$$\hat{\Sigma}(p) = \frac{1}{T} \sum_{t=p+1}^{T} \hat{\epsilon}_t(p)\, \hat{\epsilon}'_t(p)$$

The model selection criteria are $[\log(|\hat{\Sigma}(p)|) \;\approx\; -(2/T)\, \ell\ell(p)]$

$$AIC(p) = \log(|\hat{\Sigma}(p)|) + 2\, p\, k^2/T$$

$$SBC(p) = \log(|\hat{\Sigma}(p)|) + \log(T)\, p\, k^2/T$$

Model validity

As in the univariate case the residuals of the chosen model should be (multivariate) white noise. Here the $\boldsymbol{\rho}_\ell^{(\epsilon)}$ for $\ell > 0$ should be zero.

The multivariate Ljung-Box test may be applied with number of freedoms as the number of cross-correlation coefficients under test reduced for the number of estimated parameters:

$$k^2(m-p)$$

Forecasting

Forecasting

Existing concurrent structural relationships between the endogenous variables are demanding to interpret. However, due to the unique representation of the reduced/standard form the model is suitable for forecasting.

The **1-step ahead forecast** (conditional on data up to T) is

$$\boldsymbol{y}_T(1) = \phi_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_T + \ldots + \boldsymbol{\Phi}_p \boldsymbol{y}_{T-p+1}$$

The 1-step ahead **forecast error**, $\boldsymbol{e}_T(1) = \boldsymbol{y}_{T+1} - \boldsymbol{y}_T(1)$, is

$$\boldsymbol{e}_T(1) = \epsilon_{T+1}$$

Forecasting

The **2-step ahead forecast** is

$$\boldsymbol{y}_T(2) = \phi_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_T(1) + \boldsymbol{\Phi}_2 \boldsymbol{y}_T + \ldots + \boldsymbol{\Phi}_p \boldsymbol{y}_{T-p+2}$$

The 2-step ahead forecast error is

$$\boldsymbol{e}_T(2) = \boldsymbol{y}_{T+2} - \boldsymbol{y}_T(2) = \epsilon_{T+2} + \boldsymbol{\Phi}_1[\boldsymbol{y}_{T+1} - \boldsymbol{y}_T(1)] = \epsilon_{T+2} + \boldsymbol{\Phi}_1 \epsilon_{T+1}$$

The ℓ-step ahead forecast $\boldsymbol{y}_T(\ell)$ converges to the mean vector $\boldsymbol{\mu}$ as the forecast horizon ℓ increases.

The covariance matrices are

- for $\boldsymbol{e}_T(1)$ $\quad \boldsymbol{\Sigma}$,
- for $\boldsymbol{e}_T(2)$ $\quad \boldsymbol{\Sigma} + \boldsymbol{\Phi}_1 \boldsymbol{\Sigma} \boldsymbol{\Phi}_1'$.

The covariance matrix of the ℓ-step ahead forecast error converges to the covariance matrix of $\boldsymbol{y}_t$ as ℓ increases.

Impulse responses

Impulse responses

Analogous to the Wold representation of univariate weakly stationary process the VAR(p) can be written in terms of concurrent and past innovations.

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\epsilon}_{t-2} + \dots$$

This moving average representation makes the impact of an innovation $\boldsymbol{\epsilon}_{t-i}$ at $(t-i)$ on $\boldsymbol{y}_t$ at t explicit. It is $\boldsymbol{\Psi}_i$.

This impact is the same as of $\boldsymbol{\epsilon}_t$ on $\boldsymbol{y}_{t+i}$.

$\{\boldsymbol{\Psi}_i\}$ is called **impulse response function**, IRF.

Impulse responses

Since the components of ϵ_t are correlated the separating out of the effects of single innovations is difficult. Thus we consider

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\eta}_t + \boldsymbol{\Psi}_1^* \boldsymbol{\eta}_{t-1} + \boldsymbol{\Psi}_2^* \boldsymbol{\eta}_{t-2} + \ldots$$

where the components of $\boldsymbol{\eta}_t$ are uncorrelated.

$$\boldsymbol{\Psi}_j^* \boldsymbol{\eta}_{t-j} = [\boldsymbol{\Psi}_j \boldsymbol{L}] \boldsymbol{\eta}_{t-j} = \boldsymbol{\Psi}_j [\boldsymbol{L} \boldsymbol{\eta}_{t-j}] = \boldsymbol{\Psi}_j \boldsymbol{\epsilon}_{t-j}$$

$\boldsymbol{L}$ is the Cholesky factor of $\boldsymbol{\Sigma}$: $\quad \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}'$ $\qquad$ and $\qquad \boldsymbol{\epsilon}_t = \boldsymbol{L}\boldsymbol{\eta}_t$

The diagonal matrix $\boldsymbol{D}$ is the covariance matrix of $\boldsymbol{\eta}_t$. $\mathrm{Cov}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_t) = \boldsymbol{D}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}$.
Then the covariances of $\boldsymbol{L}\boldsymbol{\eta}_{t-j}$ are those of $\boldsymbol{\epsilon}_{t-j}$.

$$\mathrm{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = \mathrm{E}(\boldsymbol{L}\boldsymbol{\eta}_t)(\boldsymbol{L}\boldsymbol{\eta}_t)' = \mathrm{E}(\boldsymbol{L}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t')\boldsymbol{L}') = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}' = \boldsymbol{\Sigma}$$

Impulse responses

Element $\psi^*_{ij}(\ell)$ of $\mathbf{\Psi}^*_\ell$ is the impact of innovation $\eta_{j,t}$ of size 1 on $y_{i,t+\ell}$.

In practice instead of size 1, the standard deviation of the shock is used. Then the diagonal elements of $\boldsymbol{D}$ are transformed to 1:

$$\boldsymbol{LDL}' = \tilde{\boldsymbol{L}}\boldsymbol{I}\tilde{\boldsymbol{L}}'$$

and $\tilde{\mathbf{\Psi}} = \mathbf{\Psi}\tilde{\boldsymbol{L}}$ is used instead of $\mathbf{\Psi}^*$.

Transformation from the representation of $\boldsymbol{y}$ in $\boldsymbol{\epsilon}$ to the representation in $\boldsymbol{\eta}$ is a trial to infer from the errors of the reduced form to the structural errors. However, they are not unique in general.
The impulse responses *will change* when the Cholesky algorithm does not start with element (1,1) of $\mathbf{\Sigma}$ - as is commonly done - but e.g. with (k, k).