Macroecon. - 0

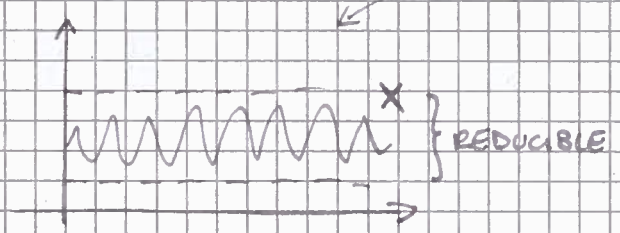
- Stochastic Process : $\{X_t\}_{t=0}^{\infty}$
  - Realization of Stoc. process : $\{X_t(\omega)\}_{t=-\infty}^{\infty}$ → event realized.

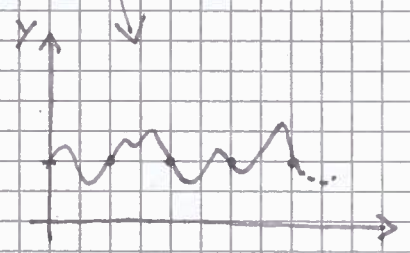
  Recall the issue:

  ~ goal is to estimate the distribution or the moments of any given X_t. Usually we can use the law of large numbers to estimate the moments $(\frac{1}{N}\sum^{N} X_i)$ under the assumption of a large number of iid observations. In TS we don't have iid since observations are taken in different time steps.
  Still, we introduce two concepts that will allow us to estimate parameters of the TS distribution.

- Stationarity : - $E(X_t) = \mu$ & $Cov(X_t, X_{t+h}) = \gamma_h$ WEAK
  - $f(X_t, X_{t+h}) = f(X_{t+k}, X_{t+k+h})$ STRONG

- Ergodicity ⇝ i.e "asymptotically independent"
  - Equivalent to : Irreducible & Aperiodic





~ i.e the TS is not contained in a closed set.

~ i.e Returns to the same y with a costant frequency ⇝ periodically.

- White Noise : $E(\varepsilon_t) = 0$ $V(\varepsilon_t) = \sigma^2$ $Cov(\varepsilon_t, \varepsilon_{t+h}) = 0$ $\forall h \neq 0$
  ↳ OR Strong version ⇝ Series of iid variables.

- ARMA Model ⇝ use ϕ_i for the AR and θ_i for the MA

- LAG Operator ⇝ linear operator ⇝ i.e self-map, it has all the properties :
  - $L X_t = X_{t-1}$
    $L^2 X_t = L L X_t = X_{t-2}$
    $L^i X_t = X_{t-i}$
    $L^{-i} X_t = X_{t+i}$
    $L(X_t + y_t) = X_{t-1} + y_{t-1}$
    $L \beta X_t = \beta L X_t = \beta X_{t-1}$
  - ARMA becomes:
    $(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 \cdots^{\infty}) X_t =$
    $= (1 + \theta_1 L + \theta_2 L^2 + \cdots^{\infty}) \varepsilon_t$
    OR
    $\phi(L) X_t = \theta(L) \varepsilon_t$

- Usefull to remember:

$\phi(L) = (1 + \phi_1 L + \phi_2 L^2 + \phi_3 L^3 + \dots)$

$\theta(L) = //$

Macroecon - 1

- **ARMA are not unique**

i.e The same process can be represented by different ARMA process

e.g let $X_t = \phi X_{t-1} + \varepsilon_t$ AR(1) OR ARMA(1, 0)

notice $X_t = \phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$

$X_t = \phi^2 X_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \rightsquigarrow$ ARMA(2, 1)

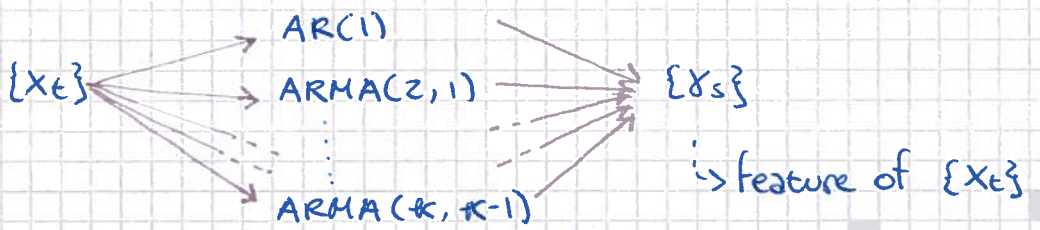
$\vdots$

$X_t = \phi^k X_{t-k} + \dots \rightsquigarrow$ ARMA(k, $k-1$)

- **The Autocovariance function is unique**

i.e any Autcov. funct. $\{\gamma_s\}$ has only one Stoc. process $\{X_t\}$ and viceversa

idea: (ex.)



- **Solving an AR(1)** (2 equivalent ways)

- Consider: $X_t = \phi X_{t-1} + \varepsilon_t$ $\rightsquigarrow$ by "solving" we mean deriving the form where X_t does not depend on its past values, but depends uniqly on the "forcing term" ε

as just seen

- $X_t = \phi^k X_{t-k} + \varepsilon_t + \phi \varepsilon_{t-1} + \dots + \phi^k \varepsilon_{t-k}$
- let $k \to +\infty$
- Assume $|\phi| < 1$ s.t $\phi^k \to 0$
- then: $\boxed{X_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}}$

Q: Can we solve for X_t using the Lag Operator?

notice: $X_t = \phi X_{t-1} + \varepsilon_t \Rightarrow$

$\Rightarrow X_t - \phi L X_t = \varepsilon_t$

$(1 - \phi L) X_t = \varepsilon_t$

[Recall that: $1/(1-z) = \sum_{i=0}^{\infty} z^i$ (Geometric series)

↳ IFF $|z| < 1$]

$\Rightarrow$ $(1-\phi L)$ is invertible IFF $|\phi| < 1$ (same condition)

then: $X_t = (1-\phi L)^{-1} \cdot \varepsilon_t$

$$\Rightarrow \quad X_t = \sum_{i=0}^{\infty} \phi^i \cdot L^i \cdot \varepsilon_t$$

Combining the two ways: $\sum_{i=0}^{\infty} \phi^i \cdot \varepsilon_{t-i} = \sum \phi^i L^i \cdot \varepsilon_t$ ✓

---

* Special case for $|\phi| > 1$ ⇝ how do we solve?

idea: instead of replacing the lag term with past values, replace with future values.

i.e $X_t = \phi X_{t-1} + \varepsilon_t \Rightarrow X_{t-1} = \frac{1}{\phi} X_t - \frac{1}{\phi} \varepsilon_t$

$$\Rightarrow X_{t-1} = \frac{1}{\phi}\left(\frac{1}{\phi} X_{t+1} - \frac{1}{\phi}\varepsilon_{t+1}\right) - \frac{1}{\phi}\varepsilon_t$$

$\Rightarrow$ k repeat. ⇝ $\left(\frac{1}{\phi}\right)^k X_{t+k} - \dots$

notice $\left|\frac{1}{\phi}\right| < 1 \Rightarrow \left(\frac{1}{\phi}\right)^k \to 0$ as $k \to \infty$

$$\Rightarrow X_t = \sum_{i=0}^{\infty} -\left(\frac{1}{\phi}\right)^i \cdot \varepsilon_{t+i}$$

This is called a "Non-Causal ARMA" ✓

- Solving for an AR(p) (From AR(p) to MA(∞))

idea: We'll see how to solve an AR(2), the procedure is the same for any value of p.

AR(2): $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$

write it in matrix form: (OR Companion form)

$$\text{(eq.1)} \quad \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \underbrace{\begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}}_{A} \cdot \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} = \begin{cases} X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} \\ X_{t-1} = X_{t-1} \end{cases}$$

- Recall, any square matrix can be rewritten using its Eigenvalues and Eigenvectors ⇝ Spectral decomposition.

$$\Rightarrow \quad A = \underset{2\times2}{B} \; \underset{2\times2}{\Lambda} \; \underset{2\times2}{B^{-1}}$$

where: - B is the matrix composed by A's Eigenvectors (i.e B columns are A's Eigenvectors)

- $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ → A's Eigenvalues

Now, premultiply eq.1 by B^{-1} to obtain:

$$\underbrace{B^{-1}\begin{bmatrix} X_t \\ X_{t-1}\end{bmatrix}}_{\tilde{X}_t} = \underbrace{B^{-1}\cdot B}_{1}\cdot \Lambda \cdot \underbrace{B^{-1}\cdot\begin{bmatrix} X_{t-1} \\ X_{t-2}\end{bmatrix}}_{\tilde{X}_{t-1}} + \underbrace{B^{-1}\cdot\begin{bmatrix} \varepsilon_t \\ 0\end{bmatrix}}_{\tilde{\varepsilon}_t}$$

equivalent to the system:

$$\begin{cases} \tilde{X}_{1t} = \lambda_1 \tilde{X}_{1,t-1} + \tilde{\varepsilon}_{1,t} & \rightsquigarrow AR(1) \\ \tilde{X}_{2t} = \lambda_2 \tilde{X}_{2,t-1} + \tilde{\varepsilon}_{2,t} & \rightsquigarrow AR(1) \end{cases}$$

Assuming that: $|\lambda_1| < 1 \Rightarrow \tilde{X}_{1t} = \sum_{i=0}^{\infty} \lambda_1^i \tilde{\varepsilon}_{t-i}$

$|\lambda_2| < 1 \Rightarrow \tilde{X}_{2t} = \sum_{i=0} \lambda_2^i \cdot \tilde{\varepsilon}_{2t-i}$

↳ using the previous result for AR(1)

To go from $\tilde{X}_t$ to X_t recall that:

$$\tilde{X}_t = B^{-1}\cdot X_t$$

$$\Rightarrow X_t = B\cdot \tilde{X}_t$$

$$\Rightarrow X_t = \begin{bmatrix} b_{11}\tilde{X}_{1t} + b_{12}\tilde{X}_{2t} \\ b_{21}\tilde{X}_{1t} + b_{22}\tilde{X}_{2t}\end{bmatrix} \begin{matrix} = X_t \\ = X_{t-1}\end{matrix} = B\cdot\tilde{X}_t$$

See how the same logic applies to any AR(p)

~~Again, notice~~

Notice how here the existence of the solution is conditioned on the Eigenvalues of A (λ_1 & λ_2) being smaller than 1.

There are actually two equivalent ways to obtain the conditions for the existence of a solution in an ARMA

- Condition for the existence of a Solution.

  idea: two interchangeable and equivalent conditions

① Eigenvalues of (A) ⇝ Comp Form matrix (still considering an AR(2))

- Consider A from the Companion form
- The Eigenvalues of A solve for: $(A-\lambda\cdot I)\cdot b = 0$
- IF $(A-\lambda I)$ is invertible ⇒ trivial solution for $b=0$
- i.e we don't want to allow the trivial solution
- $(A-\lambda I)$ is not invertible IFF $\det(A-\lambda I) = 0$ ⇒

$$\Rightarrow \det\begin{bmatrix} \phi_1-\lambda & \phi_2 \\ 1 & -\lambda\end{bmatrix} = -\lambda(\phi_1-\lambda) - \phi_2 \Rightarrow \lambda^2 - \lambda\phi_1 - \phi_2 = 0$$

(eq. 1)

② Characteristic Equation

- Recall: AR(2) $\rightarrow \underbrace{(1-\phi_1 L-\phi_2 L^2)}_{\text{"roots"}} \cdot X_t = \varepsilon_t$

- substitute L with any variable z; and set it equal to 0.
  (eq.2) $\quad 1 - z\phi_1 - z^2\phi_2 = 0 \quad \rightsquigarrow$ Charct. equation
- solve for z and check that the roots, in this case z_1 and z_2 are s.t $|z_1| > 1$ and $|z_2| > 1$

▷ Compare eq.1 and eq.2 : (Equivalence)

eq.1	eq.2
$\lambda^2 - \lambda\phi_1 - \phi_2 = 0$	$1 - z\phi_1 - z^2\phi_2 = 0$

- eq.1 = eq.2 when $\lambda = 1/z$ → Eigenvalue

$$\Rightarrow |z| > 1 \iff |\lambda| < 1$$

∘ Finding the Coefficients

Recall: $\quad \phi(L) X_t = \theta(L) \cdot \varepsilon_t$

$$\Rightarrow \quad X_t = \frac{\theta(L)}{\phi(L)} \cdot \varepsilon_t \quad \text{for } \phi(L) \text{ invertible}$$

$$= \psi(L) \cdot \varepsilon_t$$

$$\Rightarrow \quad \theta(L) = \psi(L) \cdot \phi(L) \qquad \text{eq.3}$$

e.g Consider an ARMA(1,1) $\rightarrow (1-\phi L) X_t = (1-\theta L)\varepsilon_t$

$$\Rightarrow X_t = (\psi_0 + \psi_1 L + \psi_2 L^2 \cdots) \cdot \varepsilon_t$$

$$\Rightarrow (1+\theta L) = (\psi_0 + \psi_1 L + \psi_2 L^2 \cdots) \cdot (1-\phi L) \quad \text{by eq.3}$$

idea: we have estimated ϕ and θ through the data and we want to obtain $\psi_0 \cdots$ for the solution $X_t(\varepsilon_t)$.

The series $\psi_0, \psi_1, \ldots$ is found using the Alogorithm.:

- starting from period 0 $\rightsquigarrow L = 0$
  ↳ $\psi_0 = 1$ * → and without $L = 0$
- then period 1 $\rightsquigarrow L = 1$ (notice $L^2, L^3 \cdots = 0$)
  ↳ $\theta L = \psi_1 L - \phi L \psi_0$
  $L = 1 \rightarrow \theta = \psi_1 - \phi\psi_0 \quad \rightsquigarrow$ find by sub. $\psi_0 = 1$ *
- go on with $L^2 = 1$ (otherwise 0) and sub the previously found coefficients.

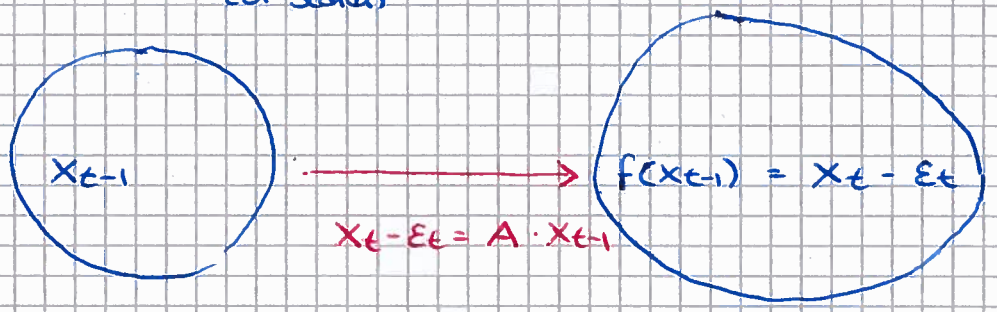
- Deriving the Autocovariance Structure
  - Assume Weak Stationarity
  - Consider MA(1) : $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$
  - $E(X_t) = E(\varepsilon_t) + \theta \cdot E(\varepsilon_{t-1}) = 0$
  - $V(X_t) = Cov(\varepsilon_t + \theta \varepsilon_{t-1} ; \varepsilon_t + \theta \varepsilon_{t-1}) = \sigma^2 + \theta^2 \cdot \sigma^2$

    σ^2 σ^2

    by $\varepsilon_t \perp \varepsilon_{t-1}$
  - $Cov(X_t, X_{t-1}) = Cov(\varepsilon_t + \theta \varepsilon_{t-1} ; \varepsilon_{t-1} + \theta \varepsilon_{t-2}) = \theta^2 \sigma^2$
  - $Cov(X_t, X_{t-2}) = Cov(\varepsilon_t + \theta \varepsilon_{t-1} ; \varepsilon_{t-2} + \theta \varepsilon_{t-3}) = 0$
  - Notice: Cov doesn't depend on t vs Stationarity

* More on Eigenvectors and Eigenvalues

idea: an AR process, in its essence, is a <u>linear map</u> connecting a vector X_{t-1} (or scalar) to its "prediction" X_t



The Matrix A is the "Transformation matrix":

- in an AR(1), A is a scalar
- in an AR(p), A is the Companion Form matrix

Eigenvectors are vectors belonging to the (X_{t-1}) space s.t $\exists \lambda$ s.t $\lambda \underline{v} = A \cdot \underline{v}$

i.e vectors that get streched or squished during the transformation, but do not rotate = remain on their span

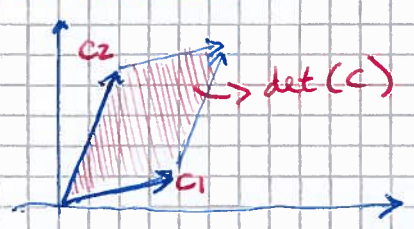
notice:
- λ is the "stretching" scalar = Eigenvalue
- Each Eigenvector has its own eigenvalue

notice: $\lambda v = A \cdot v \iff (A - I \cdot \lambda) \cdot v = 0$

- $\exists$ a trivial solution for $v = \underline{0}$
- FOR $v \neq \underline{0}$: $\exists v$ s.t $(A - I\lambda) \cdot v = 0$ IFF $\det(A - I\lambda) = 0$

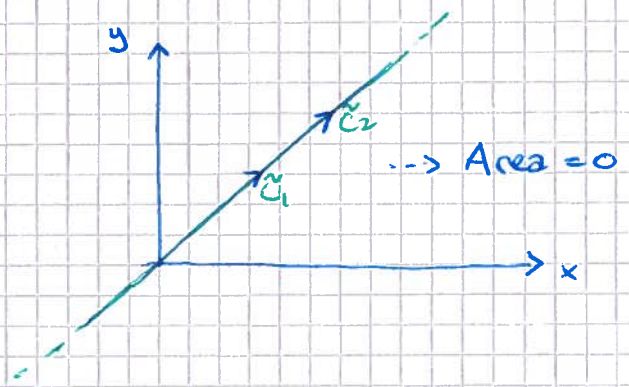
idea: In order for $(A - I \cdot \lambda) \cdot v = 0$ the matrix $A - I \cdot \lambda$ must be such that it lowers the dimension of the resulting product with the vector v.
In other words: $\det(A - I\lambda) = 0$

Recall that the $\det(C)$ = Areas formed by the vectors inside of C. let $C \in \mathbb{R}^2$



To say that $\det(C - \lambda \cdot I) = 0$ is to look for a scalar λ s.t

given $C - \lambda \cdot I = \begin{bmatrix} c_{11} - \lambda & c_{12} \\ \underbrace{c_{21}}_{\tilde{c}_1} & \underbrace{c_{22} - \lambda}_{\tilde{c}_2} \end{bmatrix}$



We already showed how an AR(p) can be inverted into its MA Form IFF the eigenvalues (λ) of the companion matrix are smaller than 1 in absolute value:

$$|\lambda| < 1$$

BUT $\exists$ one wrong solution: the Eigenvalue corresponding to the Eigenvector zero ($v = \underline{0}$) can be anything, ~~although~~ idea: we want to eliminate this solution

How?

We know that all the Eigenvalues s.t $\det(A - \lambda \cdot I) = 0$ are connected to Eigenvectors $\neq 0$

In Summary:

Thm. An AR(p) is invertible IFF TFAT:

1) $\exists \lambda \in \mathbb{R}$ s.t $\det(A - \lambda \cdot I) = 0$
2) $|\lambda| < 1$

◦ Conditional Moments for ARMA process

Consider a MA(1) : $y_t = \varepsilon_t + \theta\, \varepsilon_{t-1}$

- **Conditional Mean**

$$E(y_t \mid \underbrace{I_{t-1}}_{\text{Information up to } t-1}) = E(\varepsilon_t + \theta \varepsilon_{t-1} \mid I_{t-1}) = E(\varepsilon_t \mid I_{t-1}) + \theta \cdot \underbrace{E(\varepsilon_{t-1} \mid I_{t-1})}_{\varepsilon_{t-1}} = *$$

we need a new Assumption:

- Zero Conditional mean, $\Rightarrow E(\varepsilon_t \mid I_{t-1}) = 0$
- Recall Assumptions already established: i.e White-noise Assumpt. 2

  1) $E(\varepsilon_t) = 0$ 2) $V(\varepsilon_t) = \sigma^2$ 3) $Cov(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$

$* = \theta \cdot \varepsilon_{t-1} \quad \Rightarrow \quad E(y_t \mid I_{t-1}) = \theta \cdot \varepsilon_{t-1}$

- **Conditional Variance**

$V(y_t \mid I_{t-1}) = V(\varepsilon_t + \theta \varepsilon_{t-1} \mid I_{t-1}) = \cdots = \sigma^2$

Quick reasoning: ε_t & $\varepsilon_t \rightsquigarrow \sigma^2$; ε_t & $\varepsilon_{t-1} \rightsquigarrow 0$; $\varepsilon_{t-1} \rightsquigarrow$ Known

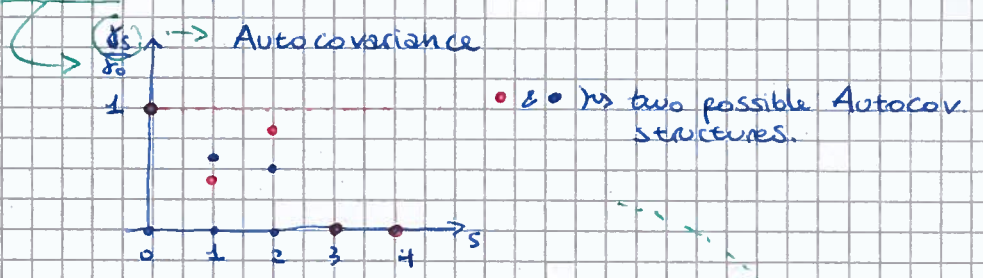
- Remarks: 1) Conditional mean is a function of past observation

  2) Conditional Variance is not

i.e when Conditioning on the past the most likely expectation for y_t becomes the weighted past error.

IF NOT Condit. then expectation goes back to zero.

Consider a MA(2) : $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$



take them as given: Unconditional moments

$E(y_t) = 0$

$V(y_t) = \gamma_0 = \sigma^2 + \theta_1^2 \sigma^2 + \theta_2^2 \sigma^2 = \sigma^2(1 + \theta_1^2 + \theta_2^2)$

$\gamma_1 = \sigma^2(\theta_1 + \theta_1 \theta_2)$

$\gamma_2 = \theta_2 \sigma^2$

Conditional moments:

$E(y_t \mid I_{t-1}) = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$

$\gamma_0 = \sigma^2$

notice: $\gamma_1, \gamma_2 \ldots$ are ... calculate.

↳ $\gamma_1 > 0$ $\gamma_2 > 0$ $\gamma_3 = 0$ $\gamma_4 = 0$

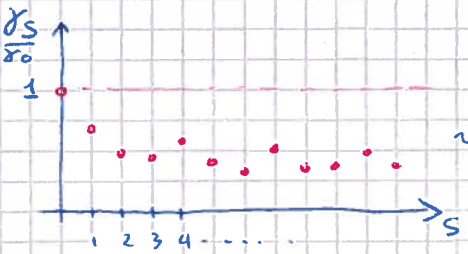
and so on...

Consider MA(∞) : $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots$

- Unconditional moments : $E(y_t) = 0$

$$V(y_t) = \sigma^2 \underbrace{(1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots)}_{*}$$

- $\rho_s = \frac{\gamma_s}{\gamma_0}$



never truncated

{Autocov / Autocorr. can also have negative values.} $\leftarrow$

$*$ notice for $\theta_i < 1 \Rightarrow (1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots) < +\infty$

Consider AR(1) : $X_t = \phi X_{t-1} + \varepsilon_t$

notice: an AR can always be rewritten as a MA(∞)

$X_t = \phi X_{t-1} + \varepsilon_t \Leftrightarrow (1 - \phi L) \cdot X_t = \varepsilon_t \quad (1 - \phi L)^{-1}$

$\Rightarrow X_t = \sum_{i=0}^{\infty} \phi^i L^i \cdot \varepsilon_t$

- Unconditional moments

idea: we have already established how moments work for the MA(·).

$E(X_t) = 0 \qquad V(X_t) = \sigma^2 (1 + \phi^2 + \phi^4 + \phi^6 + \cdots) \quad \leftarrow$ Geom. sum

$= \sigma^2 / 1 - \phi^2 \qquad \dashrightarrow \sum_{i=0}^{\infty} \phi^{2i} = \frac{1}{1-\phi^2}$

$\gamma_1 = \phi^1 \sigma^2 + \phi^3 \sigma^2 + \phi^5 \sigma^2 + \cdots = \phi \sigma \cdot (1/1-\phi^2)$

$\gamma_2 = \phi^2 \sigma^2 \cdot (1/1-\phi^2)$

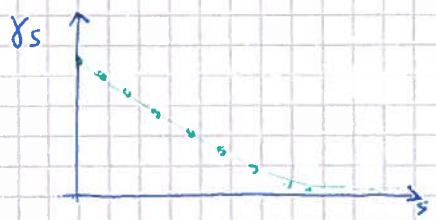
$\gamma_3 = \phi^3 \sigma^2 \cdot (1/1-\phi^2)$

$\vdots$

$\gamma_k = \phi^k \cdot \sigma^2 (1/1-\phi^2) = \phi (\phi^{k-1} \cdot \sigma^2 (1/1-\phi^2)) = \phi \gamma_{k-1}$

$(\phi > 0)$

For $\phi < 1$ $\searrow$

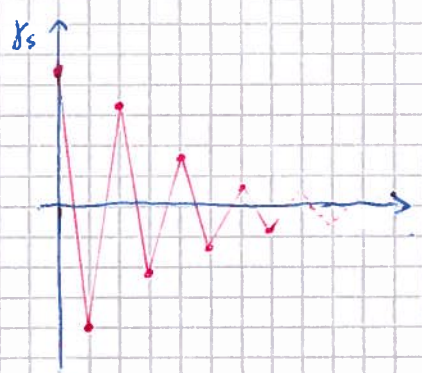


$\Rightarrow \gamma_k \geq \gamma_{k+1}$

never truncated

For $\phi < 0$ $\quad (\phi > -1)$

$\hookrightarrow$



$\Rightarrow |\gamma_k| \geq |\gamma_{k+1}|$

idea: The Autocovariance for an AR(1) is a Geometrically declining sequence.

note that :

1 - $VAR(X_t) = V(\phi X_{t-1} + \varepsilon_t) = \phi^2 V(X_{t-1}) + \sigma^2$

This holds since there is no time correlation : $X_{t-1} \perp \varepsilon_t$

2 - By Stationarity : $V(X_t) = V(X_{t-1})$

$\Rightarrow \quad \gamma_0 = \phi^2 \gamma_0 + \sigma^2 \quad \Rightarrow \gamma_0 = \sigma^2 \cdot (1/1-\phi^2)$

↳ same results without transforming the AR into a MA(∞)

- Conditional moments

$\cancel{E(X_t)} = E(\phi X_{t-1} + \varepsilon_t \mid \mathcal{I}_{t-1}) = \phi X_{t-1}$

$E(X_t \mid \mathcal{I}_{t-1})$

$V(X_t \mid \mathcal{I}_{t-1}) = E(X_t^2 \mid \mathcal{I}_{t-1}) - [E(X_t \mid \mathcal{I}_{t-1})]^2$

$= \cdots = \sigma^2$ (Recall that $E(\varepsilon_t \mid \mathcal{I}_{t-1}) = 0$)

Consider an AR(2) : $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$

Unconditional moments

$E(X_t) = 0$ ⇝ equivalent to AR(1) derivation from MA(∞)

For the Autocovariance, the Yule-Walker equations are used.

i.e Multiply the process by X_t and take the expectation

$\gamma_0 = E(X_t \cdot X_t) = \underbrace{\phi_1 E(X_t \cdot X_{t-1})}_{\phi_1 \cdot \gamma_1} + \underbrace{\phi_2 (X_t \cdot X_{t-2})}_{\phi_2 \gamma_2} + \underbrace{E(X_t \cdot \varepsilon_t)}_{\sigma^2}$

function of $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2} \cdots$ but $\varepsilon_t \perp \varepsilon_{t-h}$

Recall: $E(X_t) = 0 \ \forall t$

$= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2$

Same idea for : $\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1$ → (i.e premult. by X_{t-1} & $E(\cdot)$)

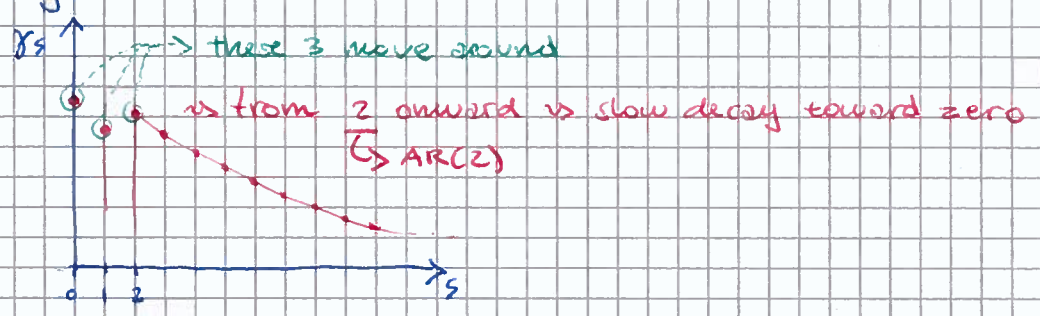
$\gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0$ → " X_{t-2}

$\gamma_3 = \phi_1 \gamma_2 + \phi_2 \gamma_1$

$\vdots$

Generalization $\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$

Graphically :



notice, if $\phi < 0$ ⇝ look before

idea: The decay toward zero begins after after the kth degree Autocov. , where k is AR(k)

General Statement for Autocovariance function.

▷ MA(k) : $|\gamma_0| > 0$ ∗ $|\gamma_1| > 0$ ∗ ... ∗ $|\gamma_k| > 0$, $|\gamma_{k+h}| = 0$ for $h > 0$

→ postive (in absolute values) but declining values until truncation at $k+1$

▷ AR(k) : $|\gamma_0| > |\gamma_1| > \cdots > |\gamma_k| > \bullet |\gamma_{k+1}| > \cdots > 0$

→ declining values but never truncated.

○ Partial Autocorrelation function (PACF)

Consider the following Regressions:

$$y_t = \pi_1 y_{t-1} + v_{1t}$$
$$y_t = \alpha_1 y_{t-1} + \pi_2 y_{t-2} + v_{2t}$$
$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \pi_3 y_{t-3} + v_{3t}$$

{ skipped explanation of OLS } (pg. 15)

we could go onward to get the sequence $\{\pi_i\}_{i=1}^{N}$

notice:

→ since y_{t-k} no exp. power

For an AR(1) ⇒ $\pi_1 = \phi_1$ and $\pi_k = 0$ for $k > 1$

For an AR(2) ⇒ $\pi_2 = \phi_2$ and $\pi_k = 0$ for $k > 2$

Now, consider the MA case:

MA(1) : $x_t = \varepsilon_t + \theta \varepsilon_{t-1}$

⇒ $x_t = \varepsilon_t + \theta(x_{t-1} - \theta \varepsilon_{t-2})$

$x_t = \cdots$ so on until you get:

$x_t = \varepsilon_t + \theta x_{t-1} + \theta^2 x_{t-2} + \theta^3 x_{t-3} + \cdots$ ⇝ AR(∞)

So, For an MA(1) ⇒ $\pi_k \neq 0 \ \forall k$

General Summary:

AR(p)
- → Auto-corr function (ACF) decaying toward zero after p-th term → Geometrically decaying
- → Partial ACF truncated after p-th term

MA(q)
- → ACF truncated after q-th term
- → PACF Geom. decaying toward zero after q-th term

ARMA(p,q)
- → ACF Geom. decaying after q-th term
- → PACF Geom. decaying after p-th term

So far: we identified ARMA models properties

Now: Given a TS we want to find the appropiate ARMA Model

hint: the appropriate model will be the one coherent with the Autocovariance and Partial Autocovariance of the TS.

FIRST ISSUE: True parameters vs Estimated parameters

let $\hat{\rho}_s$ be the estim. Autocorr. of order s

↳ from the data set.

$\rho \sim$ Rho

let ρ_s be true pop. parameter.

One can show that:

- $$\sqrt{T}\,(\hat{\rho}_s - \rho_s) \xrightarrow{d} N\left(0, \sqrt{1 + 2 \cdot \sum_{j=1}^{s-1} \rho_j^2}\right) \quad \text{as } T \to \infty$$

Sample Size

This Result follows from Central limit Theorem

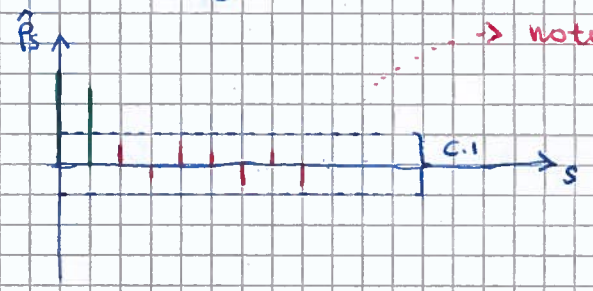
Example

Consider MA(1) $\Rightarrow \sqrt{T}(\hat{\rho}_1 - \rho_1) \xrightarrow{d} N(0, \sqrt{1 + 2\rho_1^2})$

Given the distribution we can build Confidence Intervals around the estimated values:

$$\left[\hat{\rho}_s \pm \left(\frac{1}{T}\right) \cdot \sqrt{1 + 2 \cdot \rho_1^2} \cdot 1.96\right]$$

So, we can test wether any TS can be modeled as a MA(1) by checking if $\hat{\rho}_s$ is statistically different from zero for $s \geq 2$



→ notice: if the Series is indeed a MA(1) the red Autocorr. values will be inside the C.I approxim. 95% of the times:

Finding a value outside the C.I does not imply that the Series is not a MA(1).

Rem. Same idea holds for $\hat{\pi}_s$ (PACF)

Alternative Tests:

- Q-Test or Box-Pierce $\Rightarrow$ H_0: white Noise & $Q_s = T \cdot \sum^{s} \hat{\rho}_k^2 \to \chi^2_s$

For Small Samples

- Ljung-Box $\Rightarrow$ H_0: white Noise & $T \cdot (T+2) \sum_{k=1}^{h} \hat{\rho}_k / T-k \to \chi^2_h$

where h is the n. of lags being Tested

- Estimating a Model ⇝ The Box-Jenkins approach

Objective: find a good Forecasting Model for Stationary TS.

Steps:
1- Specification ⇝ choosing p and q
2- Estimation ⇝ estimate the param: $\hat{\Phi}(L)$ & $\hat{\Theta}(L)$
3- Diagnostic Check
4- Forecasting

---

① Boils down to checking the shapes of the ACF and PACF and choosing p & q accordingly.

Rem. While both the ACF and PACF are unique, ARMAs are not ⇝ In principle there might be several ARMAs that can fit the series. The priority is then to find the ARMA which has less parameters to estimate.

note: The best ARMA in terms of Forecasting might be different from the ARMA that fits the ACF and PACF the better.

---

② Estimation of the parameters is carried by MLE.

note: We cannot use OLS since the MA part doesn't allow it.

dependence on past error terms

↳ ε_t is not exogenous

- Maximum likelihood estimation:

Recall: likelihood = Joint density of the data = Prob. of set of realizat.

max. likelihood = choosing $\hat{\theta}$ s.t prob. of the realizat. is max.

FIRST ISSUE: In TS, observations are not iid

⇒ Joint density ≠ Product of densities

So, we exploit decomposition:

$$f(X_t, X_{t+1}) = f(X_{t+1} \mid X_t) \cdot f(X_t)$$

For T observations:

$$f(X_1, \dots, X_T; \theta) = f(X_T \mid \mathcal{I}_{T-1}; \theta) \cdot f(X_{T-1} \mid \mathcal{I}_{T-2}; \theta) \cdot \dots \cdot f(X_1; \theta)$$

where $\mathcal{I}_{T-k} = \{X_1, \dots, X_{T-k}\}$

Consider an MA(1): $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$ with $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$.

notice: X_t is a lin. comb. of Normally dist. rv

⇒ X_t is also Normally dist.

Furthermore: by properties of the Markovian process

⇒ X_{t-k} has NO influence on X_t for $k \geq 2$

and for X_{t-1} known

Recall that : $X_{t-1} \mid \mathcal{I}_{t-2} \sim N(\theta \varepsilon_{t-2} ; \sigma^2)$

and : $X_1 \mid \mathcal{I}_0 \sim N(\theta \varepsilon_0 ; \sigma^2)$

$\varepsilon_0 \rightsquigarrow$ not observed

$\Downarrow$

most likely value for ε_0 ? $\Rightarrow E(\varepsilon_t) = 0 \quad \forall t$

$\Rightarrow X_1 \mid \mathcal{I}_0 \sim N(0, \sigma^2)$

{ Keep Considering a MA(1)
Remember from Conditional moments

Given these iteration rules we get:

$$X_1 = \varepsilon_1 + \theta \varepsilon_0 = \varepsilon_1 \quad \text{since} \quad \varepsilon_0 = 0$$

$$X_2 = \varepsilon_2 + \theta \varepsilon_1 = \varepsilon_2 + \theta X_1$$

$$X_3 = \varepsilon_3 + \theta \varepsilon_2 = \varepsilon_3 + \theta (X_2 - \theta X_1)$$

$$\Downarrow$$

$$X_1 \sim N(0, \sigma^2)$$

$$X_2 \sim N(\theta X_1, \sigma^2) = N(\theta \varepsilon_1 ; \sigma^2)$$

$$X_3 \sim N(\underbrace{\theta (X_2 - \theta X_1)}, \sigma^2) = N(\theta \varepsilon_2 ; \sigma^2)$$

$\hookrightarrow$ function of past observations.

In General likelihood becomes : $\prod f(X_t \mid X_{t-1} ; \theta, \sigma^2)$

Consider an AR(1) : $X_t = \phi X_{t-1} + \varepsilon_t$

Similarly to the MA, Recall :
- $E(X_t \mid \mathcal{I}_{t-1}) = \phi X_{t-1}$
- $V(X_t \mid \mathcal{I}_{t-1}) = \sigma^2$
- AR(1) $\rightsquigarrow$ MA(∞) $\rightsquigarrow$ lin. comb. of normally dist. r.v. $\Rightarrow$ Norm. dist.

$\Rightarrow \quad X_t \mid \mathcal{I}_{t-1} \sim N(\phi X_{t-1} ; \sigma^2)$

$$f(X_t \mid \mathcal{I}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \cdot \underbrace{(X_t - \phi X_{t-1})}_{= \varepsilon_t}{}^2 \right\}$$

then, the Lik. hood :

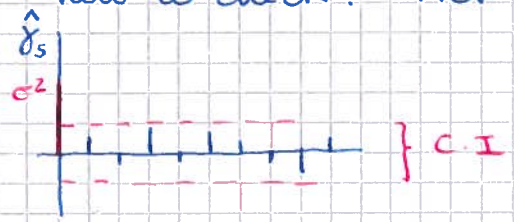
$$\prod_{t=0}^{T} f(X_t \mid \mathcal{I}_{t-1}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^T \cdot \exp\left\{ -\frac{1}{2\sigma^2} \cdot \sum_{t=0}^{T} \varepsilon_t^2 \right\}$$

i.e.
- to maximize the likelihood (or log lik.) is equivalent to minimize (notice the negative in the exponential) the Residuals sum of squares. (RSS)
- This is true for all AR processes
- minimizing the RSS $\approx$ OLS procedure
- This makes sense as the coeff. for AR processes can also be estimated through OLS (not true for MA)

③ Diagnostic checks

1- IF the model fits the data well, then the main assumption of ARMAs should be true → error terms (residuals) do not have time dependence ⇝ i.e white noise

how to check? ACF and PACF



⇝ no time dependence

⇝ Visual Test.

Formal Test: Ljung-Box Test

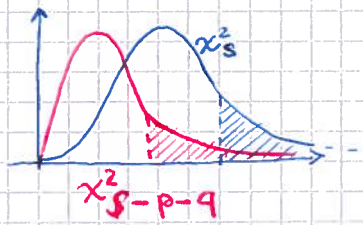
Consider the Statistics $Q = T(T+2)\sum_{k=1}^{s} \hat{\rho}(k)^2 / (T-k)$

where $\hat{\rho}(k)$ is the ACF estim.

Under H_0: $\hat{\rho}(k) = 0 \quad \forall k \geq 1 \quad \Rightarrow \quad Q \xrightarrow{d} \chi^2_{s-p-q}$

notice, the d.o.f of the Chi-Square are adjusted by the number of parameters in the model ⇝ we'll reject the Null with much lower values of the Chi-square

visually:



2- Information Criteria

We want the model to minimize the RSS (measure of goodness of fit) but, at the same time, we don't want to over-parametrize the model.

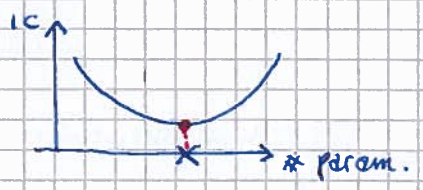
Over-parametrization ⇝ Reduces drastically the RSS but makes the model too specific for the estimation sample ⇝ effect: the model will perform very badly when it is asked to estimate new data (Forecasting)

IC: Info. Criterias weight the RSS and the loss in the d.o.f

Here, two IC:

① AKAIKE : $T \cdot \ln\left(\frac{RSS}{T}\right) + 2 \cdot (1+p+q)$

② SCHWARTZ : $T \cdot \ln\left(\frac{RSS}{T}\right) + \ln(T) \cdot (1+p+q)$



- Goal: we want to minimize the IC
- notice: The SCHWARTZ penalizes more an excessive number of parameters since $\log(T) > 2$ for almost all T.
- In General: AKAIKE ⇝ better when the data generating process is outside the class of models I'm considering.

when we are confident on the given specification | SCHWARTZ ⇝ better when the data gen. process is inside the c.o. models I'm considering.

④ Forecasting

AR(1)

$E(X_{t+1} \mid \mathcal{I}_t) = \phi X_t$

↳ Cond. Exp is the Forecast, notation used is: $X_{t+1|t}$

$X_{t+2|t} = \phi \cdot X_{t+1|t} = \phi^2 X_t$ (two step ahead)

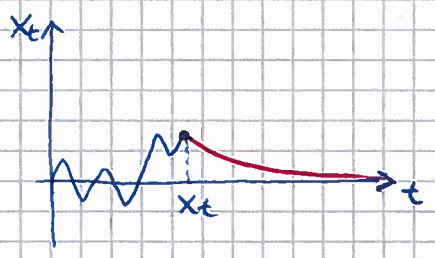
$X_{t+h|t} = \phi^h X_t$ (h-step)

- notice, as $h \to +\infty \Rightarrow \phi^h X_t \to 0 = E(X_t)$

i.e Forecasts far away in the future will approximate the Unconditional Expectation of X_{t+1}

↳ X_t looses importance in estim./forecasting X_{t+h} when h is too big. (of course only for $|\phi| < 1$ stationary)

Graphically:



- Forecast error ↳ realization ↳ Forecast

1 step: $X_{t+1} - \phi X_t = \phi X_t + \varepsilon_{t+1} - \phi X_t = \varepsilon_{t+1}$

[notice: when using population parameters (assume known) the Forecast error variance = Forecast variance]

$V(X_{t+1|t}) = V(\varepsilon_{t+1}) = \sigma^2$

2 step: $\phi X_{t+1} + \varepsilon_{t+2} - \phi^2 X_t = \phi(\phi X_t + \varepsilon_{t+1}) + \varepsilon_{t+2} - \phi^2 X_t$

$= \cdots = \varepsilon_{t+2} + \phi \varepsilon_{t+1}$ and $V(X_{t+2|t}) = (1+\phi^2)\sigma^2$

k step: $V(X_{t+k|t}) = (1+\phi^2+\phi^4+\phi^6+\cdots)\cdot\sigma^2$

So, for $k \to +\infty$: $V(X_{t+k|t}) \approx \dfrac{\sigma^2}{1-\phi^2} = V(X_{t+1})$

Again, as the forecast step goes long into the future the variance of the Forecast approximates the Unconditional Variance.

▷ Forecast moments converge to Unconditional moments.

We can Generalize this statement also for models with constant terms: $X_t = c + \phi X_{t-1} + \varepsilon_t$ AR(1) + Constant

notice:

let $\mu = E(X_t)$

Consider an AR(1) with X_t taken as deviation from the mean:

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t$$

$$X_t = \mu - \phi\mu + \phi X_{t-1} + \varepsilon_t$$

eq.1 $\rightarrow$ $$X_t = (1-\phi)\mu + \phi X_{t-1} + \varepsilon_t$$ → AR(1) + cost. term

Reconsider AR(1) with cost. term:

$$X_t = c + \phi X_{t-1} + \varepsilon_t = c + \phi(c + \phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$= \cdots = (1+\phi+\phi^2+\cdots)\cdot c + \sum_{j=0}^{\infty} \varepsilon_{t-j}\cdot\phi^j$$

$$E(X_t) = E((1+\phi+\phi^2+\cdots)\cdot c) + \underbrace{E(\quad)}_{=0}$$

$$= (1+\phi+\cdots)\cdot c$$

$$= \frac{1}{1-\phi}\cdot c$$

Therefore $\mu = E(X_t) = \dfrac{1}{1-\phi}\cdot c \Rightarrow (1-\phi)\cdot\mu = c$

i.e AR(1) + cost. term $\Leftrightarrow$ AR(1) with X_t being dev from mean

We just derived: $\boxed{E(X_t) = \dfrac{1}{1-\phi}\cdot c}$

Conditional Exp: $E(X_{t+1}|\mathcal{I}_t) = \phi X_t + c$ 1 step

$E(X_{t+2}|\mathcal{I}_t) = (1+\phi)\cdot c + \phi^2 X_t$ 2 step

$\vdots$

$$E(X_{t+k}|\mathcal{I}_t) = \underbrace{(1+\phi+\phi^2\cdots)\cdot c}_{\to \frac{1}{1-\phi}\cdot c} + \underbrace{\phi^k\cdot X_t}_{\to 0}$$

As $k \to +\infty$: $X_{t+k|t} \to \dfrac{1}{1-\phi}\cdot c$ → Unconditional mean

- **Confidence Intervals of the Forecast**

notice: the forecast error (1-step) is $\varepsilon_{t+1} \sim N(0, \sigma^2)$

So, the conf. interval for 1-step ahead forecast:

$$(c + \phi X_t) \pm 1.96\, \sigma$$

2-step :

$$(1+\phi)\cdot c + \phi^2 X_t \pm 1.96 \sqrt{(\phi^2+1)\cdot\sigma^2}$$

so on...

MA(1) : $X_t = \varepsilon_t + \theta\, \varepsilon_{t-1}$

Forecast $\Rightarrow$ $X_{t+1|t} = E(\varepsilon_{t+1} + \theta\cdot\varepsilon_t \mid \mathcal{I}_t) = \theta\cdot\varepsilon_t$

Forc. err. $\Rightarrow$ $X_{t+1} - X_{t+1|t} = \varepsilon_{t+1}$ and $V(\varepsilon_{t+1}) = \sigma^2$

2 step

$X_{t+2|t} = E(\varepsilon_{t+2} + \theta\cdot\varepsilon_{t+1} \mid \mathcal{I}_t) = 0$

$X_{t+2} - X_{t+2|t} = \varepsilon_{t+2} + \theta\, \varepsilon_{t+1}$ and $V(\cdot) = (1+\theta^2)\cdot\sigma^2$

- **Pseudo out of Sample Forecast**

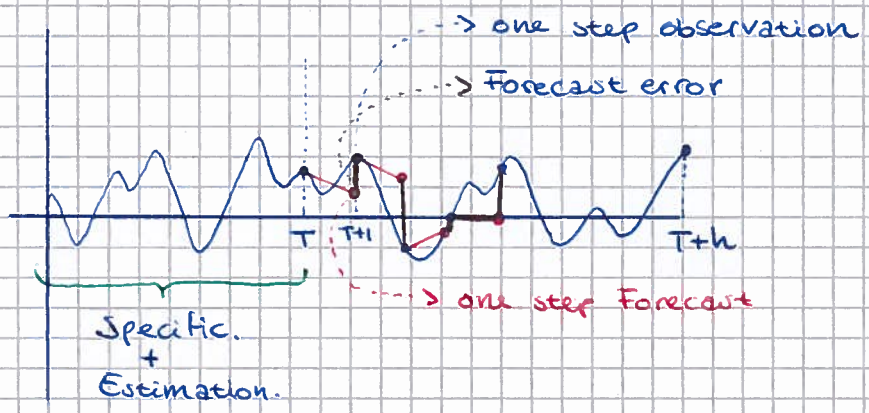
idea: We want to test the Forecast ability of the model

Step 1 $\rightarrow$ Divide the Sample in two parts

Step 2 $\rightarrow$ Specification + Estimation + Diagnostic on the first half

Step 3 $\rightarrow$ Forecast on the Second half.

e.g : We want to test the 1-step ahead Forecast ability of the model $\Rightarrow$ how? run recursively the forecast using the prev. specified model, by integrating at each step the new observation in the forecast conditioning. Then, use a <u>measure of Forecast precision</u>



Measures of Forecast precisions.

1) RMSFE $= \sqrt{\sum_{i=T}^{T+h} (y_i - \hat{y}_i)^2 / h}$

2) MAE $= \sum_{i=T}^{T+h} |y_i - \hat{y}_i| / h$

- **Impulse Response**

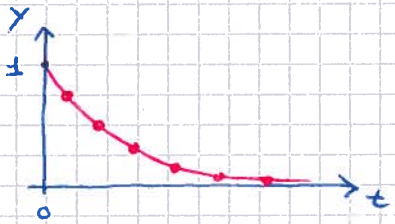
Assume there is a shock at time t . i.e $\varepsilon_t \uparrow$

How do the process go back in equilibrium?

(notice, the shock is not permanent, after t $\Rightarrow \varepsilon_{t+1} = 0$ . . . .

AR(1)

t	-1	0	1	2	3
ε_t	0	1	0	0	0
y_t	0	1	ϕ	ϕ^2	ϕ^3

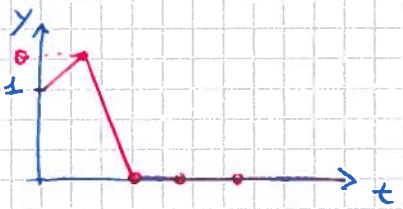


notice: the effect of the shock persits in time :

$y_{t-h} = 0 \quad \forall h$ while $y_{t+h} = \underset{\to 0}{\phi^h} \quad \forall h$

MA(1)

t	-1	0	1	2	3
ε_t	0	1	0	0	0
y_t	0	1	θ	0	0



notice: the effect of the shock vanishes after the first period

Remark

▷ An MA process is always Stationary $\forall \theta$

∘ Wold Theorem

- **Preliminaries**

  Invertibility of an MA process

Q: What is Invertibility?

It's a key property of any MA process. It allows to rewrite the residual (OR Forecast error) as an infinite lin. combination of past observations.

Formal definition

A MA process ($y_t = C(L) \cdot \varepsilon_t$) is invertible IF:

$\exists \{\partial_i\}_{i=0}^{\infty}$ s.t $\sum_{i=0}^{\infty} \partial_i \, y_{t-i} = \varepsilon_t$

Recall: Invertibility can be checked through the Charcteristic eq.

We'll make an example, then provide the proposition.

Consider an MA(1): $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$

$y_t = (1 + \theta L) \cdot \varepsilon_t$ using the lag operator

To invert the process we want to have: $\varepsilon_t = (1+\theta L)^{-1} \cdot y_t$

but we don't know whether the polynomial $(1+\theta L)$ is indeed invertible ⇝ we can check this using the characteristic equation, which is derived by:

1- substitute L with a general variable z in the polynomial:

$$1 + \theta \cdot z$$

2- Set it equal to zero ⇝ i.e find the roots wrt z

$$1 + \theta \cdot z = 0$$

$$z = -1/\theta$$

3- IF the MA process is invertible $\Rightarrow |z| > 1$

equivalent to: $|\theta| < 1$

Proposition

An MA process is invertible IFF $C(z) \neq 0 \;\; \forall z$ s.t $|z| < 1$

(Alt. $C(z) = 0$ only for $|z| > 1$)

Remark

- Invertibility (of MA) allows us to transform an MA(p) into AR(∞)
- Also, fundamental Assumption of Wold Theorem
- if we consider a non-invert. MA, we get the wrong Impulse responses.

Q: What if the MA process is not invertible? no worries

Recall: A MA(p) describes a unique Autocovariance (or PACF) but the MA(p) is not unique in itself.

i.e Unique ACF $\Rightarrow$ several potential ARMA

In particular we'll see that for any non-invertible MA(p) process able to explain an ACF, there exists a counter-face invertible MA(p) able to explain the same exact ACF

e.g

Consider the MA(1): $y_t = \varepsilon_t + m \cdot \varepsilon_{t-1}$

with $\varepsilon_t \sim W.N$ and $|m| > 1 \Rightarrow$ non-invertible

The Autocovariance structure is:

$\gamma_0 = (1+m^2)\cdot\sigma^2$ (known form, no deriv. needed)

$\gamma_1 = m \cdot \sigma^2$

$\gamma_2 = 0$

$\gamma_3 = 0$ and so on...

Now, consider the Alternative formulation:

$$y_t = \varepsilon_t + (1/m)\cdot\varepsilon_{t-1} \quad \text{s.t } V(\varepsilon_t) = m^2\sigma^2$$

notice that $|1/m| < 1$ (for $|m| > 1$) $\rightsquigarrow$ invertible

and, the Autocovariance structure:

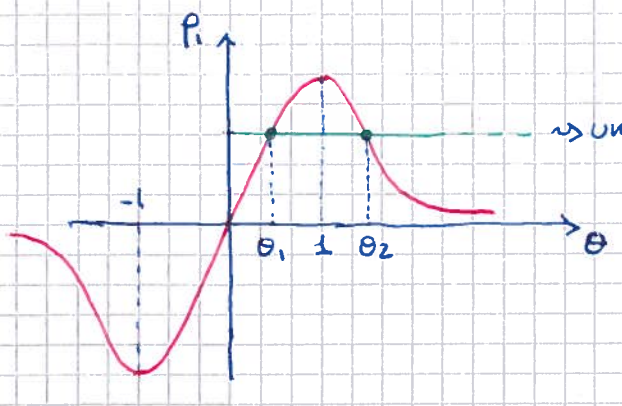
$\gamma_0 = E(y_t^2) = (1 + 1/m^2)\cdot m^2 \cdot \sigma^2 = (1+m^2)\cdot\sigma^2$

$\gamma_1 = \cdots = m \cdot \sigma^2$

$\gamma_2 = 0$ and so on.

Notice: Identical ACF.

In general, for a MA(1) we can represent the first order Autocorrelation function (ρ_1) as a function of possible values of θ.



$\rightsquigarrow$ unique ACF, 2 possible MA repr.:

- one with $\theta_2 > 1$ $\rightsquigarrow$ non-invertible
- one with $\theta_1 < 1$ $\rightsquigarrow$ invertible
- and $\theta_1 = 1/\theta_2$

◦ Wold Theorem

PROP. Any Zero-mean Covariance Stationary Stochastic process $\{X_t\}_{t=-\infty}^{+\infty}$ can be represented as:

$$X_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i} + \eta_t$$

where:

1. $\varepsilon_t \equiv X_t - Proj(X_t \mid X_{t-1}, X_{t-2}, \cdots)$
2. $Proj(\varepsilon_t \mid X_{t-1}, X_{t-2} \cdots) = 0$ & $E(\varepsilon_t \cdot X_{t-i}) = 0$
3. W.N: $E(\varepsilon_t) = 0$; $V(\varepsilon_t) = \sigma^2$; $E(\varepsilon_t \cdot \varepsilon_{t-i}) = 0$
4. ε_t is the limit of linear combinations of X_{t-s}, $s \geq 0$
5. $\theta_0 = 1$ and $\sum_{i=0}^{\infty} \theta_i^2 < +\infty$
6. $\{\theta_i\}$ and $\{\varepsilon_i\}$ are unique
7. η_t is linearly deterministic: $\eta_t = Proj(\eta_t \mid X_{t-1}, X_{t-2} \cdots)$

notice: Proj = Projection ⇝ i.e linear regression of y on X : $Proj(y \mid X)$
So, $Proj(y \mid X)$ is equivalento to $\hat{y}$ ⇝ fitted value / Forecast.

Remarks:

1. ε_t is clearly the Forecast error ⇝ $\varepsilon_t \equiv X_t - X_{t|t-1}$
2. Properties of the F.E : Orthogonal to the regressors
   ↳ we cannot have that past info explains the forecast error!
3. white-Noise Assumptions
4. ε_t is a function of past and present X_t ⇔ Invertibility
   ↳ note: this is not in contradiction with point 2.
5. guarantees (together with $V(\varepsilon_t) = \sigma^2$) that X_t has finite Variance
6. Any ACF has a unique Wold decomposition
   ↳ notice that this is a step toward wrt ARMA.

   Now, the unique features of a process are:

   1. ACF / PACF
   2. Wold Decomposition
   3. Spectral Decomposition

7. There may be parts of the process which are perfectly forecastable on the basis of past X_t.

Some intuition:

- A stochastic process can be considered as an infinite vector of random variables $\Rightarrow$ $\{X_t\}_{-\infty}^{+\infty} \in \mathbb{R}^{\infty}$
- Recall that in math, any vector can be expressed as a unique linear combination of the basis of the vector space
- IF we consider the vector space of stationary stoc. process, by Wold, we have that the basis of such vector space are the infinite series of forecast errors
- forecast errors are orthogonal to each other $\Rightarrow$ so they are eligible as basis
- Furthermore, F.E are white noises

$\Rightarrow$ Any stochastic process can be expressed as the linear combination of white noises.

$\triangleright$ notice: not all Wold representations (Stoc. Stat. process) can be represented as ARMAs

Wold : $X_t = \Theta^{W}(L) \cdot \varepsilon_t$ $\Rightarrow$ W is just to distinguish from MA

ARMA : $X_t = \dfrac{\Theta(L)}{\phi(L)} \cdot \varepsilon_t$

i.e $\Theta^{W}(L)$ not always representable as $\dfrac{\Theta(L)}{\phi(L)}$

Back to Invertibility

- In order to use Wold decomposition we must be able to express ε_t as a lin. combination of past and present observation i.e Invertibility of the MA
- One peculiar case, consider: $X_t = \varepsilon_t + \varepsilon_{t-1}$
  - $\hookrightarrow$ Polynomial is not invertible BUT $X_t = \varepsilon_t + \varepsilon_{t-1}$ is a Wold decomposition

Last note: Fundamental innovation $\Rightarrow$ ε_t

i.e ε_t is called F.I since it cannot be predicted on the basis of the past.

Non-Stationary Series

Let us see different kinds of N-S Series:

- **Random Walks + noise**

$X_t = X_{t-1} + \varepsilon_t$

- moments:

$E(X_t) = E(X_{t-1} + \varepsilon_t) = E(X_{t-2} + \varepsilon_{t-1} + \varepsilon_t) =$

$= \dots = E(X_0 + \sum_{i=0}^{\infty} \varepsilon_i) = X_0$ ⇝ Static

$\gamma_0 = V(X_t) = (1+t) \cdot \sigma^2$ ⇝ time dependent

$\vdots$

$\gamma_s = (t-s) \cdot \sigma^2$ & $\rho_s = (t-s) / \sqrt{t(t-s)}$

- Impulse Response

$\frac{\partial X_{t+k}}{\partial \varepsilon_t} = 1$ ⇝ Permanent and persistent effect on the process

- Forecast: $X_{t+1|t} = X_t \quad \dots \quad X_{t+k|t} = X_t$

⇝ last observation

- **Random Walks + Drift** ⇝ "Unit root process"

$X_t = \delta + X_{t-1} + \varepsilon_t$ ⇝ $X_t = X_0 + \delta \cdot t + \sum_{i=0}^{t} \varepsilon_i$

- moments:

$E(X_t) = X_0 + \delta \cdot t$ ⇝ time dependent

$\gamma_0 = (t+1) \cdot \sigma^2$

$\rho_s = (t-s) / \sqrt{t(t-s)}$

- Impulse Response

$\frac{\partial X_{t+k}}{\partial \varepsilon_t} = 1$

→ Variance, ACF and Input resp. remain as in the RW+noise.

- Forecast: $X_{t+1|t} = \delta + X_t \quad \dots \quad X_{t+k|t} = \delta \cdot k + X_t$

- Only difference is that the process has a trend dictated by δ

- **Deterministic Time Trend**

$X_t = \alpha + \beta \cdot t + \varepsilon_t$

- moments

$E(X_t) = \alpha + \beta \cdot t$

$V(X_t) = \sigma^2$

⇝ $F.E = \sigma^2$

- Impulse Response

$\frac{\partial X_{t+k}}{\partial \varepsilon_t} = 0$

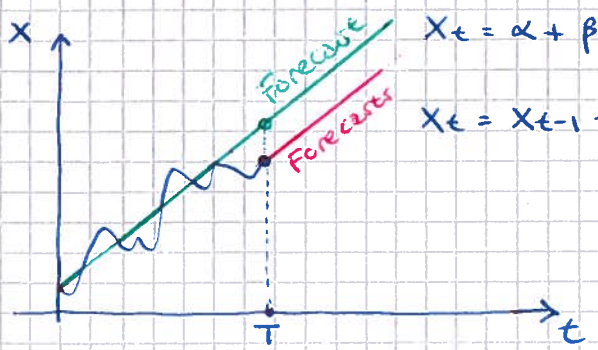
↓

no impulse

- Forecast

$X_{t+k|X_t} = \alpha + \beta(t+k)$

Visually : RW + Drift vs Deterministic Trend.



$X_t = \alpha + \beta t$

$X_t = X_{t-1} + \beta \cdot t$

Can we make these processes Stationary? two methods :

1) First difference

2) Subtract from the TS a costant and a time trend

notice : 2) is equiv. to Regressing the TS on a cost. and a trend

---

- First difference method.

Consider $X_t = \delta + X_{t-1} + \varepsilon_t$ Unit root AR

re-write as $\underbrace{X_t - X_{t-1}}_{Y_t} = \delta + \varepsilon_t$

$Y_t = \delta + \varepsilon_t$ ⇒ Stationary

notice that $Y_t = (1-L) X_t$

i.e First difference on Unit roots works!

Consider $X_t = \alpha + \beta \cdot t + \varepsilon_t$ Determ. time trend

First diff. ⇒ $Y_t = (1-L) X_t = \cancel{\alpha} + \beta t + \varepsilon_t - \cancel{\alpha} - \beta(t-1) - \varepsilon_{t-1}$

$Y_t = \beta + \varepsilon_t - \varepsilon_{t-1} = \underbrace{\beta + (1-L) \cdot \varepsilon_t}_{\text{just like a MA(1) + costant}}$

A Moving Average process with Unit root (as in this case) is :

* ↳ non-Invertible BUT Stationary ⇒ i.e it works!

- Regressing on costant + time trend

Consider $X_t = \delta + X_{t-1} + \varepsilon_t$

Rewrite as : $X_t = X_0 + \delta \cdot t + \sum_{i=0}^{t} \varepsilon_i$

Regress X_t on $(\underset{\text{cost.}}{X_0}, \underset{\text{time trend}}{\delta t})$: $Y_t = \sum_{i=1}^{t} \varepsilon_i$

notice that Y_t is an AR with Unit root ⇒ non-Stationary

why? ⇒ Variance of the process explodes as $t \uparrow$ ⇒ $V(Y_t) = t \cdot \sigma^2$

e.g let $t \to \infty$ ⇒ $\sum^{\infty} \varepsilon_i$ ⇒ $V(\sum^{\infty} \varepsilon_i) \to +\infty$

Consider $X_t = \alpha + \beta \cdot t + \varepsilon_t$

Regress on cost. & time trend $\Rightarrow$ $y_t = (X_t - \alpha - \beta \cdot t) = \varepsilon_t$

Stationary!

Remark:

- The only way to make a Unit root process (+ drift or not) Stationary is to take the first difference
- BUT Sometimes the first difference is not enough

- **Generalized Version of a Random Walk** (Unit root process)

Consider $\mu_t \sim ARMA(p, q)$

s.t $\phi(L)\mu_t = \theta(L)\varepsilon_t$

$\phi(L)$ ↳ Stationary ⇓ the Polynomial has roots < 1 in abs term.

$\theta(L)$ ↳ Invertible Polynomial

$$\Rightarrow \quad \mu_t = \frac{\theta(L)}{\phi(L)} \cdot \varepsilon_t$$

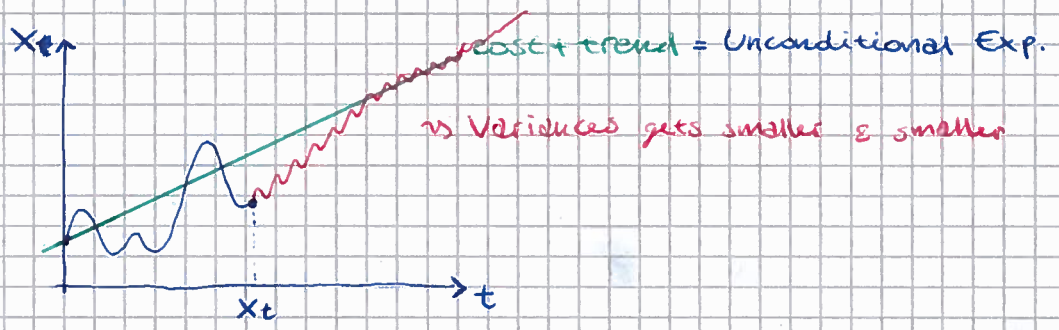
Now Consider: $X_t = \alpha + \beta \cdot t + \mu_t$

$$= \alpha + \beta \cdot t + \theta(L)/\phi(L) \cdot \varepsilon_t \qquad (1)$$

How does this Model behave?

Rewrite as: $y_t = X_t - \alpha + \beta \cdot t = \frac{\theta(L)}{\phi(L)} \cdot \varepsilon_t$

Then y_t is an ARMA process too and, the important thing for us to know is that the Forecast converges to the Unconditional expectation of the process (by prop. of ARMAs) $\Rightarrow$ Then the Forecast of X_t is exactly the Forecast of y_t BUT with a Trend and a costant



Notice that, we can multiply (1) by $\phi(L)$ and still get an equivalent process:

(1) becomes $\phi(L) X_t = \underbrace{\phi(L)\alpha}_{\text{Costant}} + \underbrace{\phi(L)\beta \cdot t}_{\text{Term}} + \underbrace{\theta(L) \cdot \varepsilon_t}_{\text{error term}}$

error term = ARMA

How does this help us?

In (1) we assumed the AR process $\phi(L)$ to be Stationary, meaning that All the roots of the polynomial $(\phi(L))$ are assumed to be less than 1 in module.

We can rewrite $\phi(L)$ as:

by collecting

$$\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots)$$
$$= (1-\lambda_1 L)\cdot(1-\lambda_2 L)\,(1-\lambda_3 L)\cdot \dots \cdot(1-\lambda_p L)$$

where these are the p roots of the AR(p)

i.e in (1) All λ_i with $i \leq p$ where s.t $|\lambda_i| \leq 1$

Now, Consider:

$$\phi(L) = \underbrace{(1-L)}_{\text{Unit root.}} \cdot \underbrace{(1-\lambda_2 L) \cdots (1-\lambda_p L)}_{\phi^*(L) \rightsquigarrow \text{Stationary}}$$

notice that $\phi(L)$ can be written as:

$$\Rightarrow \quad \phi(L) = \underbrace{(1-L)}_{\text{U.r}} \cdot \underbrace{\phi^*(L)}_{\text{Stationary AR poly.}}$$

So, the ARMA in this case would be:

$$(1-L)\cdot\phi^*(L)\cdot \mu_t = \theta(L)\cdot \varepsilon_t$$
$$(1-L)\,\mu_t = \underbrace{\frac{\theta(L)}{\phi^*(L)}\cdot \varepsilon_t}_{\text{Stationary ARMA}}$$

idea: We can rewrite the Original Model:

$$X_t = \alpha + \beta\cdot t + \underbrace{\mu_t}_{\text{not Stationary}}$$

1) Multiply by $(1-L)$ to get:

$$(1-L)\,X_t = (1-L)\,\alpha + (1-L)\,\beta\, t + \underbrace{(1-L)\,\mu_t}_{\text{Stationary}}$$

2) ~~By subtracting constant and trend we get a~~ Stationary process

Remarks

1) μ_t is called: ARIMA(p, d, q)

where d is the number of unit roots in the AR Polynomial

2) with d unit roots in the AR polynomial we multiply by $(1-L)^d$ $\rightsquigarrow$ $(1-L)^d\cdot X_t = (1-L)^d\alpha + (1-L)^d\cdot\beta t + (1-L)^d\mu_t$

Notice that :

$(1-L)\cdot\alpha = \alpha - \alpha = 0$

$(1-L)^d\cdot\alpha = (1-L)^{d-1}\cdot(1-L)\alpha = (1-L)^{d-1}\cdot 0 = 0$

& :

$(1-L)\cdot\beta\cdot t = \beta t - \beta(t-1) = \beta$

$(1-L)^d\cdot\beta t = (1-L)^{d-1}\cdot\beta = (1-L)^{d-2}\cdot(1-L)\cdot\beta = 0$

$\Rightarrow$ For $d \geq 2$:

$$\underbrace{(1-L)^d\cdot X_t}_{} = \underbrace{(1-L)^d\cdot\mu_t}_{}$$

$$\Rightarrow \quad \Delta^d X_t = \frac{\theta(L)}{\phi^*(L)}\cdot\varepsilon_t$$

$$\Rightarrow \quad \phi^*(L)\,\Delta^d X_t = \theta(L)\cdot\varepsilon_t$$

Stationary ARMA

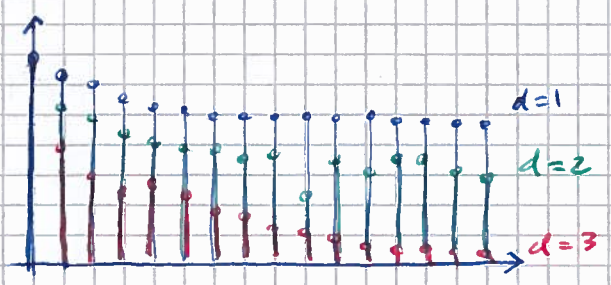
$\Rightarrow$ For $d=1$ $\Rightarrow$ $\phi^*(L)\cdot\Delta X_t = \underbrace{\phi^*(L)\cdot\beta}_{\text{Costant}} + \theta(L)\cdot\varepsilon_t$

$\hookrightarrow$ Still, Stationary ARMA + Costant.

Remarks

- $X_t \sim ARIMA(p,d,q) = \Delta^d X_t \sim ARMA(p,q)$
- In General, we have $\mu_t \sim ARIMA$ and $X_t = \text{cost.} + \text{trend} + ARIMA$ $\rightarrow$ once we take the difference $\Delta^d X_t \sim ARMA$

- Visual Test for Stationarity
  - Given a non-Stationary process, take the difference (increase d) until the ACF converges to zero.



◦ Unit Roots Tests

idea: We want to create Tests to determine wether the AR process has a unit root or not.

Consider $X_t = \phi X_{t-1} + \varepsilon_t$

One can proove that: $\sqrt{T}\cdot(\hat{\phi} - \phi) \xrightarrow{d} N(0, 1-\phi^2)$

How? Informal Proof:

by definition, $\hat{\phi}$ is the MLE = OLS estimator

$$\Rightarrow \quad \hat{\phi} = \left(\frac{1}{T}\right)^{-1}\left(\sum^T X_{t-1} X_{t-1}\right)^{-1} \cdot \left(\sum^T X_{t-1} \cdot X_t\right) \cdot \frac{1}{T}$$

We can rewrite this as a function of X_{t-1} exclusively

$$\hat{\phi} = \left(\frac{1}{T}\right)^{-1} \cdot \left(\sum^T X_{t-1}^2\right)^{-1} \cdot \underbrace{\left(\sum^T X_{t-1} \cdot (\phi X_{t-1} + \varepsilon_t)\right)}_{\phi \sum^T X_{t-1}^2 + \sum^T X_{t-1} \cdot \varepsilon_t} \cdot \frac{1}{T}$$

$$\hat{\phi} = \left[\cancel{\left(\frac{1}{T}\right)^{-1}} \cancel{\left(\sum X_{t-1}^2\right)^{-1}} \cdot \phi \cdot \cancel{\sum^T X_{t-1}^2}\right] + \left[\left(\frac{1}{T}\right)^{-1} \cdot \left(\sum^T X_{t-1}^2\right)^{-1} \cdot \sum^T X_{t-1} \cdot \varepsilon_t \cdot \frac{1}{T}\right]$$

$$\hat{\phi} = \phi + \underbrace{\left(\frac{1}{T}\sum^T X_{t-1}^2\right)^{-1}} \cdot \underbrace{\left(\frac{1}{T} \cdot \sum^T X_{t-1} \cdot \varepsilon_t\right)}$$

notice $E(X_{t-1}) = 0$ since it's an AR process

$$\Rightarrow \left(\frac{1}{T} \cdot \sum^T X_{t-1}^2\right) \rightarrow V(X_{t-1})$$

by Slutsky

where $V(X_{t-1}) = \dfrac{\sigma^2}{1-\phi^2}$

notice that by The Central limit Theor.:

$$\sqrt{T} \cdot \left(\frac{1}{T} \cdot \sum^T \overset{\bar{X}_T}{X_{t-1} \cdot \varepsilon_t} - \underset{\overset{\shortparallel}{0}}{\overset{\mu}{E(X_{t-1} \cdot \varepsilon_t)}}\right) \xrightarrow{d} N(0, V(X_{t-1} \cdot \varepsilon_t))$$

$$\Downarrow$$

$$\sqrt{T} \cdot \frac{1}{T} \cdot \sum^T X_{t-1} \cdot \varepsilon_t \rightarrow N\left(0, \underbrace{\frac{\sigma^2}{1-\phi^2}}_{V(X_{t-1})} \cdot \underbrace{\sigma^2}_{V(\varepsilon_t)}\right)$$

This term doesnt allow the variance to converge to zero

Also called ⇝ Stabilizing Factor

Why we don't want Variance to Converge to zero?

~ i.e it's expected that V converges to zero since the estimator is consistent, but in order to construct a test we need a positive Asymptotic Variance

So, we have:

$$\sqrt{T}\cdot(\hat{\phi}-\phi) = \left(\frac{\sigma^2}{1-\phi^2}\right)^{-1}\cdot\left(\frac{\sqrt{T}}{T}\cdot\sum^{T} x_{t-1}\cdot\varepsilon_t\right) \xrightarrow{d} N(0, \underline{\underline{V}})$$

notice that the Variance is equal to

$$\left(\frac{\sigma^2}{1-\phi^2}\right)^{-2}\cdot\left(\frac{\sigma^2}{1-\phi^2}\right)\cdot\sigma^2 = (1-\phi^2)$$

$$\Rightarrow \quad \sqrt{T}\cdot(\hat{\phi}-\phi) \xrightarrow{d} N(0, 1-\phi^2)$$

idea:

- notice that this result holds IFF $\phi < 1$
- For $\phi = 1$ (say $H_0: \phi = 1$) the Asymptotic Variance of $\hat{\phi}$ is zero ⇒ Again, can't construct a test!
- This is because, the Stibilizing Factor $\sqrt{T}$ is not enough (converges to infinity too slowly)
- To correct the Sampling error we need a new Stab. factor: T

- Dickey and Fuller (DF)

$$T\cdot(\hat{\phi}-1) \rightarrow \text{DF distribution}$$

So, imagine the Following Test (t-Test)

$H_0: \phi = 1$

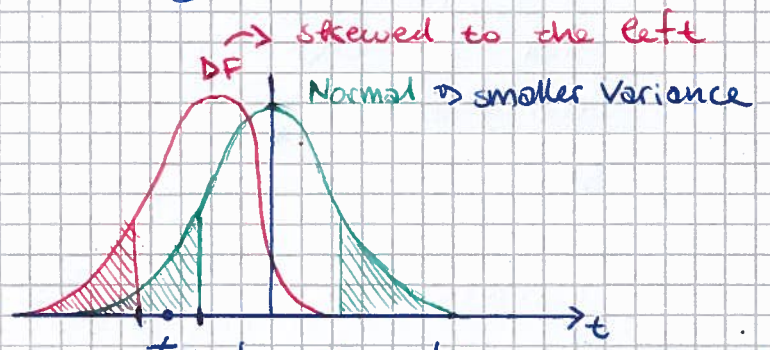
$H_1: \phi < 1$

$$t = \frac{\hat{\phi}-1}{\sqrt{Var(\hat{\phi})}}$$

↳ First advantage of the DF is that the Test is one sided. Using a t-test with a normal dist would require a two sided ~~test~~ Test s.t $H_0: \phi = 1$ $H_1: \phi \neq 1$

But in this case we are interested in the case of $\phi < 1$

Visually: DF vs Normal



↳ for a value of t in this range the Normal dist. would have rejected the Null, indicating that the proc. is Stationary

i.e

~ The Normal distribution becomes downward bias, meaning that it tends to Reject the Unit roots even if the Unit root is present

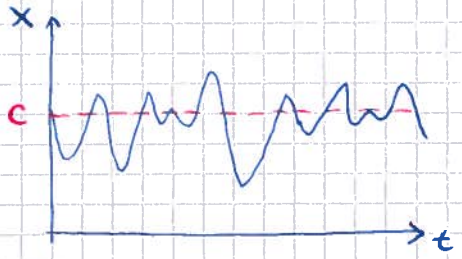
~ By Accepting that a process is Stationary, we istantly model it as a deterministic Time Trend, leading Forecasts always to be biased toward converging to the old trend.

~ In reality, a lot less Series are Stationary (by using the DF) meaning that they must be modeled as Random Walks + drifts and made Stationary by using first differences.

~~Conclusion~~

- F-Test ! DF_3 & DF_4

Consider the following case



DF_3

- let the process be modeled as an AR(1), we want to check whether the AR has a unit root or not.
  Visually, we have an issue ...
- IF $\phi < 1 \Rightarrow X_t = \phi X_{t-1} + \varepsilon_t$ but then the Series should converge, or move around, its unconditional expectation $\rightsquigarrow 0$ while its clear that the true Stationary Model would be: $X_t = c + \phi X_{t-1} + \varepsilon_t$
- IF $\phi = 1$ and we adopt $X_t = c + \phi X_{t-1} + \varepsilon_t$ we also have a problem as the model $X_t = c + X_{t-1} + \varepsilon_t$ would imply a trend, which is clearly not the case

i.e we want to Test $\underline{X_t = c + \phi X_{t-1} + \varepsilon_t}$ vs $\underline{X_t = X_{t-1} + \varepsilon_t}$

$\Rightarrow$ H_0 : $c = 0$, $\phi = 1$ unit root

H_1 : not H_0 Stationary Series.

- small Recap: Tests for Unit roots

idea: We have seen that the DF distribution is better to test for unit roots as it doesn't have the downward biasness of the Normal.

▷ We have 4 kinds of DF distributions!

1) $T(\hat{\phi} - \phi) \rightarrow DF_1$ with $\phi = 1$: H_0

2) $\dfrac{\hat{\phi} - \phi}{\sqrt{V(\hat{\phi})}} \rightarrow DF_2$ $H_0: \phi = 1$

↳ here: we have a single variable to test on, therefore we can use a normal t-test. The key idea is that the t-test is applied on the DF distribution.

Let us introduce DF_3 and DF_4 (F-tests)

- DF_3 (already seen):

$H_0: \phi = 1, \ \alpha = 0$

H_1: not H_0 $\rightarrow DF_3$

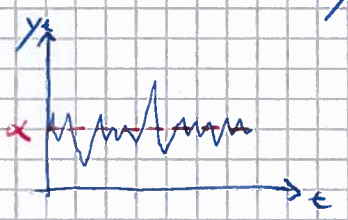
$$Y_t = \alpha + \phi Y_{t-1} + \varepsilon_t$$

OR

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

i.e we are testing whether we have:

for a pattern like:



- DF_4 (Trend)

$H_0: \phi = 1, \ \beta = 0$

H_1: not H_0 $\rightarrow DF_4$

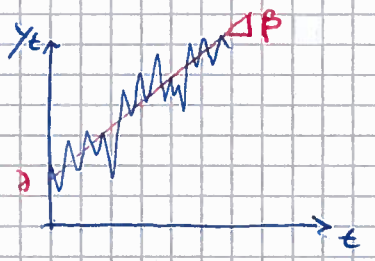
$$Y_t = \delta + \beta \cdot t + \phi Y_{t-1} + \varepsilon_t$$

OR

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

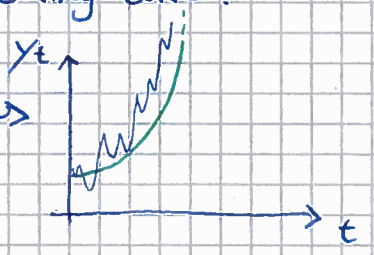
i.e we are testing whether we have:

for a pattern like:



notice: IF $\phi = 1$ and we allow $\beta > 0$ (or in general $\beta \neq 0$) then we get a process looking like:

Quadratic Trend ⇝



▷ Again, for DF_3 and DF_4 we have multiple variables in the Null, this requires the application of a F-test.

▷ Final Remark. as $T \to \infty$, the DF distribution will converge to a Normal, but they'll never exactly overlap.

- notice that the tests remain exactly the same when we consider first difference models.

e.g for DF_3, consider: $X_t = \alpha + \phi X_{t-1} + \varepsilon_t$

by subtracting X_{t-1} $\rightsquigarrow$ $\Delta X_t = \alpha + (1-\phi) X_{t-1} + \varepsilon_t$

Then, the test becomes:

H_0: $\alpha = 0$, $(1-\phi) = 0$

H_1: not H_0

e.g for DF_4: $X_t = \delta + \beta \cdot t + \phi X_{t-1} + \varepsilon_t$

$\rightsquigarrow$ $\Delta X_t = \delta + \beta \cdot t + (1-\phi) X_{t-1} + \varepsilon_t$

H_0: $\beta = 0$, $(1-\phi) = 0$

H_1: not H_0

Remark: when we start considering AR(p) instead of AR(1) matters become more complex $\rightsquigarrow$ i.e we need to capture the presence of a Unit root in the polynomial:

$\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \cdots \phi_p L^p)$

How to spot the Unit root among $\phi_1 \cdots \phi_p$?

$\Downarrow$

- Augmented DF Test.

notice that the polynomial $\phi(L)$ can be decomposed as:

$$* \quad \phi(L) = (1 - \rho L) - (\xi_1 L + \xi_2 L^2 + \cdots + \xi_{p-1} L^{p-1}) \cdot (1 - L)$$

where: $\rho = \phi_1 + \phi_2 + \phi_3 + \cdots + \phi_p$

$\xi_i = -[\phi_{i+1} + \cdots + \phi_p]$ with $i = 1, 2, \cdots p$

Claim: IF $\rho = 1$ $\Rightarrow$ There is a unit root in the polynomial $\phi(L)$

* Beveridge-Nelson decomposition

proof...

- Recall that the characteristic equation was:

$$\phi(z) = 0$$

where: if the equation is solved by $|z| > 1$ then the process is Stationary. On the other hand, if $|z| = 1 \Rightarrow$ we have a unit root!

i.e: IF $\phi(1) = 0 \Rightarrow$ unit root $\rightsquigarrow$ non-Stationary.

notice that:

$$\phi(1) = 1 - \phi_1 - \phi_2 \cdots - \phi_p$$

$$\phi(1) = 0 \quad \Rightarrow \quad 1 = \underbrace{\phi_1 + \phi_2 + \phi_3 + \cdots + \phi_p}_{\rho}$$

$\Rightarrow$ IF $\rho = 1 \Rightarrow$ U.R

Now, let us implement this decomposition of the AR polynomial into the process. Consider an AR(p): (with constant)

$$\phi(L) X_t = \delta + \varepsilon_t$$

$$\Rightarrow [(1 - \rho L) - (\xi_1 L + \xi_2 L^2 + \cdots + \xi_{p-1} L^{p-1}) \cdot (1 - L)] X_t = \delta + \varepsilon_t$$

$$[(1 - \rho L) - \xi(L)(1 - L)] X_t = \delta + \varepsilon_t$$

$$X_t - \rho X_{t-1} - \xi(L) X_t + \xi(L) X_{t-1} = \delta + \varepsilon_t$$

$$X_t = \rho X_{t-1} + \underbrace{\xi(L) \Delta X_t}_{\text{new term}} + \delta + \varepsilon_t$$

$\Rightarrow$ one U.R needs $d = 1$

↳ the polynomial $\xi(L)$ does not contain the U.R just as long there is only one in the entire process (in the original $\phi(L)$)

Test for Unit Roots in the AR(p) process becomes:

H_0: $\rho = 1$ + addit. given the deterministic components

H_1: not H_0 ↳ e.g $\delta = 0$, $\beta = 0$

Remarks.

1. the same thing hold when testing on ΔX_t, as before $\rightsquigarrow$ H_0: $(\rho - 1) = 0$

2. Key Ass: There is only one Unit Root but we don't know where

- **Alternative Test : KPSS Test**

idea: Instead of having the U.R as the Null Hypothesis, we let Stationarity be the H_0.

Consider : $X_t = \alpha + \beta \cdot t + \phi X_{t-1} + \varepsilon_t$

we then run the regression: $X_t = \hat{\alpha} + \hat{\beta} \cdot t + \hat{\varepsilon}_t$

notice : $\hat{\varepsilon}_t$ captures the behavior of ϕX_{t-1}

let $S \equiv \sum_{t=1}^{T} \hat{\varepsilon}_t$ ⇝ variance of $\hat{\varepsilon}_t$

IF S is small → Stationarity

IF S is Big → Unit Root

Here : H_0 : Stationarity

↳ We reject the Null IFF S is big enough.

In this Test we use the statistic η, where:

$$\eta \equiv \frac{1}{T^2 S^2} \cdot \sum^{T} S_t^2 \quad \text{where} \quad S_t = \sum^{t} \varepsilon_i$$

IF η is too big ⇒ Reject the Null.

- **Additional issues related to the presence of Unit Roots : Spurious Regressions.**

Consider two independent U.R processes:

1) $X_t = X_{t-1} + \varepsilon_t^x$
2) $Y_t = Y_{t-1} + \varepsilon_t^y$

with $cov(\varepsilon_t^x, \varepsilon_{t-s}^y) = 0 \ \forall s$

$\varepsilon^x \perp \varepsilon^y$

Say that we were interest in spotting some relationship between Y_t and X_t, in order to do so we run the following regression:

$$Y_t = \alpha + \beta \cdot X_t + \varepsilon_t$$

IF the two process are independent we should expect:

$\hat{\beta} = 0$ and $R^2 = 0$

It turns out that this is not the case, why?

⇒ The Residuals are not Stationary

notice: $\varepsilon_t = Y_t - \beta X_t$ (let $\alpha = 0$ for simplicity)

$$= \sum_{i=0}^{\infty} \varepsilon_i^y - \beta \sum_{i=0}^{\infty} \varepsilon_i^x \quad (AR(1) \Rightarrow MA(\infty))$$

↳ but $\phi = 1$

$$= \sum_{i=0} \varepsilon_i^y - \beta \varepsilon_i^x$$

⇒ Non-Stationary

notice that having Stationary residuals is one of the main assumptions for OLS. If this is not the case the relation between y_t and X_t becomes spurious and β is biased. ?

How do we adjust for this?

idea: Include the process in the regression and instead run:

$$y_t = \alpha + \beta X_t + \gamma X_{t-1} + \delta y_{t-1} + \varepsilon_t$$

If the data generating process is $y_t = y_{t-1} + \varepsilon_t$ then the OLS will minimize the squared residuals only in setting $\alpha = 0$ $\beta = 0$ $\gamma = 0$ $\delta = 1$

i.e OLS replicates the data generating process.

- **Beveridge - Nelson Decomposition**

idea: We want to decompose a non-Stationary process into a Trend and into a Cycle

Consider a non-Stationary process (Unit Root)

↳ $\Phi(L)$ contains a U.R

We can take the difference to get a Stationary process:

$$\Delta X_t = \Psi(L) \cdot \varepsilon_t$$

and write the Stat. process in its Wold decomposition.

CLAIM: $\Psi(L) = \underset{p}{\Psi(1)} + \underset{p-1}{\Psi^*(L)} \cdot \underset{1}{(1-L)}$ ↳ of order (order of the poly)

PF.

denote $A(L) \equiv \Psi(L) - \Psi(1)$

notice that, by definition $A(1) = \Psi(1) - \Psi(1) = 0$

and $A(1) = 0$ ⇒ we have a Unit Root in $A(L)$

now, $A(L)$ can be rewritten as:

$A(L) = \Psi^*(L) \cdot (1-L)$ → Unit Root Component

Stationary poly.

↳ see that the same kind of decomposition was carried previously.

⇒ by def. of $A(L)$ ⇒ $\Psi(L) - \Psi(1) = \Psi^*(L) \cdot (1-L)$

⇒ $\Psi(L) = \Psi(1) + \Psi^*(L) \cdot (1-L)$ ✓

Therefore we have:

$$\Delta X_t = \Psi(1) \cdot \varepsilon_t + \Psi^*(L)(1-L) \cdot \varepsilon_t$$

From : $\Delta X_t = \Psi(1) \cdot \varepsilon_t + \Psi^*(L)(1-L) \cdot \varepsilon_t$ eq.1

multiply by $(1-L)^{-1}$

What is $(1-L)^{-1}$? Consider $y_t = y_{t-1} + \varepsilon_t$

Recall that $y_t = y_0 + \sum^t \varepsilon_i$ (we can ignore y_0)

or alternatively
$$y_t - y_{t-1} = \varepsilon_t$$
$$y_t (1-L) = \varepsilon_t$$
$$y_t = (1-L)^{-1} \cdot \varepsilon_t$$

notice: since we are assuming that $\phi = 1$, the polynomial $(1-L)$ cannot be inverted! ⇝ mathematically $(1-L)^{-1}$ has no meaning. But the idea is that:

$$* \quad (1-L)^{-1} \cdot \varepsilon_t \Leftrightarrow \sum^t \varepsilon_i$$

we can denote $(1-L)^{-1}$ as Δ^{-1} since:

$$y_t - y_{t-1} = \varepsilon_t \rightsquigarrow \Delta y_t = \varepsilon_t \rightsquigarrow y_t = \Delta^{-1} \cdot \varepsilon_t$$

Back to equation 1...

$$\Delta^{-1} \Delta X_t = (\Psi(1) \cdot \varepsilon_t + \Psi^*(L)\overbrace{(1-L)}^{\Delta} \varepsilon_t) \cdot \Delta^{-1}$$

$$X_t = \Psi(1)(1-L)^{-1} \cdot \varepsilon_t + \Psi^*(L) \cdot \varepsilon_t$$

notice that $(1-L)^{-1} \cdot \varepsilon_t = \sum^t \varepsilon_i$ *

$$\Rightarrow \quad X_t = \Psi(1) \cdot \sum^t \varepsilon_i + \Psi^*(L) \cdot \varepsilon_t$$

where:

$\sum^t \varepsilon_i$ is simply a Random Walk ⇝ $S_t = S_{t-1} + \varepsilon_t$

↳ by substituting $S_{t-1} \rightsquigarrow \sum^t \varepsilon_i$

$$\Rightarrow \quad X_t = \underbrace{\Psi(1) \cdot S_t}_{} + \underbrace{\Psi^*(L) \cdot \varepsilon_t}_{} + X_0$$

- Unit Root part
- it basically represent the "trend" part of the process.
- the polynomial $\Psi(1)$ indicates the relevance of the non-stationary component:

  IF $\Psi(1) = 0$ ⇝ $X_t = \Psi^*(L) \cdot \varepsilon_t + X_0$ ⇝ since $\Psi^*(L)$ is the stationary component ⇒ X_t is already stationary.

- Stationary part: it should represent the "cycle" part which moves around the trend.

Remark

- There are infinite ways in which we can separate the process into a trend and a cycle. But, Beveridge-Nelson assumes that the cycle and the trend are driven by the same shocks : ε_t

Example of a B-N decomposition.

Consider a process in which we have already taken differences to make it stationary:

$$\Delta y_t = \underbrace{(1+\theta L)}_{\psi(L)} \cdot \varepsilon_t$$

Recall: $\psi(L) = \psi(1) + (\psi_0^* + \psi_1^* L + \psi_2^* L^2 + \cdots)(1-L)$

So, we have :

$$(1+\theta L) = (1+\theta) + (\psi_0^* + \psi_1^* L + \psi_2^* L^2 + \cdots)(1-L)$$

how do we determine $\{\psi_0^*, \psi_1^*, \cdots\}$?

↳ Recursively :

no L 1. elements with zero lag : $1 = 1 + \theta + \psi_0^* \Rightarrow \psi_0^* = -\theta$

L 2. first lag : $\theta = -\psi_0^* + \psi_1^* \Rightarrow \psi_1^* = \psi_0^* + \theta = 0$

L^2 3. second lag : $0 = -\psi_1^* + \psi_2^* \Rightarrow \psi_1^* = \psi_2^* = 0$

L^3 ⇝ $0 = -\psi_2^* + \psi_3^* \Rightarrow \psi_2^* = \psi_3^* = 0$

L^4 ⇝ $0 = -\psi_3^* + \psi_4^* \Rightarrow \psi_3^* = \psi_4^* = 0$

and so on...

So, $(1+\theta) + (\psi_0^* + \psi_1^* L + \psi_2^* L^2 + \cdots)(1-L)$

$= (1+\theta) - \theta(1-L)$

$$\Rightarrow \quad \Delta y_t = [(1+\theta) - \theta(1-L)]\,\varepsilon_t$$

Δ^{-1} $\qquad$ $\parallel$ $\qquad$ Δ^{-1}

$$\Rightarrow \quad y_t = \underbrace{(1+\theta)\cdot\sum^{t}\varepsilon_i}_{\text{Unit Root.}} - \theta\varepsilon_t + y_0$$

- **Multivariate Analysis**

idea: instead of having scalar AR(p) process, we now have vectors of AR $\Rightarrow$ VAR Models with $\{\underset{n\times 1}{X_t}\}_{-\infty}^{+\infty}$

- **Covariance - Stationary** (definition)

1) $E(\underset{n\times 1}{X_t}) = \underset{n\times 1}{\mu}$

2) $E[(\underset{n\times 1}{X_t} - \mu)(\underset{1\times n}{X_{t-s}} - \mu)'] = \underset{n\times n}{\Gamma_s}$ $\Rightarrow$ Variance - Covariance matrix

lag.

let's look at Γ_s :

$$\Gamma_s = \underset{n\times n}{\begin{bmatrix} \gamma_s^1 & \square & & \square \\ \square & \gamma_s^2 & & \\ & & \gamma_s^3 & \\ & & & \ddots \\ \square & & & \gamma_s^n \end{bmatrix}}$$

$\square \to cov(X_t^1, X_{t-s}^2)$

$\square \to cov(X_t^1, X_{t-s}^n)$

$\square \to cov(X_{t-s}^1, X_t^2)$

$\square \to cov(X_{t-s}^1, X_t^n)$

where :

- γ_s^i = covariance of the same variable at two periods with lag s $\Rightarrow$ $cov(X_t^1, X_{t-s}^1)$
- The matrix is not symmetric. $\Rightarrow$ $\underline{\Gamma_s \neq \Gamma_s'}$
- symmetry only for Γ_0 $\Rightarrow$ $\underline{\Gamma_0 = \Gamma_0'}$
- In general $\underline{\Gamma_s = \Gamma_{-s}'}$

- **Vector White Noise**

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

2) $E(\varepsilon_t \varepsilon_t') = V(\varepsilon_t) = \underset{n\times n}{\Sigma}$

3) $E(\varepsilon_t \cdot \varepsilon_{t-k}') = 0$ for $k \neq 0$

- **Vector Autoregression $\Rightarrow$ VAR**

$$\underset{n\times 1}{X_t} = \underset{n\times 1}{c} + \underset{n\times n}{\phi_1}\,\underset{n\times 1}{X_{t-1}} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$$

idea: X_t^1 (first element of the vector X_t) is not just auto-regressed on X_{t-1}^1 etc.., but also on $X_{t-1}^2 \cdots, X_{t-1}^3 \cdots, \cdots, X_{t-1}^n$.

So, the term $\phi_1 X_{t-1}$ is equal to:

$$\underset{n\times 1}{\begin{bmatrix} \phi_{11} X_{t-1}^1 + \phi_{12} X_{t-1}^2 + \phi_{13} X_{t-1}^3 + \cdots + \phi_{1n} X_{t-1}^n \\ \phi_{21} X_{t-1}^1 + \phi_{22} X_{t-2}^2 + \phi_{23} X_{t-1}^3 + \cdots + \phi_{2n} X_{t-1}^n \\ \vdots \\ \phi_{n1} X_{t-1}^1 + \cdots + \phi_{nn} X_{t-1}^n \end{bmatrix}} = \underset{n\times n}{\phi_1}\,\underset{n\times 1}{X_{t-1}}$$

Q: How many parameters do we have in total?

- In ϕ_i we have n^2 parameters
- In $\phi_1 \cdots \phi_p$ then the total is $n^2 \cdot p$
- We also have the Variance-Covariance matrix Σ with $\frac{n \cdot (n+1)}{2}$ parameters.

In total: $\underset{c}{n} + \underset{\phi_1 \cdots \phi_p}{n^2 \cdot p} + \underset{\Sigma}{n \cdot (n+1)/2}$

- **Stationarity**

Consider a VAR(1) model: $X_t = c + \phi X_{t-1} + \varepsilon_t$

idea: Similiarly to the AR process, stationarity can be assesed either by looking at the characteristic equation or by looking at the Eigenvalues.

① **Characterist equation**: $\phi(z)$

Before: Stat. IFF $\phi(z) = 0$ for $|z| > 1$

↳ in an AR(1) ⇝ $\phi(z) = (1 - z\phi) = 0$

Now: For a VAR(1):

Stat. IFF $\det(I - \phi z) = 0$ for $|z| > 1$

notice that z is not a scalar anymore, but (potentially) a $n \times 1$ vector, so each element of z need to fulfill the condition.

e.g for a VAR(2): $\det(I - \phi_1 z - \phi_2 z^2) = 0 \quad |z| > 1$

↳ in this case we have $2n$ roots.

② **Write down the Companion Form and check the Eigenvalues.**

Consider a VAR(2) in Comp. Form:

$$\underset{2n \times 1}{\begin{bmatrix} \underline{X}_t \\ \underline{X}_{t-1} \end{bmatrix}} = \underset{2n \times 2n}{\begin{bmatrix} \phi_1 & \phi_2 \\ I & 0 \end{bmatrix}} \cdot \underset{2n \times 1}{\begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix}} + \underset{2n \times 1}{\begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}}$$

$$\downarrow$$

$$A$$

- Check Stationarity by making sure that all the Eigenvalues of A lie within the Unit Circle ⇒ i.e < 1
- notice that when ϕ is a $n \times n$ matrix, it doesn't make sense to impose $\phi < 1$ ⇝ we need to look at Eigenvalues.

- Recall that the matrix A can be rewritten as the product of its Eigenvectors and Eigenvalues, using the Spectral decomposition :

$$A = B \Lambda B^{-1}$$

matrix of Eigenvectors

Eigenvalues

- **Impulse Response**

Similiarly to the scalar case, for a VAR(1)

$$X_t = \phi X_{t-1} + \varepsilon_t$$

let there be a shock $\underset{n \times 1}{A}$ at time zero

First period $(t=1)$: ϕA

$t = 2$: $\phi^2 A$

$t = 3$: $\phi^3 A$

so on . . .

- **Estimation of VAR**

Consider a very simple VAR Model :

$$\begin{cases} y_t = a\, y_{t-1} + b\, x_{t-1} + \varepsilon_t^y \\ x_t = c\, y_{t-1} + d\, x_{t-1} + \varepsilon_t^x \end{cases}$$

with $V(\varepsilon_t) = \Sigma$

notice : The VAR above is also called "Seemingly unrelated regressions" (SUR). The idea is that ε_t^x and ε_t^y might be correlated and estimation through the OLS might appear to be bias.

Apparent Solution : GLS estimation

Actually, it has been proved that when the regressors are exactly the same, the OLS is not biased :

Regress X_t on X_{t-1} and y_{t-1}

Regress Y_t on X_{t-1} and y_{t-1}

$\Rightarrow$ We can also use OLS to estimate a, b, c, d

$\Rightarrow$ the only choice is : the number of lags.

- **Diagnostic Checks** (same as scalar case)

↳ Information Criteria

1) AIC : $T \cdot \ln|\hat{\Sigma}| + 2(T^2 \cdot p)$

2) BIC : $T \cdot \ln|\hat{\Sigma}| + (T \cdot p) \ln T$

↳ Shwartz-Bayesian I.C

⇝ once our model has been estimated and checked, we can move to Forecasting.

- **Forecasting**

Consider a VAR(p) : $y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdot + A_p y_{t-p} + \varepsilon_t$

clearly :

$$E(y_{t+1} \mid I_t) = y_{t+1|t} = c + A_1 \underbrace{y_{t|t}}_{y_t} + A_2 \underbrace{y_{t-1|t}}_{y_{t-1}} + \cdots + \underbrace{\cancel{\varepsilon_{t+1|t}}}_{0}$$

Forecast error : ε_{t+1}

Remember, in general
- ↗ Forecast contains info up to t
- ↳ Forecast error cont. info after t

Forecast err variance : $V(\varepsilon_{t+1}) = \Omega$ ⇝ var-cov matrix

2-step Forecast :

- $y_{t+2|t} = c + A_1 \underbrace{y_{t+1|t}}_{\text{replace with previous forecast.}} + A_2 y_{t|t} + \cdots$
- error : $A_1 \varepsilon_{t+1} + \varepsilon_{t+2}$
- error variance : $A_1 \cdot \Omega \cdot A_1' + \Omega$

by iterating :

Forecast converges to Unconditional Expectation

Error Variance converges to Unconditional Variance → $V(y_t)$

Consider the same case for a VMA : $y_t = \Psi(L) \cdot \varepsilon_t$

1-step : $y_{t+1|t} = \underbrace{\varepsilon_{t+1|t}}_{\times} + \underbrace{\varepsilon_{t|t}}_{\varepsilon_t} \cdot \underbrace{\Psi_1}_{\Psi_1} + \cdots$

error : ε_{t+1}

error variance : $V(\varepsilon_{t+1}) = \Omega$

2-step : $y_{t+2|t} = \underbrace{\varepsilon_{t+2|t}}_{\times} + \Psi_1 \underbrace{\varepsilon_{t+1|t}}_{\times} + \underbrace{\Psi_2}_{\Psi_2} \underbrace{\varepsilon_{t|t}}_{\varepsilon_t} + \cdots$

error : $\varepsilon_{t+2} + \Psi_1 \cdot \varepsilon_{t+1}$

error variance : $\Omega + \Psi_1 \Omega \Psi_1'$

- Again, convergence is to the Unconditional moments.
- error variance for the VAR and VMA is exactly the same.

- **Wold Decomposition for a VAR**

Any ~~stat~~ Covariance - Stationary Stochastic process can be written as:

$$y_t = \Psi(L) \cdot \varepsilon_t$$

s.t $\varepsilon_t \equiv y_t - Proj(y_t | y_{t-1}, y_{t-2} \dots)$

and $\Psi(L) = I - \Psi_1 L - \Psi_2 L^2 - \dots$

notice: The Wold decomposition is essentially the VAR form, this tells us that a ~~MVA~~ VMA can always be rewritten as a VAR (Through the Wold decomposition).
i.e given an Autocovariance structure* we can model it either by VAR or ~~MVA~~ VMA (which are equivalent through the Wold decomposition) but it is not implying that the two processes will yield the same path of residuals $\{\varepsilon_t\}_{-\infty}^{+\infty}$. The latter depends on invertibility.

- **Invertibility of MVA**

def. A ~~MVA~~ VMA of the form $y_t = C(L) \cdot \varepsilon_t$ is Invertible IFF $\exists \{A_j\}_{j=0}^{\infty}$ s.t:

$$\sum_{j=0}^{\infty} A_j y_{t-j} = \varepsilon_t$$

We can check for Invertibility by using the same intuition as for the scalar case:

$\det(C(z)) = 0$ when $|z| > 1$ outside the unit circle

or altern. $\det(C(z)) \neq 0$ when $|z| \leq 1$ inside the unit circle.

notice: let $y_t = C(L) \cdot \varepsilon_t$ be an invertible VMA, then $\exists$ a VAR $\tilde{\varepsilon}_t = A(L) y_t$ s.t:

$$\varepsilon_t = \tilde{\varepsilon}_t$$

intuition:

An invertible VMA can be rewritten as a VAR (in its Wold decomposition) with the same shocks $\varepsilon_t \ \forall t$

e.g. Checking the invertibility of a VMA

Consider:

VMA
$$\begin{bmatrix} y_t^1 \\ y_t^2 \end{bmatrix} = \underbrace{\overset{A}{\begin{bmatrix} 1 & L \\ 0 & \theta - L \end{bmatrix}}}_{\text{check determinant}} \cdot \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \rightsquigarrow \begin{aligned} y_t^1 &= \varepsilon_{t-1}^2 + \varepsilon_t^1 \\ y_t^2 &= \theta \varepsilon_t^2 + \varepsilon_{t-1}^2 \end{aligned}$$

$$\det(A) = \theta - L$$

then replace L with z and set equal to zero:

$$\theta - z = 0 \Rightarrow \theta = z$$

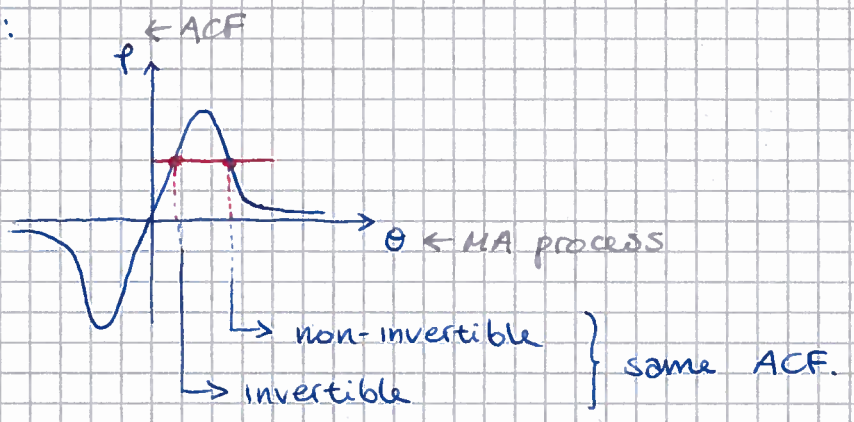
Invertibility for $|z| > 1 \Rightarrow |\theta| > 1$

The inverted process is:

VAR
$$\begin{bmatrix} 1 & -L/\theta - L \\ 0 & 1/\theta - L \end{bmatrix} \begin{bmatrix} y_t^1 \\ y_t^2 \end{bmatrix} = \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix}$$

notice: If the VMA process is not invertible, we can still obtain a VAR describing the same ACF, BUT the output residuals won't be the same.

Recall:



When deriving the Wold decomposition of a non-invertible VMA, what we are doing is to invert the invertible counterface of the original VMA to obtain the VAR.

OR, equivalently, by inverting the Wold decomposition of a non-invertible VMA, we obtain the invertible counterface.

o Granger Causality

def. A variable X_t "Granger Cause" y_t IF past values of X_t help predicting y_t, given the past of y_t

idea: how to we spot Granger Causality?

intuitive way

Consider :

$$X_t = a(L) X_{t-1} + b(L) y_{t-1} + \varepsilon_t^x$$
$$y_t = c(L) X_{t-1} + d(L) y_{t-1} + \varepsilon_t^y$$

IF there is NO G.C then by running an F-test on $C(L)$ with $H_0: C(L) = 0$ we should accept the Null.

Now, a key result can be obtained by first assuming that X_t does not G.C y_t. In this case we have $C(L) = 0$, rewriting:

$$\begin{cases} X_t = a(L) X_{t-1} + b(L) y_{t-1} + \varepsilon_t^x \\ y_t = \quad 0 \quad\quad + d(L) y_{t-1} + \varepsilon_t^y \end{cases}$$

In VAR Form.

$$\begin{bmatrix} X_t \\ y_t \end{bmatrix} = \begin{bmatrix} a(L) & b(L) \\ 0 & d(L) \end{bmatrix} \begin{bmatrix} X_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix}$$

$$\begin{bmatrix} 1-a(L) & \tilde{b}(L) \\ 0 & 1-d(L) \end{bmatrix} \begin{bmatrix} X_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix}$$

Assume Stationarity and obtain the corresponding VMA:

$$\begin{bmatrix} X_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1-a(L) & \tilde{b}(L) \\ 0 & 1-d(L) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix}$$

$$\begin{bmatrix} X_t \\ y_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{21}(L) \\ 0 & C_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix}$$

So, we can have a second definition of G.C, given by Sims.

Sims 1972 ~ G.C

def. A variable X_t does not "G.C" Y_t IF

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{21}(L) \\ 0 & C_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t^X \\ \varepsilon_t^Y \end{bmatrix}$$

where: - ε_t^X and ε_t^Y are white noise
- $\varepsilon_t^X \perp \varepsilon_s^Y \quad \forall t,s$ (strong independence)

From this definition: $\begin{cases} X_t = C_{11}(L) \cdot \varepsilon_t^X + C_{21}(L) \cdot \varepsilon_t^Y \\ Y_t = C_{22}(L) \cdot \varepsilon_t^Y \end{cases}$

notice that: $C_{22}(L)^{-1} \cdot Y_t = \varepsilon_t^Y$

by substituting:

$$X_t = \underbrace{C_{11}(L) \cdot \varepsilon_t^X}_{\text{Residuals}} + \underbrace{C_{21}(L) \cdot C_{22}(L)^{-1} \cdot Y_t}_{\text{Regressors}}$$

$\perp$

Key result: we can regress X_t on the past and present of Y_t using OLS. This is because we proved that the residuals are $\perp$ to the regressors.

notice: if we regress X_t on the past, present, future values of Y_t, we get:

$$X_t = \sum_{j=-\infty}^{+\infty} b_j Y_{t-j} + V_t$$

BUT by absence of G.C we should get that:
$b_j = 0 \quad \forall j < 0 \rightsquigarrow$ Future Y_t can't be correlated with present X_t.

So, test for G.C is running an F-test on the b_j's and checking: $H_0: b_j = 0 \quad \forall j < 0$

First Order Taylor Expansion tells us:

$$\boxed{\underbrace{F(v_t, \omega_t)}_{} \approx \underbrace{F(v, \omega)}_{} + \underbrace{\frac{\partial F}{\partial \tilde{v}_t} \cdot \tilde{v}_t}_{} + \underbrace{\frac{\partial F}{\partial \tilde{\omega}_t} \cdot \tilde{\omega}_t}_{}}$$

now notice:

$U_t \approx U \quad +$

note that:

∂F

- Applications of Macroeconometrics

Quick History of the evolution:

1963: Friedman - Schwartz: "Monetary History of the US"
main idea: every time the FED changes the Money supply (ΔM) there is a recession ($\Delta y \downarrow$)
notice: the result was purely based on a qualitative analysis of the TS of M and y.
At the time they had no VAR

1970: St. Louis Equations
essentially: $\Delta y_t = \alpha + \beta(L) \cdot \Delta M_t + \gamma(L) \cdot \Delta y_{t-1} + v_t$ eq.1

idea: by running eq.1 we should spot the relation between Δy_t and ΔM_t

Major problem: the validity of eq.1 is threatened by a possible endogeneity of Monetary policy
i.e Monetary policy (ΔM) might be responding to Δy
s.t:

$$\Delta M = \lambda(L) \Delta y_{t-1} + \omega_t \quad \text{eq.2}$$

IF Monetary policy is indeed endogenous then there is a feedback relationship between ΔM and Δy_{t-1} in eq.1
$\Rightarrow$ OLS is clearly biased.
Sim $\rightarrow$ let us verify whether Δy_{t-1} Granger Cause ΔM
$\hookrightarrow$ some support on the view that there is no Granger Causality.

- Limits of G.C

1) G.C does not spot real causality but more like a lagged correlation, e.g:
"Xmas cards G.C Xmas" $\rightsquigarrow$ TRUE

2) Results (implications on G.C) strongly depend on the variables set. (variables included in the VAR)

3) Forecastability cannot be used by policymakers since causation mechanism is flawed
AKA "POST HOC, ERGO PROPTER HOC" Fallacy

Now we move to a central issue in Macroeconometrics which Sim (1980) helped to resolve by introducing VAR Models.

- Identification

Key paper: Sim (1980) "Macroeconomics and Reality"

Before the 80's: Simultaneous equations Models

- large systems of equations taken simultaneously and with some level of economic intuition.

e.g $C = C_0 + C_1 (y - T)$ ⇝ Consumption eq.

* to allow estimability of parameters, 2 specifics were included:

1) Some variables were determined inside the model (Endogenous) and other outside the model (Exogenous)
2) Not all lags were included: Apriori exclusion of lagged variables.

* to see why we need to consider a simple version of a Macroeconometric model before Sim:

Consider a SINGLE MARKET MODEL in its Structural Form:

$$q_t^d = \alpha_1 p_t + \alpha_2 y_t + \varepsilon_t^d$$

$$q_t^s = \beta_1 p_t + \varepsilon_t^s$$

$$q_t^s = q_t^d = q_t$$

+ usual assumptions: $E(\varepsilon_t^d) = E(\varepsilon_t^s) = 0$

$V(\varepsilon_t^d) = \sigma_d^2$; $V(\varepsilon_t^s) = \sigma_s^2$

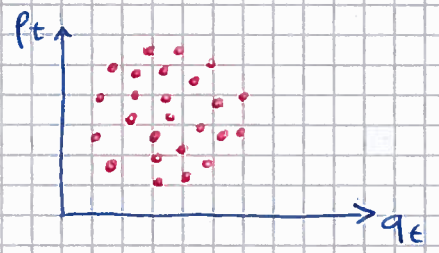
$E(\varepsilon_t^d \cdot \varepsilon_{t-h}^s) = 0$

where: q_t and p_t are the Endogenous variables.

y_t is the Exogenous variable.

Intuitively, our objective is to determine α_1, α_2 and β_1 such that we can derive both the Supply and Demand curve...

From data we might get points describing equilibrium prices and quanties s.t :



For simplicity let $\underline{y=0}$ s.t the Structural form is:

$$q_t = \alpha_1 p_t + \varepsilon_t^d \qquad \text{demand curve}$$
$$q_t = \beta_1 p_t + \varepsilon_t^s \qquad \text{supply curve}$$

notice: a key assumption to estimate both equations is to know that we have Exogeneity:

$$p_t \perp \varepsilon_t^d$$
$$p_t \perp \varepsilon_t^s$$

Is this the case?

Let us rewrite the Structural form into its Reduced Form:

$$p_t = \frac{\varepsilon_t^d - \varepsilon_t^s}{\beta_1 - \alpha} = v_{p_t} \quad \text{and} \quad q_t = \frac{\beta_1 \varepsilon_t^d - \alpha_1 \varepsilon_t^s}{\beta_1 - \alpha_1}$$

Clearly, it is not the case that $p_t \perp \varepsilon_t \Rightarrow$ <u>OLS is biased!</u>
If we run the regression anyway, we get:

$$\alpha_1 = \beta_1 = \frac{cov(p_t, q_t)}{V(p_t)}$$

where, $$cov(p_t, q_t) = \frac{\beta_1 \sigma_d^2 + \alpha_1 \sigma_s^2}{(\beta_1 - \alpha_1)^2}$$

$$V(p_t) = \frac{\sigma_s^2 + \sigma_d^2}{(\beta_1 - \alpha_1)^2}$$

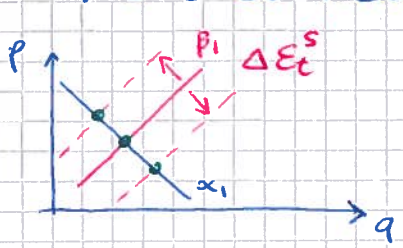
$$\Rightarrow \text{OLS estimator} = \frac{\sigma_d^2}{\sigma_d^2 + \sigma_s^2} \beta_1 + \frac{\sigma_s^2}{\sigma_d^2 + \sigma_s^2} \cdot \alpha_1$$

Supply slope Demand slope.

$\Rightarrow$ So, the OLS will estimate a weighted average of the Supply and Demand slopes. This is called: Simultaneous Equations bias

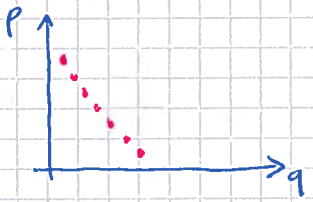
The story drastically changes when we try to impose restrictions on the parameters.

- e.g 1 let $\varepsilon_t^d = 0$ ⇝ no variance on the demand curve
  ⇒ OLS estimator becomes α_1 ⇝ not bias.



⇝ from shocks to the supply curve only we get equilibrium (q, p) (green dots) from which it's easy to estimate the slope α_1.

i.e we would observe data looking like:



e.g 2 let $\varepsilon_t^s = 0$ ⇝ OLS estim. is β_1
i.e reversed case above.

intuition: why do we need to impose restrictions?

Unrestricted model has:

- ~ 2 dependent variables : p, q
- ~ 4 indep. variables: α_1 β_1 σ_s^2 σ_d^2

problem is that there are infinite combinations of the 4 ind. variables that yield p and q.

Only solution is to impose restrictions.

- Consider now a restricted version of the original Single market Model

STRUCTURAL FORM:

$$p_t + \varepsilon_t^s = 0 \qquad \text{(no q in Supply)}$$
$$q_t = \alpha_1 p_t + \varepsilon_t^d$$

REDUCED FORM:

$$p_t = -\varepsilon_t^s \qquad S$$
$$q_t = -\alpha_1 \varepsilon_t^s + \varepsilon_t^d \qquad D$$

notice: $p_t \perp \varepsilon_t^d$ ⇒ I can OLS estimate the demand curve.

- no restriction, but introduction of an exogenous variable affecting only one curve

STRUCTURAL:

$$q_t = \alpha_1 p_t + \alpha_2 y_t + \varepsilon_t^d \qquad \text{Demand}$$

$$q_t = \beta_1 p_t + \varepsilon_t^s \qquad \text{Supply}$$

REDUCED:

$$p_t = \frac{\alpha_2}{\beta_1 - \alpha_1} y_t + \frac{\varepsilon_t^d - \varepsilon_t^s}{\beta_1 - \alpha_1} = \boxed{\pi_1 y_t + v_t^p}$$

$$q_t = \frac{\beta_1 \alpha_2}{\beta_1 - \alpha_1} y_t + \frac{\beta_1 \varepsilon_t^d - \alpha_1 \varepsilon_t^s}{\beta_1 - \alpha_1} = \boxed{\pi_2 y_t + v_t^q}$$

notice: I can run the OLS for:

$$\begin{cases} p_t = \pi_1 y_t + v_t^p \\ q_t = \pi_2 y_t + v_t^q \end{cases}$$

Ortogonality Condition

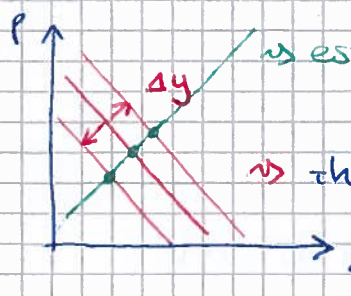
since $y_t \perp v_t^q, v_t^p$

$\Rightarrow$ I obtain OLS estimates for $\hat{\pi}_1$ and $\hat{\pi}_2$

From $\hat{\pi}_1$ and $\hat{\pi}_2$ we can estimate the indirect least square estimate for β_1, by noticing that:

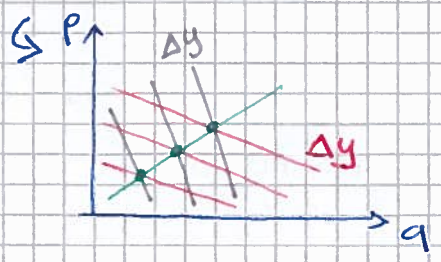
$$\hat{\beta}_1 = \frac{\hat{\pi}_2}{\hat{\pi}_1}$$

intuition: when we have shocks in y, the demand curve shifts but the supply curve actually remains fixed. The resulting equilibrium p & q will map the slope of the supply curve:



$\rightsquigarrow$ estimated Supply curve

$\rightsquigarrow$ the slope of the demand actually remains unknown, but it doesn't matter



notice: We are able to estimate the slope of the supply (β_1) given that $y_t \perp \varepsilon_t^s$

i.e when y_t is subject to a shock, the supply curve stays fixed.

Analytical proof:

given $y_t \perp \varepsilon_t^s \Rightarrow E(y_t \cdot \varepsilon_t^s) = 0$

$$\Rightarrow E(y_t \cdot (q_t - \beta_1 p_t)) = 0$$

$$\Rightarrow E(y_t \cdot q_t) - \beta_1 E(y_t \cdot p_t) = 0$$

$$\Rightarrow \beta_1 = E(y_t \cdot p_t)^{-1} \cdot E(y_t \cdot q_t)$$

By assumption, we have also $y_t \perp \varepsilon_t^d$, does this mean that in principle we should be able to estimate α_1? NO

proof.

$$E(y_t \cdot \varepsilon_t^d) = 0$$

$$E(y_t \cdot (q_t - \alpha_1 p_t - \alpha_2 y_t)) = 0$$

2 unknowns, one eq.

↳ we are not able to estimate α_1 or α_2

idea: we need to impose an Exclusion Restriction (i.e y does not enter the Supply equation) to be able to estimate the Structural parameters.

Now, consider the same kind of problem in matrix notation (one step forward in generalizing the result)

STRUCT. FORM : $B\, y_t + \Gamma x_t = u_t$

y_t → endogenous, x_t → exogenous OR pre-determined vs lags

deriving the reduce form: $y_t = \underbrace{-B^{-1} \cdot \Gamma}_{\Pi} x_t + \underbrace{B^{-1} \cdot u_t}_{V_t}$

REDUCED FORM $y_t = \Pi x_t + V_t$

where : $\Pi = -B^{-1} \cdot \Gamma$ & $V_t = B^{-1} \cdot u_t$

notice: $V(V_t) = \Omega = B^{-1} \cdot \Sigma B^{-1'}$

General Statement

The Structural parameters of eq. jth (jth row of the system generated by the matrix form) are IDENTIFIED IF given $\{\Pi, \Omega\}$ we can uniquely recover B_j and Γ_j

The Conditions to do so are:

1) **ORDER CONDITION**

number of exogenous variables excluded (notice *)

$\geq$

number of endogenous variables included

2) **RANK CONDITION**

The instruments must contain non-redundant info.

Let's understand this...

Consider the STRC. FORM Again:

$$B \cdot y_t + \Gamma x_t = u_t$$

notice that: $y_t' \neq (y_0, y_1 \cdots y)$

$$y_t' = [\underset{n_0 \times 1}{y_{0t}'} \,\vdots\, \underset{n_1 \times 1}{y_{1t}'} \,\vdots\, \underset{n_2 \times 1}{y_{2t}'}] \quad \text{where } \underset{n_i \times 1}{y_{it}}$$

$$x_t' = [\underset{m_1 \times 1}{x_{1t}'} \,\vdots\, \underset{m_2 \times 1}{x_{2t}'}]$$

↳ Partitioning

just an example

Consider the case where are interested in identifying the Structural parameters of equation zero. Here, y_{2t} are going to be the endogen. variables EXCLUDED ⇝ y_{1t} and y_{0t} are the included endog. variables in equation zero. On the other hand x_{2t} will be the EXCLUDED exogenous variables. Then, the system of equations is:

$$\begin{bmatrix} 1 & B_{01} & 0 \\ B_{10} & B_{11} & B_{12} \\ B_{20} & B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} y_{0t} \\ y_{1t} \\ y_{2t} \end{bmatrix} = - \begin{bmatrix} \Gamma_{01} & 0 \\ \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + u$$

& equation zero: $\underset{1 \times n_0}{1'} \underset{n_0 \times 1}{y_{0t}} + \underset{1 \times n_1}{B_{01}} \underset{n_1 \times 1}{y_{1t}} + \underset{1 \times m_1}{\Gamma_{01}} \underset{m_1 \times 1}{x_{1t}} = u_{0t}$

Key REMARK:

EXCLUDED EXO. VAR: $X_{2t} \rightsquigarrow m_2$ variables

INCLUDED END. VAR: $y_{1t} \rightsquigarrow n_1$ variables

* notice, even if y_{0t} is in the equation, the fact that is normalized by a coefficient makes it so that we consider it as excluded $\rightsquigarrow$ it's not going to threaten our ability to estimate B_{01} and Γ_{01}.

The reduced form of the system is:

$$\begin{array}{c} \\ n_1 \\ \\ \end{array}\begin{bmatrix} y_{0t} \\ y_{1t} \\ y_{2t} \end{bmatrix} = \underbrace{\begin{bmatrix} \Pi_{01} & \Pi_{02} \\ \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix}}_{n \times m} \underbrace{\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}}_{m \times 1} + V$$

where: $\Pi = -B^{-1} \cdot \Gamma$ OR $B \cdot \Pi = -\Gamma$

$$\begin{bmatrix} 1 & B_{01} & 0 \\ B_{10} & B_1 & B_{12} \\ B_{20} & B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} \Pi_{01} & \Pi_{02} \\ \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} = - \begin{bmatrix} \Gamma_{01} & 0 \\ \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}$$

focus on equation zero $\rightsquigarrow$ we obtain 2 equations:

$$\begin{cases} (\hat{\Pi}_{01} + B_{01} \cdot \hat{\Pi}_{11} = -\Gamma_{01})' \\ (\hat{\Pi}_{02} + B_{01} \hat{\Pi}_{12} = 0)' \end{cases}$$

$1 \times m_1$

$1 \times n_1$

! Here our unknowns are B_0 and Γ_{01}, while Π can be estimated through data $\rightsquigarrow \hat{\Pi}$

From the system $\rightsquigarrow$ we can solve for B_{01} in the second equation and plug the result in the first equation to get Γ_{01}.

look at the second eq.:

$$\underset{m_2 \times n_1}{\Pi_{12}'} \; \underset{n_1 \times 1}{B_{01}'} = \underset{m_2 \times 1}{-\Pi_{02}}$$

the equation involves: m_2 equations with n_1 unknowns

For the equation to have a solution:

- at least as many eq. as unknowns: $m_2 \geq n_1$ ORDER
  $\hookrightarrow$ necessary but not sufficients
- we need the equations to be linearly independent:
  $\text{Rank}(\Pi_{12}') = n_1$ provides a unique solution.

In Summary:

- RANK + ORDER Conditions make sure that the unknows are uniquely identified.

Alternative understanding using IV logic.

Given the restriction:

eq.0 $\Rightarrow$ $y_{0t} + B_{01} \cdot y_{1t} + \Gamma_{01} X_{1t} = u_{0t}$

By orthogonality of: $u_{0t} \perp X'_t$ (exogeneity ass.)

$$\Rightarrow \quad E(u_{0t} \cdot X'_t) = 0$$

notice, here $X'_t = [X'_{1t} \;\; X'_{2t}]$

$$\Rightarrow \quad E(u_{0t} \cdot X'_t) = E[(y_{0t} + B_{01} y_{1t} + \Gamma_{01} X_{1t}) \cdot X'_t] = 0$$

$$E[y_{0t} \cdot X'_t] + \underbrace{B_{01} \cdot E[y_{1t} \cdot X'_t] + \Gamma_{01} E[X_{1t} \cdot X'_t]}_{[B_{01} \;\; \Gamma_{01}] \cdot \begin{bmatrix} E(y_{1t} X'_t) \\ E(X_{1t} \cdot X'_t) \end{bmatrix}} = 0$$

By taking the transpose:

$$[E(y_{1t} \cdot X'_t)' \;\; E(X_{1t} \cdot X'_t)'] \cdot \begin{bmatrix} B'_{01} \\ \Gamma'_{01} \end{bmatrix} = -\underbrace{E[y_{0t} X'_t]'}_{(1 \times m)'} \rightarrow \boxed{m \times 1}$$

$$\| $$

$$\begin{bmatrix} \overset{n_1 \times 1}{E(y_{1t}} \; \overset{1 \times m_1}{X'_{1t})} & \overset{n_1 \times 1}{E(y_{1t}} \; \overset{1 \times m_2}{X'_{2t})} \\ \underset{m_1 \times 1}{E(X_{1t}} \; \underset{1 \times m_1}{X'_{1t})} & \underset{m_1 \times 1}{E(X_{1t}} \; \underset{1 \times m_2}{\cdot X'_{2t})} \end{bmatrix}' \rightarrow (1 \times (n_1 + m_1))' \rightarrow \boxed{(n_1 + m_1) \times 1}$$

$$(n_1 + m_1) \times \underbrace{(m_1 + m_2)}_{m} \rightarrow \boxed{m \times (n_1 + m_1)}$$

in order to have at least one solution:

※ equations: $m \geq n_1 + m_1$ ※ variables

$$\cancel{m_1} + m_2 \geq n_1 + \cancel{m_1} \quad \Rightarrow \quad m_2 \geq 1 \qquad \text{ORDER}$$

In order to have one solution: RANK

$$\text{Rank}\left(\begin{bmatrix} E(y_{1t} X'_{1t}) & E(y_{1t} X'_{2t}) \\ E(X_{1t} X'_{1t}) & E(X_{1t} \cdot X'_{2t}) \end{bmatrix} \right) = n_1 + m_1$$

equivalent to $\text{Rank}(\Pi'_{12}) = n_1$

So, before Sim critique, models where identified by imposing restrictions of two kinds:

1- Contemporaneous exclusion restrictions
2- Exogeneity Assumptions.

Both restrictions were imposed in a AD HOC way.

Then came Sims

2 major ~~concepts~~ concepts in economics were introduced:

1) Rational Expectations:

Equations in the system are all related + imposing some lags to be zero is unlikely

2) General Equilibrium: important one

ALL variables must be determined inside the system

↳ no distinction between exogen. and endog.

NO ZERO LAGS : Each variable depends on all other variables.

Solution found by Sims: Structural VAR (SVAR)

- no distinction between X_t and Y_t

SVAR

Pre-determined but still endogenous.

STRUCT. FORM $\quad A \cdot X_t = B X_{t-1} + u_t$

$$u_t \sim N(0, \Sigma)$$

REDUCED FORM: $\quad X_t = C X_{t-1} + V_t$

$$V_t \sim N(0, \Omega)$$

Also: only one lag is shown here, this is just for simplicity, it poses no limitat. in its generalization

where: $C = A^{-1} \cdot B$; $V_t = A^{-1} \cdot u_t$; $V(V_t) = A^{-1} \cdot \Sigma A^{-1} = \Omega$

Still, to identify the model we need some RESTRICTIONS, but:

- we don't want zero lags ⇝ dynamics should be left unrestricted
- we don't want instruments as before (Y_t vs X_t)

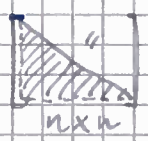
On the other hand, restrictions can still be imposed on:

1) Contemporaneous relations (matrix A)
2) Covariance between shocks (Σ)

The idea is to impose as little restrictions as we possibly can. How many do we need?

From data we can estimate C and Ω (red. form)

In C there are n^2 parameters $\underset{n \times n}{C}$

In Ω there are $\frac{n(n+1)}{2}$ distinct elements  $= \Omega$

In the Structural Form: $A \rightsquigarrow n^2$ param.

$B \rightsquigarrow n^2$ param.

$\Sigma \rightsquigarrow \frac{n(n+1)}{2}$ elem.

$\Rightarrow$ n^2 more param. in the Struct. form than in the estimatable Reduced Form.

$\Rightarrow$ n^2 restrictions must be imposed on A and Σ

Assumption 1: $\boxed{\Sigma = I_n}$

shocks in μ_t are Orthogonal, uncorrelated

economic intuition: when thinking about a certain shock in the economy it is useful to not have correlations with other shocks otherwise it would be impossible to consider the effects of just that peculiar shock.

Ass. 1 generates $\boxed{\frac{n(n+1)}{2} \text{ restrictions}}$ $+ \rightarrow$ n^2 total restrict.

Assumption 2: A is lower triangular $\rightsquigarrow \boxed{\frac{n \cdot (n-1)}{2} \text{ restrict.}}$

example

$$\overset{A}{\begin{bmatrix} a_1 & 0 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & c_3 \end{bmatrix}} X_t = B X_{t-1} + \mu_t \qquad \text{STR.}$$

$$\Leftrightarrow$$

$$X_t = \begin{bmatrix} \tilde{a}_1 & 0 & 0 \\ \tilde{b}_1 & \tilde{b}_2 & 0 \\ \tilde{c}_1 & \tilde{c}_2 & \tilde{c}_3 \end{bmatrix} B X_{t-1} + \begin{bmatrix} \tilde{a}_1 & 0 & 0 \\ \tilde{b}_1 & \tilde{b}_2 & 0 \\ \tilde{c}_1 & \tilde{c}_2 & \tilde{c}_3 \end{bmatrix} \cdot \mu_t \qquad \text{RED.}$$

Implications:

- A Recursive Structure is imposed on the system.

  look at how shocks affect X_t:

$$\begin{bmatrix} \tilde{a}_1 & 0 & 0 \\ \tilde{b}_1 & \tilde{b}_2 & 0 \\ \tilde{c}_1 & \tilde{c}_2 & \tilde{c}_3 \end{bmatrix} \cdot \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \mu_{3t} \end{bmatrix}$$

Given a shock in μ_{3t} the only variable directly affected is X_{3t}, which in turn affects X_{2t} and X_{1t} but only in the period after, so:

$$\begin{array}{lll} \mu_{3t} & \rightarrow X_{3t} & \rightarrow X_{1t+1}\ X_{2t+1}\ X_{3t+1} \\ \mu_{2t} & \rightarrow X_{2t},\ X_{3t} & \rightarrow \quad '' \\ \mu_{1t} & \rightarrow X_{1t},\ X_{2t}\ \&\ X_{3t} & \rightarrow \quad '' \end{array}$$

---

estimating the STRUC. parameters

From data: $\hat{C}$ & $\hat{\Omega}$

$$\hat{C} = A^{-1} \cdot B$$
$$\hat{\Omega} = A^{-1} \cdot A^{-1\prime}$$

$\downarrow$

system on n^2 equations with n^2 unknowns

it can be shown to have a unique solution

In particular, we can exploit the following result for the estimation of A:

PROP. Any postive definite matrix Ω can be written as:

$$\Omega = G \cdot G'$$

where G is lower triangular and its called the Choleski Factor

In our case, Ω is p.d by definition of var-cov matrix.

The STRUCT. FORM is then

Chol. Decomposition: $G^{-1} X_t = G^{-1} \cdot C X_{t-1} + G^{-1} \cdot V_t$ → Choleski Factor

where $V(G^{-1} \cdot V_t) = G^{-1} \cdot \Omega \cdot G^{-1\prime} = \dots = I_n = \Sigma$ ✓

Remark

- There are an infinite number of STRC. models compatible with the same REDUCED Form
  - ↳ the choice of the STRUCT. Model is not a statistical choice, but a choice regarding economic modelling.

Standard example of SVAR

$$\begin{bmatrix} a_1 & 0 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} Y_t \\ \pi_t \\ FFR_t \end{bmatrix} = B \begin{bmatrix} Y_{t-1} \\ \pi_{t-1} \\ FFR_{t-1} \end{bmatrix} + \begin{bmatrix} \mu^1_t \\ \mu^2_t \\ \mu^{FFR}_t \end{bmatrix}$$

Y_t → GDP

π_t → INFLATION

FFR_t → Federal Fund Rate

focus on the last equation (the only one including current FFR):

Information Function

$$FFR_t = \underbrace{f(\mathcal{I}_t)}_{} + \mu^{FFR}_t$$

↳ includes:
- current levels of Y_t and π_t
- lags

intuition: • Today's decision of the FFR depends on current aggregate (Taylor Rule), lags and some unexpected component (i.e today FFR cannot be perfectly predicted by $f(\mathcal{I}_t)$).

- Also, Y_t and π_t don't respond to FFR_t, but only to FFR_{t-1} ⇝ lagged effect of Monetary policy.
  - ↳ it takes time for Y_t and π_t to adjust.
  - ↳ sticky prices intuition.

What does this model predict? Monetary Transmission

From data:

- An increase in the FFR_t generates an increase in π_t
  - ↳ Completely opposite to economic intuition
  - ↳ this is called the PRICE PUZZLE

Sims: The Information set function $f(\mathcal{I}_t)$ does not include any measure of future inflation expectation, which is taken into consideration by the CB for its policy rule.

Sims solution: include Commodity prices $\rightsquigarrow$ they signal future inflation.

- An alternative way to look at identification

Before: Instrumental variables were used for identifications i.e variables that were exogenous to all but one end. variable and therefore allowed for estimation.

problem arisise from simultaneity of variables, the IV solves by being exogenous.

Now: look at Choleski Structural Form.

1) $y^1_t = a(L) \cdot y_{t-1} + \varepsilon^1_t$

2) $y^2_t + a y^1_t = b(L)\, y_{t-1} + \varepsilon^2_t$

3) $y^3_t + a y^2_t + c y^1_t = c(L)\, y_{t-1} + \varepsilon^{3t}$

notice, in 1 $\rightsquigarrow$ no simultaneity

in 2 and 3 $\rightsquigarrow$ There is simultaneity but:

e.g in 2 I need y^1_t to be orth. to ε^2_t to correctly estimate the model: $y^1_t \perp \varepsilon^2_t$ and it is! same for 3

It turns out that the Choleskey shcema assumptions (lower triangularity) allows us to estimate without endogeneity problems.

- The Impulse response function (IRF)

We can compute the IRF from the VMA representation of the REDUCED FORM.

The RED. FORM has a VAR form, which can be transformed into a VMA through its Wold Decomposition.

RED. FORM $\quad X_t = C \cdot X_{t-1} + V_t$

VMA $\quad X_t = V_t + C^1 V_{t-1} + C^2 V_{t-2} + \cdots + C^k V_{t-k}$

VMA with struct. parameters:

$$X_t = A^{-1} \cdot u_t + \underbrace{A^{-1} \cdot B}_{C^1} \cdot \underbrace{A^{-1} \cdot u_{t-1}}_{V_{t-1}} + (A^{-1} \cdot B)^2 \cdot A^{-1} u_{t-2} + \cdots + (k)$$

- ERROR BANDS : Consider the model : $\boxed{Y_t = A_1 y_{t-1} + \cdots + u_t}$

VAR(p)

Modern method to compute error bands:

BOOTSTRAP METHOD (Runkle 1987)

↳ this method allows us to build a distribution for the observ. from a small sample

Steps:

1. Estimate VAR (RED. FORM) and save the fitted residuals:
$$\{\hat{u}_1, \ldots\ldots, \hat{u}_T\} = \textcircled{\sim}$$

2. Randomly draw from $\textcircled{\sim}$, which provides us with the realization $\tilde{u}_1^{(1)}$. Use this realization to compute:
$$y_1^{(1)} = \hat{A}_1 y_0 + \cdots + \tilde{u}_1^{(1)}$$

3. Take a second draw from $\textcircled{\sim}$ (note: first draw is with replacement so sample to draw from remains the same through out the process)
get $\tilde{u}_2^{(1)}$ and derive: $y_2^{(1)} = \hat{A}_1 \hat{y}_1^{(1)} + \cdots + \tilde{u}_2^{(1)}$

4. Repeating the process we get:
$$\{y_1^{(1)}, y_2^{(1)} \cdots y_T^{(1)}\}$$
Use the sample to compute new VAR estimates and the implied IRF (from the VMA repres.)

5. Repeat steps 3 and 4, M times to get:

$$\{Y_t^{(1)}\}_0^T ; \{Y_t^{(2)}\}_0^T ; \dots ; \{Y_t^{(M)}\}_0^T$$

⇝ i.e M processes generated by the same pool of residuals.

Finally, after having computed all the IRF generated confidence intervals for all the horizons given the chose α^{th} level.

---

lec.

- Forecast Variance Error Decomposition FEVD

idea: we are looking for the variance of a given variable from a given shock

Recall Forecast variance at each step:

step1: model ⇒ $Y_t = \Psi(L) \cdot \varepsilon_t$ ⇝ F.E ⇒ ε_t $V(FE) = \Omega$

step2: $V(FE) = \Omega + \Psi \Omega \Psi'$

step3: $V(FE) = \Omega + \Psi \Omega \Psi' + \Psi^2 \Omega \Psi^{2'}$

step4: ....

in the limit: unconditional variance

notice Ω is the variance of the ~~res~~ model residuals ε_t.
To get FV from the structural shocks we need to retrieve the structural model ⇝ Choleski identification

$$\varepsilon_t = A^{-1} \cdot \mu_t \quad \text{⇝ struct. shocks}$$

By restrictions: $V(\mu_t) = I_n$

Therefore: $V(\varepsilon_t) = A^{-1} \cdot A^{-1'} = G \cdot G' = \Omega$ (Choleski Factor)

example

Consider a 2 variables / 2 struct. shocks model

st

$$V(\varepsilon_t) = \Omega = \underbrace{\begin{bmatrix} x & 0 \\ y & z \end{bmatrix}}_{\text{Choleski Factor}} \cdot \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} = \begin{bmatrix} x^2 & yx \\ yx & y^2+z^2 \end{bmatrix}$$

notice that Ω can be decomposed as:

$$\Omega = \begin{bmatrix} x \\ y \end{bmatrix} [x \quad y] + \begin{bmatrix} 0 \\ z \end{bmatrix} [0 \quad z]$$

↳ FIRST COLUMN of G · (FIRST. COL. of G)'

in short form:

$$\Omega = G \cdot G' = \underbrace{G_{(1)} \cdot (G_{(1)})'}_{\text{var-cov due to the FIRST shock}} + \underbrace{G_{(2)} \cdot (G_{(2)})'}_{\text{var-cov due to the SECOND shock}}$$

- the FEVD of a SVAR for the one-step ahead prediction can be generalized into:

$$\Omega = G_{(1)} \cdot (G_{(1)})' + G_{(2)} \cdot (G_{(2)})' + \cdots\cdots + G_{(j)} \cdot (G_{(j)})'$$

notice that by using a Choleski identification scheme the var-cov due to the last shock will only have the bottom left value postive, since the last shock affects exclusively the last variable (in the one step ahead.

In the 2x2 example, $G_{(2)} \cdot (G_{(2)})' = \begin{bmatrix} 0 & 0 \\ 0 & z^2 \end{bmatrix}$

Notice that by looking inside the var-cov of each structural shock we can further decompose the variance of each variable due to the given shock looking at $G_{(1)} \cdot (G_{(1)})'$:

variance of the first variable due to the FIRST SHOCK

$$\begin{bmatrix} x^2 & xy \\ xy & y^2 \end{bmatrix}$$

variance of the second variable due to the FIRST shock

In particular we can write:

$$\text{diag}\left(\frac{G_{(1)} \cdot (G_{(1)})'}{\Omega} \right) = \underbrace{\frac{x^2}{\Omega_{11}}}_{} ; \underbrace{\frac{y^2}{\Omega_{22}}}_{}$$

Fraction of the first variable tot. variance caused by the FIRST SHOCK

Fraction of the second variable total variance caused by the FIRST SHOCK

in the ONE-STEP Forecast

- of course, this can be generalized to all steps in the forecasts. To do so just substitue the $G(j)$'s with the columns of the appropriate forecast variance.

e.g step 2

$$V(FE) = \Omega + \Psi \Omega \Psi'$$

$$= G \cdot G' + \Psi G \cdot G' \cdot \Psi' = \Gamma$$

$$= \Gamma(1) \cdot (\Gamma(1))' + \Gamma(2) \cdot (\Gamma(2))' + \cdots + \Gamma(j) \cdot (\Gamma(j))'$$

Same properties as before hold.

- HISTORICAL DECOMPOSITION

idea: we want to compare the relative importance of a shock in generating the data process.

Consider: $\quad I_n \hat{y}_t = A \hat{y}_{t-1} + \hat{u}_t$

i.e this is a SVAR struct. Form coinciding with its Reduced Form. From estimation we get the historical series of each residual (or str. shock)

$$\left.\begin{matrix} \hat{u}_{1t}, \hat{u}_{1\,t-1} & \cdots & \hat{u}_{1\,0} \\ \hat{u}_{2t}, \hat{u}_{2\,t-1} & \cdots & \hat{u}_{2\,0} \\ \vdots & & \\ \hat{u}_{st}, \hat{u}_{s\,t-1} & \cdots & \hat{u}_{s\,0} \end{matrix}\right\} \begin{matrix} TS_1 \\ TS_2 \end{matrix}$$

Now, Generate ~~xxx~~ the data only using A and TS_1 + Compare withe data generated using A and TS_2

↳ From this comparison one should get the relative importance of the first shock in generating the observations.

By now we have only seen the Choleski identification scheme, let us look at alternatives.

- BLANCHARD - QUAH 1989 (used similiarly in Gali 99)

idea: We want to restrict ~~[illegible]~~ shocks from having long-term effects on variables.

Consider the VMA of the following model:

$$\begin{array}{l} \text{Unemployment} \\ \%\,\Delta GDP \end{array} \begin{bmatrix} X_t \\ \Delta y_t \end{bmatrix} = \Psi(L) \cdot \begin{bmatrix} \mu_t^s \\ \mu_t^d \end{bmatrix} \quad \begin{array}{l} \text{Supply shock} \\ \text{Demand shock} \end{array}$$

↳ taken in difference to make it stationary

Restriction 1 : $V(\mu_t) = I_n$

↳ equally to the Choleski scheme: driving forces of the economy are Orthogonal.

↳ the condition imposes 3 restrictions (needed: 4)

Restriction 2 : Demand shocks (μ_t^d) have NO long-term effect on X_t

notice: if X_t was stationary then the restriction would be obvious since the effects of all shocks converge toward zero.

~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

idea: the response of a variable to a given shock can be seen as the first order derivative of the variable wrt the shock

Suppose : $\Delta y_t = (A_0 + A_1 L + A_2 L^2 + \cdots)\,\mu_t$

$\rightsquigarrow$ $y_t = y_{t-1} + (A_0 + \cdots)\,\mu_t$

$\rightsquigarrow$ IR of y_t to a μ_t shock : $\dfrac{\partial y_t}{\partial \mu_t} = A_0$

by the same logic:

$$y_{t+1} = y_t + (A_0 + \cdots)\,\mu_{t+1}$$
$$y_{t+1} = (A_0 + \cdots)\,\mu_t + (A_0 + \cdots)\,\mu_{t+1} + y_{t-1}$$

$\rightsquigarrow$ $\dfrac{\partial y_{t+1}}{\partial \mu_t} = A_0 + A_1$

We can generalize by stating that the IR of y_{t+n} to a μ_t shock is:

$$\frac{\partial y_{t+n}}{\partial \mu_t} = \sum_{i=0}^{n} A_i$$

Furthermore, notice that as $n \to +\infty$:

$$\frac{\partial y_{t+n}}{\partial \mu_t} = A(\overset{L}{1}) = (A_0 + \overset{L}{1} A_1 + \overset{L^2}{1} A_2 + \overset{L^3}{1} A_3 + \cdots)$$

Therefore, imposing no long run effect of μ_t to y is equivalent to setting : $A(1) = 0$

~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

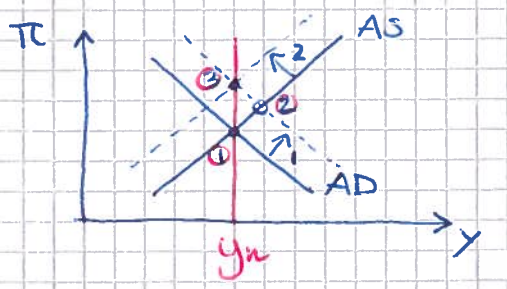
Reconsider the original model:

$$\begin{matrix} X_t \\ \Delta y_t \end{matrix} = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \mu_t^s \\ \mu_{ts}^d \end{bmatrix}$$

IR of Δy_t to a μ_t^d shock

$\Rightarrow$ no l-t effect of μ_t^d on Δy $\approx$ $\Psi_{22}(1) = 0$

- Economic intuition behind restriction



economic int. : Output will always go back to its natural level y^n

role of the shocks:

- μ_t^s ⇝ accounts for s-t supply shocks and l-t economic growth.
- μ_t^d ⇝ highlights the role of Fiscal and Monetary shocks. $\Rightarrow$ only s-t effects.

Q: How do we impose $\Psi_{22}(1) = 0$?

Let's start from the Struct. VAR Form:

$$A \cdot \Delta y_t = B \, \Delta y_{t-1} + u_t$$

$$\Delta y_t = \underbrace{A^{-1} \cdot B}_{} \cdot \Delta y_{t-1} + A^{-1} \cdot u_t$$

estimate $\hat{C}$ $\quad \Delta y_t = \hat{C} \, \Delta y_{t-1} + A^{-1} \cdot u_t$

VMA Form $\quad \Delta y_t = (I + \hat{C} L + \hat{C}^2 L^2 + \cdots) \, A^{-1} \cdot u_t$

imposing $L = 1$ we can rewrite: $(I + \hat{C} L + \hat{C}^2 L^2 + \cdots) = (I - \hat{C})^{-1}$

$$\Rightarrow \Delta y_t = \underbrace{(I - \hat{C})^{-1} \cdot A^{-1}}_{} \cdot u_t$$

notice, we want $\Rightarrow$ $\begin{bmatrix} * & * \\ * & 0 \end{bmatrix} \rightarrow \Psi_{22}(1) = 0$

- Restriction on A: (1)

  by Restriction 1 we want $V(u_t) = I_n$

  notice that $u_t = A \cdot v_t$ by construction

  $V(u_t) = V(A \cdot v_t) = A \cdot V(v_t) \cdot A'$

  $= A \cdot \hat{\Omega} \cdot A' = I_n$ IFF $\boxed{\hat{\Omega} = A^{-1} \cdot A^{-1'}}$

  $\Rightarrow$ we want $A^{-1} \cdot A^{-1'} = \hat{\Omega}$

- Restriction on $(I - \hat{C})^{-1} \cdot A^{-1}$: (2)

$$\boxed{[(I - \hat{C})^{-1} \cdot A^{-1}]_{22} = 0}$$

Method 2: exploit Choleski identification scheme

First of all, invert the variables s.t the restriction will apply to the the top righ element of the $\Psi(L)$ matrix:

$$\begin{bmatrix} \Delta y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} u_t^s \\ u_t^d \end{bmatrix}$$

here we want: $\Psi_{12}(1) = 0$

$\rightarrow$

Consider the following steps:

1) Start from the VMA form of the estimated VAR:

$$Y_t = \hat{D}(L) \cdot V_t$$

2) derive $\hat{D}(1)$ as $[I_n - \hat{D}]^{-1}$ and get the term:

$$D(1) \cdot \hat{\Omega} \cdot D(1)'$$

3) Use Choleski to the term above to get:

$$\text{Chol.}(D(1) \cdot \hat{\Omega} \cdot D(1)') = S \cdot S'$$

lower triangular.

4) Consider the term $K = D(1)^{-1} \cdot S$

5) Write the Struct. shocks as: $\mu_t = K^{-1} \cdot V_t$

6) Write the Struct. VMA: $Y_t = D(L) \cdot K \cdot \mu_t$

Notice that the two restrictions are verified:

1 $$V(\mu_t) = I_n$$

$$V(\mu_t) = V(K^{-1} \cdot V_t) = K^{-1} \cdot \hat{\Omega} \cdot K^{-1\prime}$$

$$= S^{-1} \underbrace{D(1) \cdot \hat{\Omega} \cdot D(1)'}_{S \cdot S'} \cdot S^{-1\prime}$$

$$= I_n$$

2 $$[D(1) \cdot K]_{(1,2)} = 0$$

notice $D(1) \cdot K = D(1) \cdot D(1)^{-1} \cdot S = S$

$S_{(1,2)} = 0$ by definition

↳ lower triangular.

lec.

- General Approach to Identification

AKA Identification via Rotation

idea:

1 - Keep restrictions on the var-cov of the struct. shocks:

$$V(\mu_t) = I_n$$

$$\downarrow$$

$$\frac{n(n+1)}{2} \text{ restrictions}$$

2 - pick a matrix R with $\frac{n(n-1)}{2}$ restrictions

Consider the VMA form of the reduced form:

$$Y_t = C(L) \cdot V_t$$

$$\text{with } V(V_t) = \Omega$$

1) rewrite: $Y_t = C(L) \cdot P \cdot P^{-1} \cdot V_t$

where P is s.t $(P \cdot P' = \Omega)$

so that the structural shocks are: $P^{-1} \cdot V_t = \mu_t$

& $V(\mu_t) = P^{-1} \cdot \Omega \cdot P^{-1\prime} = P^{-1} \cdot P \cdot P' \cdot P^{\prime -1} = I_n$

2) let R be a nxn matrix s.t $R'R = I_n$

$$\Rightarrow Y_t = \underbrace{C(L) \cdot P \cdot R'}_{IR} \cdot \underbrace{R \cdot \mu_t}_{\text{new STRUCT. SHOCKS}}$$

notice that:

- $V(R \cdot \mu_t) = R \cdot I_n \cdot R' = I_n$

↳ the var-cov is still Orthogonal.

- $C(L) \cdot P \cdot R'$ is the generalized IR of the SVAR

In general we can write R in a generalized form depending on a parameter a.

R changes wrt to the number of variables in the system.

For $n = 2$

$$R = \begin{array}{cc} \begin{bmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{bmatrix} & \begin{array}{c} x \\ y \end{array} \end{array}$$

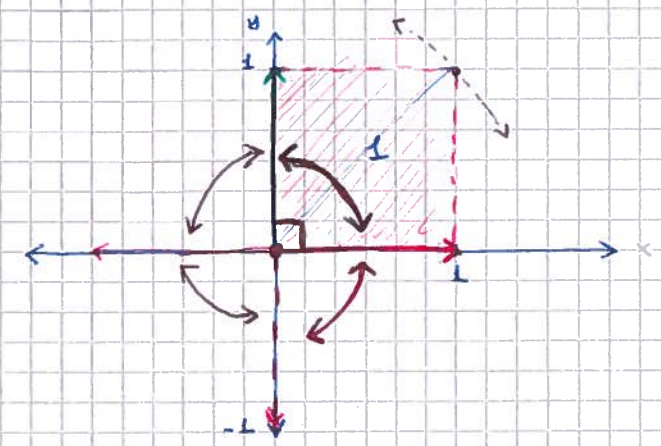
For $n = 3$

$$R = \begin{array}{cc} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos a & \sin a \\ 0 & -\sin a & \cos a \end{bmatrix} \begin{bmatrix} \cos b & \sin b & 0 \\ -\sin b & \cos b & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos c & 0 & \sin c \\ 0 & 1 & 0 \\ -\sin c & 0 & \cos c \end{bmatrix} & \begin{array}{c} x \\ y \\ z \end{array} \end{array}$$

~ ~ ~ ~ ~ ~ ~

Geometric interpretation

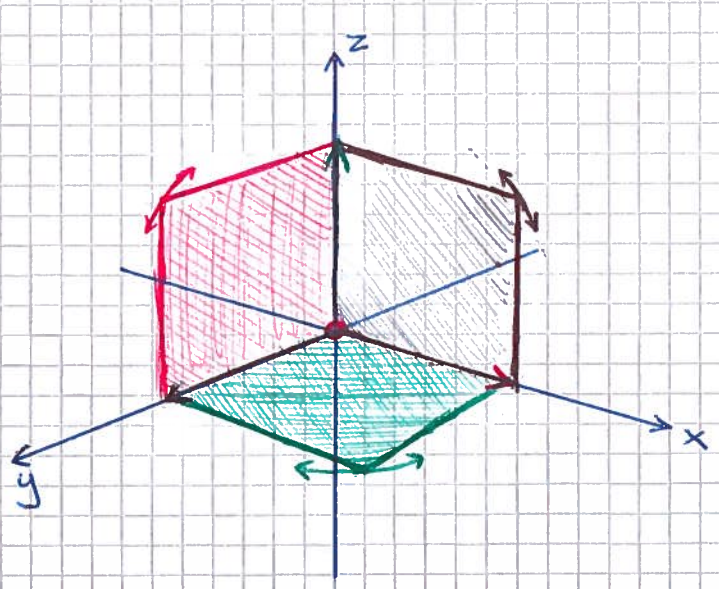
$n = 2$



$a = 0$

idea: Square of Area 1 mapping all points inside a circle with ray $\sqrt{2}$

$n = 3$



~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

- **Replicating Choleski identification scheme**

we want $R = I_n$ s.t $Y_t = \underbrace{C(L) \cdot P R'}_{C(L) \cdot P} \cdot \underbrace{R \cdot \mu_t}_{\mu_t}$

where P is s.t $P' \cdot P = \Omega$ ⇒ Choleski Factor
and can be found to be lower triangular
i.e $P = G$

e.g for $n = 2$ let $a = 0$ ⇒ $R = I_n$

- **Replicating BLANCHARD & QUAH**

Consider the known B&Q Model but with both GDP and Unemployment taken in first difference.

Structural VMA

$$\Delta X_t = \Theta(L) \cdot \mu_t$$

$$\Updownarrow$$

$$\begin{matrix} \Delta y_t \\ \Delta \text{unem.} \end{matrix} = \begin{bmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{bmatrix} \begin{matrix} \mu_t^d \\ \mu_t^s \end{matrix}$$

Impose: 1) $V(\mu_t) = I_n$ 2) $\theta_{11}(1) = 0$

As before, rewrite:
let P be an $n \times n$ s.t $P' \cdot P = \Omega$
& $R \cdot R' = I_n$

$$\Rightarrow \quad \Delta X_t = \underbrace{C(L) \cdot P \cdot R}_{\Theta(L)} \cdot \underbrace{R' \cdot P^{-1} \cdot V_t}_{\mu_t}$$

we need to impose: $[C(1) \cdot P \cdot R]_{11} = 0$

$$C(1) \cdot P \cdot R = \begin{bmatrix} C_{11}(1) & C_{12}(1) \\ C_{21}(1) & C_{22}(1) \end{bmatrix} \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{bmatrix}$$

$$[C(1) \cdot P \cdot R]_{11} = (C_{11}(1) P_{11} + C_{12}(1) \cdot P_{21}) \cdot \cos \underline{a} - (C_{12} \cdot P_{22}) \cdot \sin \underline{a} = 0$$

Solve for a.

- Sign Restrictions

idea: we want to restrict the sign of the response depending on the shock.

ex.: After a positive shock in the FFR we expect:

$$\uparrow \mu_t^{FFR} \rightarrow \begin{cases} Y_t \downarrow \\ FFR_t \uparrow \\ P_t \downarrow \end{cases}$$

Structural Model: $Y_t = \underbrace{C(L) \cdot P \cdot R}_{IR} \cdot \underbrace{\tilde{\mu}_t}_{R' \mu_t}$

How do we apply this restriction?

notice: R is a 3x3 matrix depending on three parameters: a, b, c

- RANDOMLY PICK a, b & c
- IF the IRF complies with the sign restriction keep it.
- IF NOT discard
- use Monte Carlo repetitions
- At the end we have a sample of IRFs complying with our restriction ⇝ mean + C.I

Co-Integration

For simplicity let $I(1)$ be a non-stationary process & $I(0)$ a Stationary process.

def.

Two $I(1)$ variables, x_t and y_t, are co-integrated if there exist at least one linear combination which is $I(0)$

i.e $\exists\, \alpha, \beta \in \mathbb{R}$ s.t $\beta y_t - \alpha x_t \sim I(0)$

Remark

1) notice if $\exists$ one $\Rightarrow \exists\ \infty$

PF. if $\alpha\beta$ is s.t $\beta y_t - \alpha x_t \sim I(0)$

then $\forall k \in \mathbb{R}$: $k \cdot (\beta y_t - \alpha x_t) \sim I(0)$

$\Rightarrow \quad (k \cdot \beta) \cdot y_t - (\alpha k) \cdot x_t \sim I(0)$

2) The Co-integrating vector is $[\beta \quad -\alpha]$

3) We say that there exist more than one co-integrating vector when each of them is linearly independent from the other

e.g Consider a vector of variables Y_t $(N \times 1)$ $\sim I(1)$

let there be K linearly independent combinations that yield $K \sim I(0)$ vectors

then $\Rightarrow$ let $\beta = \underbrace{\begin{bmatrix} | & | & | & | & | & | \end{bmatrix}}_{N \times K}$

where each column is a co-integrating vector and its lin. independent

notice: β is FULL RANK & $N > K$ by definition

Furthermore, we can compute the $I(0)$ processes as:

$$\beta' \cdot Y_t = \underset{K \times 1}{Z_t} \sim I(0)$$

example

idea: If in our SVAR model there are variables $I(1)$ that can be co-integrated, taking the first difference of each variable (ignoring the co-integration) makes the model non-estimable as the reduced form VAR does not exist anymore.

Consider the following SVAR Model:

$$\underbrace{\begin{bmatrix} 1 & -\gamma \\ 0 & 1 \end{bmatrix}}_{A} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}}_{B} \begin{bmatrix} y_{1\,t-1} \\ y_{2\,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

$$\Updownarrow$$

$$\begin{cases} y_{1t} - \gamma y_{2t} = u_{1t} & \text{eq.1} \\ y_{2t} = y_{2\,t-1} + u_{2t} & \text{eq.2} \end{cases}$$

1. notice: y_2 is a non-stationary variable (unit root), y_1 is a function of a non-stationary variable $\Rightarrow$ non-stationary

$\Rightarrow$ both y_1 and y_2 are $I(1)$

2. notice: From eq.1 we clearly see that:

$[1 \quad -\gamma]$ is a co-integrating vector.

By inverting A we can get the REDUCED FORM:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix}}_{C} \begin{bmatrix} y_{1\,t-1} \\ y_{2\,t-1} \end{bmatrix} + \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

1. notice: As we have mentioned, y_1 and y_2 are $I(1)$

$\Rightarrow |I - C \cdot z| = 0$ for $z = 1$ (non-stationary)

2. From the REDUCED FORM we can easily get the Forecast errors: (Preliminary step)

- Clearly: $\varepsilon_{2t} = u_{2t}$ since $y_{2\,t+1} - y_{2\,t+1|t} = \cancel{y_{2t}} + u_{2\,t+1} - \cancel{y_{2t}}$
- $y_{1\,t+1} - y_{1\,t+1|t} = u_{1,t+1} + \gamma y_{2\,t+1} - \gamma y_{2\,t+1|t}$

$= u_{1\,t+1} + \gamma (u_{2\,t+1})$

$\Rightarrow \varepsilon_{1t} = u_{1t} + \gamma u_{2t}$

By noticing that the two processes are I(1) we might want to take the first differences:

$$\begin{cases} \Delta y_{1t} = \gamma \Delta y_{2t} + \Delta u_{1t} = \gamma u_{2t} + \Delta u_{1t} \\ \Delta y_{2t} = u_{2t} \end{cases}$$

$$\Updownarrow$$

$$\begin{bmatrix} 1 & -\gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 1-L & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

VMA FORM

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 1-L & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Using the previously obtained forecast errors:

1. $u_{1t} = \varepsilon_{1t} - \gamma u_{2t} = \varepsilon_{1t} - \gamma \varepsilon_{2t}$
2. $u_{2t} = \varepsilon_{2t}$

we get to:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \underbrace{\begin{bmatrix} 1-L & \gamma L \\ 0 & 1 \end{bmatrix}}_{\Psi(L)} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

notice: This is the reduced VMA form that we should get once the model is estimated throug a VAR. But, can we get this VMA from a VAR? IFF the matrix of polynomials $\Psi(L)$ is invertible

$\Rightarrow$ check $|\Psi(z)| = 0$ for $z = 1$ non-invertible

$\Rightarrow$ A VAR for Δy_t does not exist!

- Impulse Response

For the differentiated model we had: $\begin{bmatrix} 1-L & \gamma \\ 0 & 1 \end{bmatrix}$

$\hookrightarrow$ the long run IR: $(L=1)$

- a shock in u_{1t} has NO long run effect on y_t
- a shock in u_{2t} has a long run effect of:
  - $\sim \gamma$ on y_{1t} & 1 on y_{2t} $\rightsquigarrow$ non-stationary

$\Rightarrow$ BUT their co-integration: $y_{1t} - \gamma y_{2t} \sim I(0)$

- **Implications for the VAR representation**

Recall the Reduced VAR Form:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix}}_{C} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix}}_{\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}}$$

We want to get the Canonical VAR representation:

$$\phi(L) \cdot y_t = \varepsilon_t$$

in this case:

$$\underbrace{\left(I_2 - \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix} \cdot L \right)}_{\phi(L)} y_t = \varepsilon_t$$

$$\phi(L) = \begin{bmatrix} 1 & -\gamma L \\ 0 & 1-L \end{bmatrix}$$

- **Property of Co-integration**

The number of linearly independent Co-integrating vectors is equal to the Rank of the $\phi(1)$ matrix.

$\to K$

in this case: $\begin{bmatrix} 1 & -\gamma \\ 0 & 0 \end{bmatrix}$ the Rank is 1.

Moreover, the matrix $\phi(1)$ can be rewritten as:

$$\phi(1) = B \cdot A' = \underbrace{\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{n \times k} \overbrace{\begin{bmatrix} 1 & -\gamma \end{bmatrix}}^{k \times n}$$

$\hookrightarrow$ co-int vector

- **Additional proof that a VAR for Δy_t doesn't exist.**

Recall: $y_t = C\, y_{t-1} + \varepsilon_t$

subtract y_{t-1}: $\Delta y_t = (C - I_n) \cdot y_{t-1} + \varepsilon_t$

can be rewritten as: $\Delta y_t = -\phi(1) \cdot y_{t-1} + \varepsilon_t$

$$= -B \cdot A' \cdot y_{t-1} + \varepsilon_t$$

notice: The differentiated model is explained by the variable in levels (y_{t-1})
In general, when we have co-integration, y_{t-1} brings in information that Δy_t doesn't have. ⇒ Error Correction Mechanism (ECM)

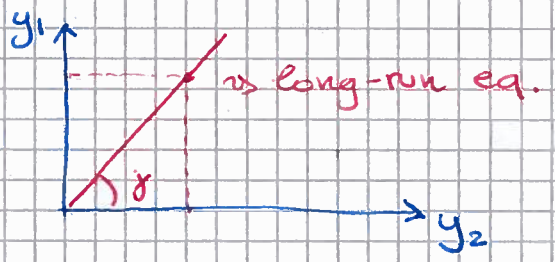
In particular:

$$\Delta y_t = -B \cdot \underbrace{\overset{\text{co-integrating vector}}{A'} \cdot y_{t-1}}_{Z_{t-1}} + \varepsilon_t$$

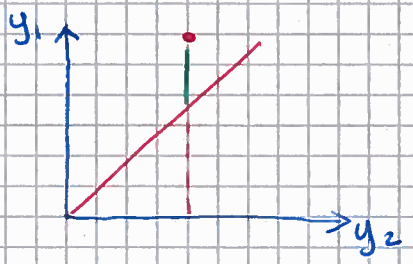
where $Z_t = y_{1,t-1} - \gamma y_{2,t-1}$ ⇒ co-integrated vector

$$\Rightarrow \quad \Delta y_t = \begin{bmatrix} -1 \\ 0 \end{bmatrix} Z_{t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

- In the long-run Δy_{1t} & $\Delta y_{2t} = 0$ (no shocks)
  this implies that $Z_t = 0$
  $\Rightarrow \quad y_{1,t-1} = \gamma\, y_{2,t-1}$



- In the short-run, when a shock happens, we move away from the red line ⇒ $|Z_{t-1}| > 0$



! The red line is the l-r eq.
⇒ any point on the red line is a l-r equilibrium.

notice, since Z_t is a Stationary variable, the system will go back in equilibrium (on the red line) and the initial shock will vanish

important remark. In the long run we only know that the $y_{1t} = \gamma\, y_{2t}$ relation will hold, but we don't know y_{1t} and y_{2t} singularly as they are non-Stationary
⇒ we go back to the red line but we don't know at which point.

- Stock and Watson common trends representation

Recall our VMA form of the differentiated model:

$$\Delta y_t = \Psi(L) \cdot \varepsilon_t$$

We can use the Beveridge-Nelson decomposition to decompose the $\Psi(L)$ into the long-run (trend) non-stat. part and the short-run (cycle) stationary part.

Recall that: $\Psi(L) = \Psi(1) + \Psi^*(L)(1-L)$

Computing, we get:

$$\Psi^*(L) = \begin{bmatrix} 1 & -\gamma \\ 0 & 0 \end{bmatrix} + \overbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} L}^{0} + \overbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} L^2}^{0} + \overbrace{\cdots}^{0}$$

$$\Rightarrow \Delta y_t = \underbrace{\begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix}}_{\Psi(1)} \varepsilon_t + \begin{bmatrix} 1 & -\gamma \\ 0 & 0 \end{bmatrix} (1-L) \cdot \varepsilon_t$$

decomposing into the cycle & trend (i.e taking y_t in levels):

$$y_t = \underbrace{\begin{bmatrix} \gamma \\ 1 \end{bmatrix} \sum_{i=0}^{t} \varepsilon_{2,i}}_{\substack{\text{trend} \\ \text{non-stat.}}} + \underbrace{\begin{bmatrix} 1 & -\gamma \\ 0 & 0 \end{bmatrix} \varepsilon_t}_{\substack{\text{cycle} \\ \text{stationary}}}$$

We can furthermore substitute the forecast errors with the structural shocks to get:

$$y_t = \underbrace{\begin{bmatrix} \gamma \\ 1 \end{bmatrix} \sum_{i=0}^{t} u_{2,i}}_{\text{single permanent component that drives the long run behavior of } y_t \text{; i.e "Random Walk" part.}} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_{1t}$$

$\Psi(1)$

In general:

- In a co-integrated system with k lin. ind. co-int. vectors there are $n-k$ trends.

- notice: By premultiplying the co-integrating vector $[1 \; -\gamma]$ we are left with the stationary part

$$\Rightarrow A' \cdot \Psi(1) = 0$$

∘ Granger Representation Theorem

- Suppose that we have: $\underset{N\times 1}{y_t} \sim I(1)$ & K co-integrations

So, our Wold decomposition is $\Delta y_t = \Psi(L)\cdot \varepsilon_t$.

Then, $\exists$ an $N\times K$ matrix A with linearly independent columns s.t: 1) $A'\cdot y_t = z_t \sim I(0)$

2) $A'\cdot \Psi(1) = 0$

- IF y_t can be written in its VAR form: $\phi(L)\cdot y_t = \varepsilon_t$

Then $\exists$ an $N\times K$ matrix B s.t:

1) $B\cdot A' = \underset{N\times N}{\phi(1)}$

[where, the Rank of $\phi(1)$ is K and the $\det(\phi(z)) = 0$ has $N-K$ unit roots.]

- Moreover, $\exists$ matrices $\zeta_1, \zeta_2 \cdots \zeta_{p-1}$ s.t:

$$\Delta y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} - \underbrace{BA' y_{t-1}}_{= B\cdot z_{t-1}} + \varepsilon_t$$

[Key: A VAR in Δy_t does NOT exist, we can notice by the presence of y_{t-1}.]

Notice, this implies that the VAR polynomial (in levels) $\phi(L)$ can be rewritten as:

$$\phi(L) = (I_n - \rho L) - (\zeta_1 L + \zeta_2 L^2 + \cdots + \zeta_{p-1} L^{p-1})(1-L)$$

where $\rho = \phi_1 + \phi_2 + \cdots + \phi_p$

and the ζ's matrices can be computed as:

$$\zeta_s = -[\phi_{s+1} + \phi_{s+2} + \cdots + \phi_p]$$

note: this is the same decomposition used in the Dickey Fuller derivation.

notice: IF $\rho = I_n \Leftrightarrow \phi_1 + \cdots + \phi_p = I_n \Leftrightarrow \cdots$

$\Rightarrow \phi(1) = I_n - (\phi_1 + \cdots + \phi_p) = 0$

$\Rightarrow$ There are NO cointegrants, the Δy_t VAR has no element in level.

Remarks

- The Co-integration Rank $\Leftrightarrow$ the number of co-integrators is s.t : $0 \leq k \leq n-1$

We can write the VAR + level of Δy_t as:

ECM Form $\rightsquigarrow$ $\zeta(L) \cdot \Delta y_t = -\phi(1) \cdot y_{t-1} + \varepsilon_t$

$$\phi(1) \cdot y_{t-1} = \underbrace{-\zeta(L) \cdot \Delta y_t + \varepsilon_t}_{\text{Stationary } *} = \omega_t$$

* Function of $\Delta y_t \sim I(0)$ and $\varepsilon_t \sim I(0)$

* As a consequence must also be Stationary.

IF $\phi(1)$ was FULL-RANK $\rightsquigarrow$ $\text{Rank}(\phi(1)) = N$

we could invert it to get:

$$y_{t-1} = \underbrace{\phi(1)^{-1} \cdot \omega_t}_{\text{lin. combination of Stationary variables.}}$$

$\Rightarrow$ then y_{t-1} would also be Stationary

Therefore:

- IF $y_t \sim I(1)$ $\Rightarrow$ $\phi(1)$ cannot be invertible
  $\Rightarrow \text{Rank}(\phi(1)) < N$ $\Rightarrow$ $K < N$

- IF $\phi(1)$ is invertible $\Rightarrow$ $y_t \sim I(0)$

- IF $\text{Rank}(\phi(1)) = 0$ $\Rightarrow$ $\phi(1) = 0$ only case which implies that the y_{t-1} term in the ECM gets cancelled out and Δy_t can be written as a VAR $\Rightarrow$ NO Co-integrants.

one extreme : y_t is already Stationary
$\Rightarrow \text{Rank}(\phi(1)) = N$

on the other extreme : y_t has no co-integrators
$\Rightarrow \text{Rank}(\phi(1)) = 0$

i.e <u>$\phi(1)$ creates the need for co-integration</u>

∘ Tests for Co-integration

(Engle-Granger 87')

notice: if co-integration implies that two variable, x_t and y_t, are $I(1)$ but $\exists\, \alpha$ s.t $y_t - \alpha x_t \sim I(0)$ then:

by estimating $y_t = \hat{\alpha} x_t + \hat{\varepsilon}_t$

we should get $\varepsilon_t \sim I(0)$

(Stock 87')

IF x_t and y_t can be co-integrated:

- α converges to the true value at rate T
- to test co-integration run:

$$\Delta \hat{\varepsilon}_t = \rho\, \hat{\varepsilon}_{t-1} + v_t$$

use DF to test $H_0: \rho = 0$ ⇝ non Stationarity

↳ Actually, best to use ADF on:

$$\Delta \hat{\varepsilon}_t = \rho\, \hat{\varepsilon}_{t-1} + a(L)\, \Delta \hat{\varepsilon}_{t-1} + v_t$$

Johansen (88', 91')

problem with original approach:

Suppose we have 3 $I(1)$ variable ⇒ then there might be more than one co-integrating vector. The OLS only finds one (the one yielding the z_t with minimal variance)

Johansen ⇝ idea: we can find a matrix that is the MLE of $\Phi(1)$. The core-idea is to make a likelihood ratio test between a restricted and an unrestricted model.

◦ Johansen Procedure

idea: develop a test to spot the number of co-integrators
i.e the Rank of $\Phi(1)$.

Again, $y_t \sim VAR(p)$ with $y_{ti} \sim I(1)$ but with linear combinations that are Stationary.

1. The model can be written in its ECM Form:

$$\Delta y_t = \Pi_0 y_{t-1} + \Pi(L)\Delta y_{t-1} + \varepsilon_t \quad \text{with } \varepsilon_t \sim N(0, \Omega)$$

given that the residuals are Gaussians, the log likelihood of the parameters can be written as:

$$L(\Omega, \Pi_0, \Pi(L)) =$$

$$= -\frac{T}{2} n \cdot \ln(2\pi) - \frac{T}{2}\ln(|\Omega|) - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t' \cdot \Omega^{-1} \cdot \varepsilon_t$$

→ we want to get MLE for: Π_0, $\Pi(L)$, Ω

2. The procedure involves 3 steps:

1) Auxiliary Regressions: estimate the VARs below.

$$\begin{cases} \Delta y_t = \hat{\Pi}_0 + \hat{\Pi}_1 \Delta y_{t-1} + \dots + \hat{\Pi}_{p-1} \Delta y_{t-p+1} + \hat{u}_t \\ y_{t-1} = \hat{\Theta}_0 + \hat{\Theta}_1 \Delta y_{t-1} + \dots + \hat{\Theta}_{p-1} \Delta y_{t-p+1} + \hat{v}_t \end{cases}$$

2) Canonical Correlations:

find (compute): $\hat{\Sigma}_{vv}$, $\hat{\Sigma}_{uu}$, $\hat{\Sigma}_{uv}$ $(= \hat{\Sigma}_{vu}')$

var-cov of the regressions

$$\left(\frac{1}{T} \cdot \sum^{T} \hat{u}_t \cdot \hat{v}_t'\right)$$

From the matrix: $\left[\hat{\Sigma}_{vv}^{-1} \cdot \hat{\Sigma}_{vu} \cdot \hat{\Sigma}_{uu}^{-1} \cdot \hat{\Sigma}_{uv}\right]$

$n \times n$

- find its eigenvalues $\hat{\lambda}_i$ and eigenvectors $\hat{a}_i$
- Order the eigenvalues: $1 \geq \hat{\lambda}_1 > \dots > \hat{\lambda}_n > 0$
- Normalize the eigenvectors s.t:

$$\hat{a}_i' \cdot \hat{\Sigma}_{vv} \hat{a}_i = 1$$

so to get the normalized eigenvectors $\tilde{a}_i$
(notice: $\tilde{a}_i' \cdot \tilde{a}_i = 1$)

Clearly: $\hat{a}_i = \tilde{a}_i \,./\, \sqrt{\underbrace{\tilde{a}_i'}_{1\times n} \underbrace{\hat{\Sigma}_{vv}}_{n\times n} \underbrace{\tilde{a}_i}_{n\times 1}}$ $\quad 1\times 1$

↳ element-by-element division

- Take the k largest eigenvalues*: $\hat{\lambda}_1 \cdots \hat{\lambda}_k$ which in turn are associated with k largest eigenvectors: $\hat{a}_1 \cdots \hat{a}_k$

* here our hypothesis is that there are k co-i relations.

3) ML Estimates

denote: $\hat{A} = [\hat{a}_1 \cdots \hat{a}_k]_{n \times k}$ ⇝ matrix of co-i vectors

and: $\hat{B} = -\hat{\Sigma}_{uv} \cdot \hat{A}$

Recall that from our ECM we have $\phi(1) = -\xi_0$

Then our estimate for ξ_0 is:

$$-\hat{\xi}_0 = \underbrace{-\hat{\Sigma}_{uv} \hat{A}}_{\hat{B}} \cdot \hat{A}' = \hat{B} \cdot \hat{A}' = \phi(1)$$

The other estimates: $\hat{\xi}_i = \hat{\Pi}_i - \hat{\xi}_0 \hat{\theta}_i$

and $\hat{\Omega} = (1/T) \cdot \sum^{T} (\hat{u}_t - \hat{\xi}_0 \hat{v}_t)'(\,''\,)$

Under the hypothesis (restriction) that there are k co-i relations, the log-likelihood is:

$$L_k^* = -\frac{n}{2} \cdot T \log(2\pi) - \frac{n}{2} T - \frac{T}{2} \ln|\hat{\Sigma}_{uu}| - \frac{T}{2} \sum_{i=1}^{k} \ln(1 - \hat{\lambda}_i)$$

Test 1 (Trace Test)

idea: IF the true number of co-i relations is k then

$$\Rightarrow \hat{\lambda}_{k+1}, \hat{\lambda}_{k+2}, \cdots, \hat{\lambda}_n \approx 0$$

we can then exploit the Statistic:

$$2(L_n^* - L_k^*) = -T \sum_{i=k+1}^{n} \ln(1 - \hat{\lambda}_i) = \lambda_{trace}(k)$$

with the Null: $H_0: \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n = 0$

i.e H_0: There are at most k co-int relations.

and the Alternative H_1: there are n co-i relationships.
⇝ y_t is fully Stationary.

notice: The distribution of the Statistic does not converge to a χ^2, but has been tabulated by Johansen.

Test 2 (Max Test) (Alternative)

Statistic: $2(L^*_{k+1} - L^*_k) = \lambda_{max}(k, k+1) = -T \ln(1-\hat{\lambda}_{k+1})$

here, the Null is: $H_0: \lambda_{k+1} = 0$ (there are k co-i relat.)
and, the Alternative $H_1: \lambda_{k+1} \neq 0$ (There are $k+1$ co-i relat.)

Remark

- In practice we both use the Trace and max test in combination.

∘ Co-integrating Relationships

By now we have only seen one type of co-i relat.:

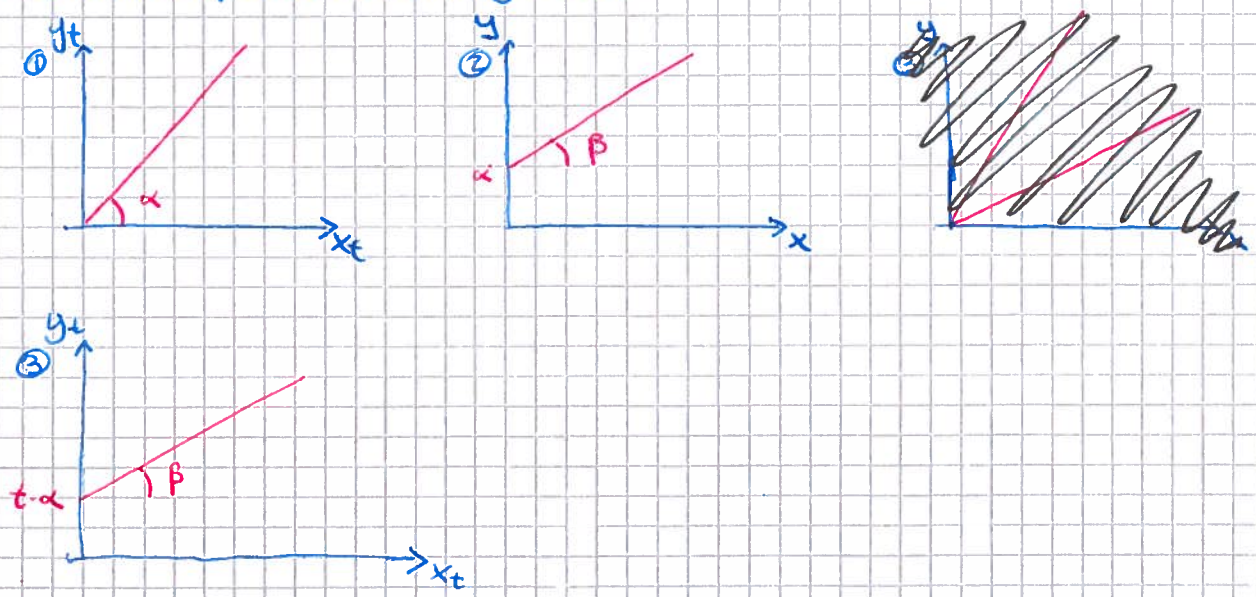
① $$y_t = \alpha x_t + \varepsilon_t \qquad \varepsilon_t \sim I(0)$$

But, in principle, co-i relat. might include both a constant and or a time trend:

② $$y_t = \alpha + \beta x_t + \varepsilon_t$$

③ $$y_t = \alpha \cdot t + \beta x_t + \varepsilon_t$$

The three co-i relat. describe 3 different long-term ($\varepsilon_t = 0$) relationships between y_t and x_t.



◦ Example of Co-integration restriction in RBC

In the RBC we are interested in the co-movement of three main variables: $\{C, I, Y\}$

notice: $\{C, I, Y\} \sim I(1)$ since they are driven by Tech shocks modeled as Random Walk.
(Function of $I(1)$ is itself $I(1)$)

BUT: From Theory we know that:

$$\left.\begin{array}{l} C/Y \sim I(0) \\ I/Y \sim I(0) \end{array}\right. \Rightarrow \left.\begin{array}{l} \log C = \log Y + \varepsilon_t^1 \\ \log I = \log Y + \varepsilon_t^2 \end{array}\right\} k = 2$$

(common trend)

one way to see this is by adopting Stock-Watson decomp. (recall, similar to Beveridge-Nelson) in which:

$$X_t = \underbrace{\Psi(1) \sum \varepsilon_i}_{\text{component driving l-r behavior}} + \text{short-run}$$

notice: Rank($\Psi(1)$) = n. of variables driving the $I(1)$ process

in this case, Tech shocks are the only variables driving $I(1)$.

$\Rightarrow$ Rank($\Psi(1)$) $= n - k = 1$ with $n = 3 \Rightarrow k = 2$

So, we should expect $A' = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$ with $X_t = \begin{pmatrix} \log C \\ \log I \\ \log Y \end{pmatrix}$

From data we get the estimate for A, where:

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

notice that $\exists \infty$ ECM since $\Phi(1) = B \cdot \mathcal{Q} \cdot \underbrace{\mathcal{Q}^{-1} \cdot A'}_{\text{co-i vector}}$

let $\mathcal{Q}^{-1} = \begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \end{bmatrix}$

s.t $\begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_1 \\ 0 & 1 & x_2 \end{bmatrix}$ $\Rightarrow$ Restriction.

*

IF Theory is correct $\Rightarrow x_1 = x_2 = -1$

* $\exists \mathcal{Q}^{-1}$ s.t $\mathcal{Q}^{-1} \cdot \hat{A} = A$ with restriction

○ PRO and CONS of different modeling procedures

Given any y_t, we can decide to adopt 3 distinct approach:

1) VAR in Δy_t (IF $\text{Rank}(\Phi(1)) = 0$)

2) ECM (IF $\text{Rank}(\Phi(1)) = k$ with $0 < k < n$)

3) VAR in y_t (IF $\text{Rank}(\Phi(1)) = n$)

The PROs and CONs of the different approaches given the Rank of $\Phi(1)$ are summarized below:

	FINITE VAR in ΔY	FINITE VAR in Y	ECM
$k=0$	✓	*	≈ ①
$k=n$	≈ ①	✓	≈ ①
$0<k<n$	✗	*	✓

↳ not-consistent

① Model remains consistent but there is a loss in power
In particular:
- an ECM can always be used, when $k=0$ the $\hat{\Phi}(1)$ OLS will be zero (BUT loss in power due to additional useless parameter to be estimated)
- In the other scenarios we are considering Δy_t for Stationary variables ⇒ unmotivated loss of information.

* Sims, Stock & Watson (1990)

↳ A VAR in levels is always consistent!

The ECM has the additional benefit of separating the long-term and short-term behavior of the process, but its not the only correct way to estimate I(1) variables that can be co-integrated.

→ Philips: estimating a VAR in level on I(1) variables can bring inconsistencies in the long-term IRFs →

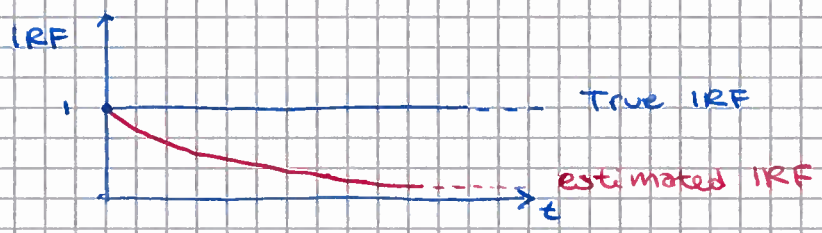
Consider a DGP with a Unit root: $X_t = X_{t-1} + \varepsilon_t$

Suppose we estimate a VAR in levels:

$$X_t = \underline{\underline{\hat{\phi}}} X_{t-1} + \varepsilon_t$$

Recall: The OLS is downward bias! $\Rightarrow$ $\hat{\phi} < 1$

This directly affects the IRFs:



lec.

- ~ Jorda (2005)
- ~ Plagborg - Møller 2022
- ~ Stock - Watson 2018

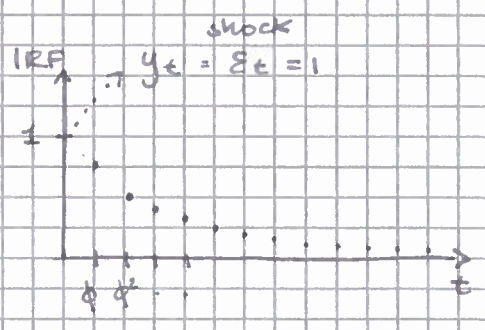
idea: We'll look at alternative ways to compute predictions and impose restrictions for IRFs.

- Linear Projections (LP)

Before: Iterated Forecasts

From, $y_t = \phi y_{t-1} + \varepsilon_t$:

$$y_{t+1|t} = \phi y_t$$
$$y_{t+2|t} = \phi^2 y_t$$
$$\vdots$$
$$y_{t+h|t} = \phi^h y_t$$



here, IRFs to a shock (unitary) in ε_t are computed as:

$$E(y_{t+h} \mid \varepsilon_t = 1, \mathcal{I}_{t-1}) - E(y_{t+h} \mid \varepsilon_t = 0, \mathcal{I}_{t-1})$$

Now: Linear Projection OR Multi-Step Forecast.

idea: directly regress y_{t+h} on y_t, so we get:

$$y_{t+1} = \beta_1 y_t + v_{t+1}$$
$$y_{t+2} = \beta_2 y_t + v_{t+2}$$
$$y_{t+3} = \beta_3 y_t + v_{t+3}$$

notice that we should get equivalent predictions s.t:

$\Longrightarrow$ $\beta_1 = \phi$, $\beta_2 = \phi^2$, $\beta_3 = \phi^3$ . . .

→ IRFs are computed considering a unitary shock in V_t.

Jorda: IRFs computed using linear Projections are more Robust to specifications & can easily accomodate non-linearities + Inference is simplier.
Although this remark might be true in small samples, asymptically the two methods are equivalent.

Remark: One way in which LP have been used has been in computing IRFs of the kind:

$$Y_{t+k} = \mathcal{I}_t(\beta_k^{(1)} Y_t + V^1_{t+k}) + (1-\mathcal{I}_t)(\beta_k^{(2)} Y_t + V^2_{t+k})$$

Indicator Function:
Takes value conditionally on an event (or State)

e.g
IF Recession: $\mathcal{I}_t = 1$
↳ Forecast using $\beta^{(1)}$
IF Boom: $\mathcal{I}_t = 0$
↳ Forecast using $\beta^{(2)}$

- External instrument identification EII

We now move on to restrict the model and find the structural shocks. By now we have used restrictions involving instruments internal to the model:
In the Choleski: Residuals of the short form VAR were assumed to be correlated to the structural shocks and a recursive structure was implemented for identification.
EII: We are given an external variable Z_t (IV) that we know it's correlated to the structural shocks.
notice: EII can be used either in the VAR Framework or in the LP Framework

1) VAR Framework

Begin with: $A(L) \cdot X_t = V_t$ SF VAR

↳ $X_t = C(L) \cdot V_t$ VMA

idea: as usual we need to identify a matrix B s.t:

$V_t = B \cdot \varepsilon_t$ ⇝ Struct. shocks.

For simplicity (it can be generalized) assume that we are interested in the effect of the first shock ε^1_t on the system $\Rightarrow$ to identify ε^1_t we need only the first column of B: B_1

Introduce the external IV:

Consider the IV for ε^1_t, Z_t s.t:

1) Z_t has finite mean $E(Z_t) = \mu$
2) Z_t has finite variance $V(Z_t) = \sigma_z^2$
3) $Corr(\varepsilon^1_t \, Z_t) = \phi \neq 0$; $Corr(\varepsilon^2_t \, Z_t) = 0$

Z_t must be ~~ex~~ have no contemporaneous correlation with the rest of the structural shocks.

Identification of B: let us partition the system s.t:

$$V_t = [B_1 \; B_2] \begin{bmatrix} \varepsilon^1_t \\ \varepsilon^2_t \end{bmatrix} \leftarrow \text{shock of interest}$$

Then:

$$\underbrace{E(V_t \cdot Z_t)}_{\text{known}} = E\Big[\underbrace{[B_1 \; B_2] \begin{bmatrix} \varepsilon^1_t \\ \varepsilon^2_t \end{bmatrix}}_{V_t} Z_t\Big]$$

$$= E\Big[\underbrace{[B_1}_{\text{first column of } B} \; B_2] \underbrace{\begin{bmatrix} \phi \\ 0 \end{bmatrix}}_{Corr(\varepsilon_t \, Z_t)}\Big] = \begin{bmatrix} B_{11} \cdot \phi \\ \vdots \end{bmatrix} = B_1 \cdot \phi$$

we can then normalize the first element of B_1, B_{11} so that:

$$B_{i1} = E(V_{it} \, Z_t) \, / \, E(V_{1t} \, Z_t)$$

By having the first column of B we can easily identify the structural shocks of the system.

Remark. in order to assess the validity of Z_t (and its corr. with ε_{it}) we can run:

$$Z_t = \alpha \cdot \varepsilon_{it} + \partial_t$$

2) LP Framework LP-IV

In the LP Framework we need stronger assumptions for the IV to be valid.

Z_t is s.t: 1) Z_t has finite mean //

2) Z_t has finite variance //

3) $E(\varepsilon_{1t} Z_t) = \phi \neq 0$ $\quad E(\varepsilon_{2t} Z_t) = 0$

Strong exogeneity 4) $E(\varepsilon_{t+j} Z_t) = 0 \quad \forall j \neq 0$

↳ OR LEAD-LAG Exogeneity: we need Z_t to be uncorrelated with all future and past shocks. why?

In the LP framework, IRFs are computed as:

$$Y_{t+h} = \beta_h \underline{Y_t} + \underline{u_{t+h}}$$

notice: by assumption Y_t is determined by the struct. shock at t (ε_t) which in turn is correl. with Z_t.

The additional element u_{t+h} (not present in the VAR fr.) is actually a residual "containing" all the rest (past and future wrt t) of the structural shocks:

~ u_{t+h} is a linear combinations (mean $=0$) of: $\{\varepsilon_{t+h}, \dots, \varepsilon_{t+1}\}$ and $\{\varepsilon_{t-1} \dots\}$

~ If we use Z_t as an IV for $\varepsilon_t \rightsquigarrow Y_t$ then $Z_t \perp u_{t+h}$ i.e uncorr. with lead and lag structural shocks.

Key idea: Y_{t+h} depends on the entire history of shocks, so Z_t is meant to capture only the effect of ε_t.

While it makes sense to assume that Z_t is not correlated with ε_{t+h} with $h>0$ since Z_t only contains variables up to t (think: Z_t is not linearly predictable given future values of ε_t) / it makes less sense to assume that Z_t is not linearly predictable given past values of ε.

↳ Restrictive assumption.

Given Assumption 4 we have that:

From $X_{t+h} = \hat{\beta}_h \tilde{X}_t + u_{t+h}$

where $\tilde{X} = \hat{\gamma}_1 Z_t \quad (X_t = \hat{\gamma}_1 Z_t + \omega_t)$

$$\Rightarrow \beta_h = \frac{E(X_{t+h}\tilde{X}_t)}{E(\tilde{X}_t^2)} = \frac{\cancel{\hat{\gamma}_1} E(X_{t+h} Z_t)}{\hat{\gamma}_1^{\cancel{2}} E(Z_t^2)} = \frac{E(X_{t+h} Z_t)}{\cancel{E(Z_t^2)}} \cdot \frac{\cancel{E(Z_t^2)}}{E(X_t Z_t)} \quad \frac{1}{\hat{\gamma}_1}$$

PROs and CONs

LP - IV	VAR
+ no need for VAR Invertibility	− Invertibility of the VAR.
− LEAD-LAG restriction on IV	+ no strong restriction on IV

CHEAT SHEET

MA(2)

$E(X_t) = 0$

$V(X_t) = \sigma^2 (1 + \theta_1^2 + \theta_2^2)$ same for MA(n)

$\gamma_1 = \sigma^2 (\theta_1 + \theta_1 \theta_2)$

$\gamma_2 = \sigma^2 \theta_2$

$\gamma_3 = 0$

AR(1)

$E(X_t) = 0$

$V(X_t) = (1 - \phi^2)^{-1} \cdot \sigma^2 = \gamma_0$

$\gamma_1 = \phi (1 - \phi^2)^{-1} \cdot \sigma^2$

$\gamma_2 = \phi^2 (1 - \phi^2)^{-1} \cdot \sigma^2$

$\gamma_s = \phi^s (1 - \phi^2)^{-1} \cdot \sigma^2$ $\rightsquigarrow$ converges to zero

AR(2)

$E(X_t) = 0$

$V(X_t) = \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2$

$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1$

$\gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0$

$\vdots$

$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$

Estimated ACF

$$\sqrt{T} (\hat{\rho}_n - \rho_n) \xrightarrow{d} N\left(0, \sqrt{1 + 2 \sum_{j=1}^{n-1} \rho_j^2}\right)$$

Likelihood of AR(1) and MA(1)

$$f(x_1 \cdots x_T ; \theta) = \prod^{T} f(X_t | X_{t-1} ; \theta, \sigma^2)$$

Diagnostic

1 - $H_0: \rho_k = 0 \ \forall k \neq 0$ then $Q = T(T+2) \sum_{k=1}^{s} \hat{\rho}_k^2 (T-k)^{-1}$

$\rightarrow \chi^2_{s-p-q}$

2 - IC: A: $T \cdot \ln(RSS/T) + 2 \cdot (1 + p + q)$

S: $T \cdot \ln(RSS/T) + \ln(T) \cdot (1 + p + q)$

Measures of Forecast precision

1) RMSFE : $\left(\sum_{i=T}^{T+h} (y_i - \hat{y}_i)^2 / h \right)^{\frac{1}{2}}$

2) MAE : $\left(\sum_{i=T}^{T+h} |y_i - \hat{y}_i| / h \right)$