Linear Time Series

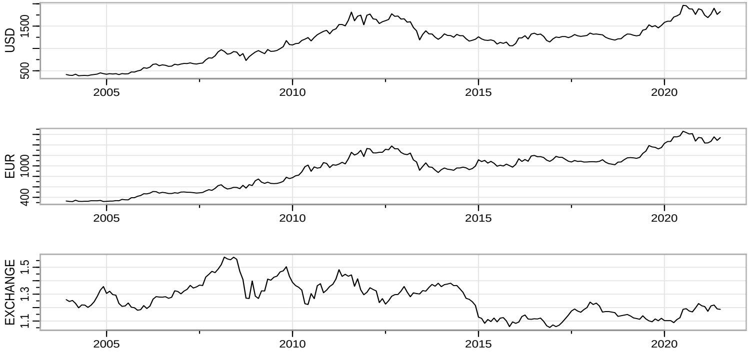
Christian Francq

CREST-ENSAE

Chapter 5: VAR and cointegration

Multivariate time series

A multivariate time series is a collection of several time-dependent variables. Each variable may depend on its past values but also may have some dependency on other variables. ok.



Gold price in USD and EUR, and exchange rate, from December 2003 to July 2021

Questions about multivariate time series

- If I want to explain one variable as function of the others, can I do linear regression "as usual" ?
  - **No**: sometimes regression makes no sense (spurious regression), sometimes it makes sense, but dynamics always matter.
- Which model/technique can I use to analyse and forecast multivariate series ?
  - In practice a Vectorial AutoRegressif (VAR) model, or its VECM (Vector Error Correction Model) form, often suffices. Three Nobel prices: Engle, Granger and Sims on these topics.

Outline

Spurious correlation

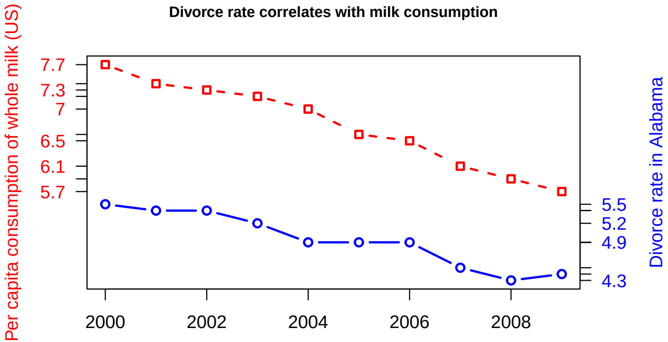
What is correlation? What is causation?

- A positive/negative correlation between 2 random variables X and Y means that they are linearly related and they tend to vary in the same/opposite direction. It does not tell us why the relationship exists.
- X causes Y if X is the cause and Y the effect (smoking causes cancer)

Correlation does not imply causation: examples

- there is a strong correlation between sales of ice cream and sunglasses (that can be due to a third variable–a "mediator variable"–like the temperature);
- CO2 concentration and obesity are empirically correlated (both are incidentally growing in recent years)

Example of spurious regression



The correlation between divorce rate and milk consumption is 0.97

Hundreds of examples of funny spurious correlations/regressions can be found here.

Example of spurious regression

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.78494    0.36594   2.145   0.0643 .
data$Milk     0.61648    0.05404  11.407 3.15e-06 ***

Residual standard error: 0.1111 on 8 degrees of freedom
Multiple R-squared:  0.9421,    Adjusted R-squared:  0.9348
F-statistic: 130.1 on 1 and 8 DF,  p-value: 3.151e-06
```

What is happening ?

- Empirical/sample correlation, beta, R2 always exist, but their theoretical/population counterparts do not when **series are not stationary**.
- CAR : Linear regression outputs are valid for independent and identically distributed (iid) observations.

(X_t, Y_t) iid $\longrightarrow$ pas le cas dès que non stationnaire !

Simulating spurious regressions

Let (ϵ_{1t}) and (ϵ_{2t}) be two independent white noises (iid centred sequences),

$$\begin{cases} X_t = X_{t-1} + a + \epsilon_{1t} = \overbrace{\sum_{i=1}^{t} \epsilon_{1i}}^{\text{tend stoch}} + \overbrace{at + X_0}^{\text{tend. déterm}}, \\ Y_t = Y_{t-1} + b + \epsilon_{2t} = \sum_{i=1}^{t} \epsilon_{2i} + bt + Y_0. \end{cases}$$

Note that (X_t) and (Y_t) are two independent random walks (with drifts if $ab \neq 0$).

ie: $\varepsilon_{1t} \perp\!\!\!\perp \varepsilon_{2t}$

A simulated spurious regression

Note that (X_t) and (Y_t) are two independent random walks.
However, on a simulation of size $n = 100$ of the two series we found $\rho_n(X, Y) = 0.79$ and

corrélation impertinente ?!

s'explique car séries non stationnaires!

on a l'impression que les variables sont liées linéairement.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.40927    0.73151   0.559    0.577
X            0.74337    0.05849  12.708   <2e-16 ***
---
Residual standard error: 2.903 on 98 degrees of freedom
Multiple R-squared:  0.6224,    Adjusted R-squared:  0.6185
F-statistic: 161.5 on 1 and 98 DF,  p-value: < 2.2e-16
```

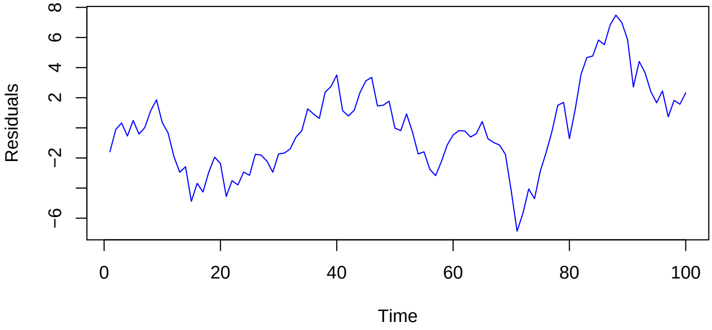
En général régression d'une var. non stat / var non stat → fallacieux.

How can I check a regression is spurious?

The residuals

$$\hat{u}_t = (Y_t - \overline{Y}_n) - \hat{\beta}_n (X_t - \overline{X}_n), \quad t = 1, \ldots, n$$

do not seem stationary:



on ne voit pas de dérive...

How can I check that a regression is spurious?

The residuals do not seem stationary:

```
Augmented Dickey-Fuller Test

data:  resid
Dickey-Fuller = -2.8715, Lag order = 4, p-value = 0.2156
alternative hypothesis: stationary
```

→ Test racine unitaire

assez élevé → pas de raison de rejeter hyp nulle = hyp racine unité ie modèle non stat

ici hyp nulle = hyp stat.

```
KPSS Test for Level Stationarity

data:  resid
KPSS Level = 0.52685, Truncation lag parameter = 4,
p-value = 0.03562
```

→ marginalement rejeté = rejeter hyp de stationnarité

Play around with `sim.spurious.R`

⇒ Résidus non stationnaires.

Regression on the components of a stationary time series

If (X_t, Y_t) is a stationary time series with second-order moments it makes sense to regress Y_t on X_t: ok.

$$Y_t = \beta X_t + c + u_t, \quad \beta = \frac{\mathrm{Cov}(X_t, Y_t)}{\mathrm{Var}(X_t)}, \quad c = EY_t - \beta EX_t.$$

However, using a standard regression routine,

- although $\hat{\beta}_n$ is generally consistent ($\hat{\beta}_n \to \beta$ when $n \to \infty$),
- the **estimated standard deviation** of $\hat{\beta}_n$ and its associated p-value are **wrong**, because the observations are not a sample. (iid)
- Moreover, such a contemporaneous regression is **useless for prediction** (X_t and Y_t are observed at the same time).

On a pas des données iid

On observe les deux à même date … + intéressant de faire des autoreg.

Bivariate Vectorial AutoRegression

modèle autoreg biv

A regression on one lagged value is written

$$\begin{cases} X_t = aX_{t-1} + bY_{t-1} + c + \epsilon_{1t}, \\ Y_t = dY_{t-1} + eX_{t-1} + f + \epsilon_{2t}. \end{cases}$$

également avec passé de l'autre variable.

In vectorial form, this $\boxed{\text{VAR(1)}}$ is written

on régresse sur 1 valeur passée.

$$Z_t = AZ_{t-1} + \nu + \epsilon_t,$$

with

$$Z_t = \begin{pmatrix} X_t \\ Y_t \end{pmatrix}, \quad A = \begin{pmatrix} a & b \\ e & d \end{pmatrix}, \quad \nu = \begin{pmatrix} c \\ f \end{pmatrix}, \quad \epsilon_t = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

If the model is correct, the error term ϵ_t should be a white noise (a centered sequence that does not contain dynamics relationships).
Sims (Economics Nobel price, 2011) popularized the "reduced-form" VAR equation in econometrics.

Times series in $\mathbb{R}^d$

Consider a vector time series:

$$X_t = \begin{pmatrix} X_{1t} \\ \vdots \\ X_{dt} \end{pmatrix}$$

df:

A white noise (WN) is a sequence (ϵ_t) of uncorrelated, centered variables with covariance matrix Σ:

$$E(\epsilon_t) = 0, \quad \text{Var}(\epsilon_t) = \Sigma, \quad \text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0, \text{ for } h \neq 0.$$

↳ matrice.

$\boxed{\text{VAR(1)}}$ models are written under the forms

($\mu = \nu - A\nu$)

$$X_t = \nu + AX_{t-1} + \epsilon_t, \ \Leftrightarrow \ X_t - \mu = A(X_{t-1} - \mu) + \epsilon_t$$ (introduisant la moyenne)

where $\nu, \mu \in \mathbb{R}^d$ are vectors of intercepts, A is the autoregression matrix.

VECM Representation of a VAR(2)

Consider for instance a VAR(2)

$$X_t = \underbrace{A_1 X_{t-1} + A_2 X_{t-2}}_{\text{reg de } X_t / X_{t-1} \text{ et } X_{t-2}} + \nu + \epsilon_t,$$

and let the differentiation operator $\underbrace{\nabla X_t}_{\text{accroissement}} = X_t - X_{t-1}$. We have the VECM

$$\begin{aligned} X_t - X_{t-1} = \nabla X_t &= (A_1 - I_d) X_{t-1} + A_2 X_{t-2} + \nu + \epsilon_t \quad \pm A_2 X_{t-1} \\ &= (A_1 + A_2 - I_d) X_{t-1} + A_2 (X_{t-2} - X_{t-1}) + \nu + \epsilon_t \\ \nabla X_t &:= \boxed{\Pi X_{t-1}} + \Pi_1 \nabla X_{t-1} + \nu + \epsilon_t. \quad \text{(cf D.F augmenté)} \end{aligned}$$

Généralisation :

Regressing the level on p past levels is equivalent to regressing the increment on one level and $p-1$ past increments. It will be seen that the error correction term ΠX_{t-1} have a nice interpretation in non stationary frameworks.

terme à correct° d'erreur.

VECM Representation of a VAR(p)

A VAR(p) model with level

$$X_t = \nu + \sum_{i=1}^{p} A_i X_{t-i} + \epsilon_t \quad \text{or} \quad X_t - \mu = \sum_{i=1}^{p} A_i (X_{t-i} - \mu) + \epsilon_t$$

can be written under the form of a Vector Error Correction Model (VECM) as

$$\nabla X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Pi_i \nabla X_{t-i} + \epsilon_t + \nu,$$

Proof

or

$$\nabla X_t = \Pi (X_{t-1} - \mu) + \sum_{i=1}^{p-1} \Pi_i \nabla X_{t-i} + \epsilon_t,$$

with $\Pi = -I_d + \sum_{i=1}^{p} A_i$ and $\Pi_i = -\sum_{j=i+1}^{p} A_j$.

1

Stationary VAR(1): $\overset{\mathbb{C}^{\mathbb{R}^d}}{X_t} - \mu = \overset{\text{matrice}}{A}(X_{t-1} - \mu) + \epsilon_t$

Quelle condit° pour modèle stationnaire ? Pour univarié il faut $|\phi| < 1$ pour stationnaire causal

$\forall z \leq 1 : 1 - \phi z \neq 0$ ↔ (solut° dépendant de val. prés. et passées de ϵ_t.

If all the eigenvalues of A are strictly less than 1 in modulus, i.e.

$$\text{if} \quad \boxed{\det(I_d - Az) \neq 0, \quad \forall |z| \leq 1,}$$

↳ $\det(\lambda I_n - A) = 0 : \underbrace{\lambda = \frac{1}{z}}_{\text{VP}} : \boxed{|\lambda(A)| < 1}$

$\forall |\lambda| \geq 1$

$\boxed{\det(\lambda I_n - A) \neq 0 \Leftrightarrow \det(I_n - Az) \neq 0 \; |z| \leq 1}$

then

$$X_t = \mu + \sum_{j=0}^{\infty} A^j \epsilon_{t-j}$$

is the stationary and non anticipative solution of the VAR(1) model.

Condition to have a stationary VAR(p)

In the sequel, we assume that the VAR(p) model

$$\boxed{X_t = \nu + \sum_{i=1}^{p} A_i X_{t-i} + \epsilon_t}$$

admits the stationarity and causality condition iff

$$\det\left(I_d - \sum_{i=1}^{p} A_i z^i\right) \neq 0, \quad \forall |z| \leq 1,$$

(the roots of the characteristic polynomial are smaller than 1 in module) which allows to write the model under MA(∞) form.

Recursive relation for the autocovariances

$$X_t = \nu + \sum_{i=1}^{p} A_i X_{t-i} + \epsilon_t$$

Using $\mathrm{Cov}(\epsilon_t, X_t) = \Sigma$ and $\mathrm{Cov}(\epsilon_t, X_{t-h}) = 0$ for all $h > 0$, we get

$$\Gamma(0) = \sum_{i=1}^{p} A_i \Gamma(-i) + \Sigma \quad \text{and} \quad \Gamma(h) = \sum_{i=1}^{p} A_i \Gamma(h-i), \quad \forall\, h > 0.$$

Knowledge of the $\Gamma(0), \ldots, \Gamma(p-1)$ allows to obtain the $\Gamma(h)$ for all h.

ok.

Using $\Gamma(-h) = \Gamma'(h)$, the system (called Yule-Walker equations) yields all the $\Gamma(h)$'s.

VAR(1)

For the VAR(1), we have $\Gamma(h) = A\Gamma(h-1) = A^h\Gamma(0)$ for all $h > 0$ and

$$\Gamma(0) = A\Gamma(-1) + \Sigma = A\Gamma'(1) + \Sigma = A\Gamma(0)A' + \Sigma.$$

To solve this equation, we introduce matrix operators.

Vec operator and Kronecker product

For a matrix A, the vector $\text{vec}A$ is obtained by stacking the columns of A.
For instance,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Rightarrow \quad \text{vec}(A) = \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix}$$

empile les colonnes d'une matrice.

The Kronecker product of 2 matrices is defined by

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1d_2}B \\ \vdots & & \\ a_{d_1 1}B & \cdots & a_{d_1 d_2}B \end{pmatrix}.$$

Properties of the operators vec and $\otimes$

Assuming that the dimensions are compatible,

- $A \otimes B \neq B \otimes A$ in general,
- $(A \otimes B)' = A' \otimes B'$,
- $A \otimes (B + C) = A \otimes B + A \otimes C$,
- if A and B invertible $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, $(A^{-1} \otimes B^{-1})(A \otimes B) = I$
- $(A \otimes B)(C \otimes D) = AC \otimes BD$,
- the eigenvalues of $A \otimes B$ are the products of those of A and B,
- $\operatorname{vec}(A + B) = \operatorname{vec} A + \operatorname{vec} B$,
- $\operatorname{vec} ABC = (C' \otimes A) \operatorname{vec} B$,
- $\operatorname{vec} AB = (I \otimes A) \operatorname{vec} B = (B' \otimes I) \operatorname{vec} A$,
- $(\operatorname{vec} B')' \operatorname{vec} A = \operatorname{tr} AB = \operatorname{tr} BA = (\operatorname{vec} A')' \operatorname{vec} B$,
- $\operatorname{vec} ab' = b \otimes a$ when a and b are 2 vectors.

Variance of a VAR(1)

The variance of a VAR(1) satisfies

$$\rightarrow \quad \Gamma(0) = A\Gamma(0)A' + \Sigma.$$

Using $\operatorname{vec} ABC = (C' \otimes A)\operatorname{vec} B$, we get

$$\operatorname{vec}\Gamma(0) = (A \otimes A)\operatorname{vec}\Gamma(0) + \operatorname{vec}\Sigma,$$

hence

$$\boxed{\operatorname{vec}\Gamma(0) = (I_{d^2} - A \otimes A)^{-1}\operatorname{vec}\Sigma}$$

noting that $I_{d^2} - A \otimes A$ is invertible under the stationarity condition.

$\text{VP} < 1$

or. (donc $\det(I - A\otimes A) \neq 0$
$I - A\otimes A$: inversible

$\det(\lambda I - A) \neq 0 \quad \forall\, |\lambda| \geq 1$ $\quad \boxed{\lambda(A) < 1}$
$(=)$
$\det(I - Az) \neq 0 \quad \forall\, |z| \leq 1$

VAR(1) representation of a VAR(p)

$$X_t = \nu + \sum_{i=1}^{p} A_i X_{t-i} + \varepsilon_t$$

If X_t is a VAR(p), the vector $Y_t = (X_t', \dots, X_{t-p+1}')'$ of size pd satisfies the VAR(1) equation

$$\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix} = \begin{pmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{\begin{pmatrix} A_1 & A_2 & \cdots & A_p \\ I_d & 0 & \cdots & \\ \vdots & & & \\ 0 & \cdots & I_d & 0 \end{pmatrix}}_{\text{↳ mat. compagnon}} \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

that is, $\underline{Y_t = \nu^* + AY_{t-1} + \epsilon_t^*}$.

↪ markovien d'ordre 1.

Autocovariances of a VAR(p)

Since

$$\Gamma_Y(0) = \begin{pmatrix} \Gamma_X(0) & \dots & \Gamma_X(p-1) \\ \vdots & & \\ \Gamma_X(p-1)' & \dots & \Gamma_X(0) \end{pmatrix},$$

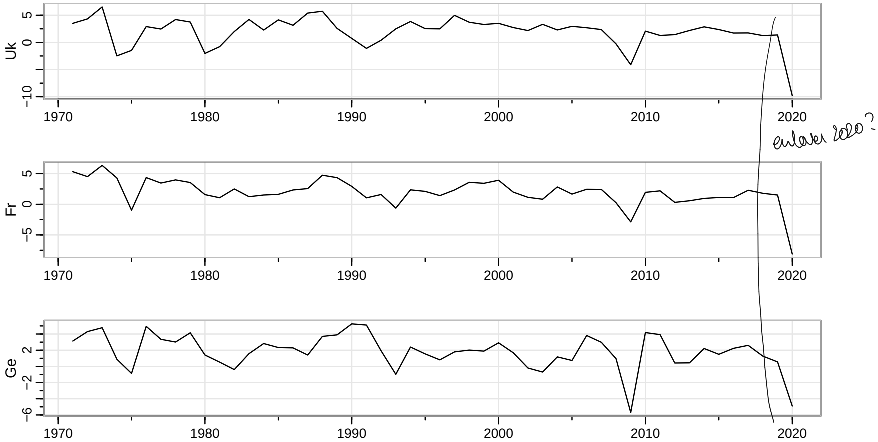
we can obtain $\Gamma_X(0), \dots, \Gamma_X(p-1)$ by → donne tous les $\Gamma_X(h)$ $\forall h \geq p$.

$$\operatorname{vec} \Gamma_Y(0) = (I_{(pd)^2} - A \otimes A)^{-1} \operatorname{vec} \Sigma^*, \quad \Sigma^* = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}.$$

GDP Growth rates of United Kingdom, France and Germany

Stationarity is plausible (when 2020 is omitted)

Tx ↗ du PNB.



Identification: choice of the order p of the VAR(p) model

Based on the observations $X_1,\ldots,X_n$, one can estimate a VAR(p) by OLS and compute the residuals $\hat{\epsilon}_1,\ldots,\hat{\epsilon}_n$.

Identification (i.e. the choice of p) relies on:

1. significance tests on the coefficients,
2. portmanteau tests on the residuals,
3. information criteria

"si tous coeffs de A_p sont petits en peu simplifier ...

$\hat{\varepsilon}_t = X_t - \hat{A}_1 X_{t-1} - \hat{A}_2 X_{t-2} - \hat{A}_3 X_{t-3}$
(VAR(3)) $\longrightarrow$ modèle correct : autocov ~ 0 .

(prend compte multi-variée) .

$$\mathrm{AIC}(p) = \log\det\hat{\Sigma}(p) + \frac{2}{n}pd^2 \text{ or } \mathrm{BIC}(p) = \log\det\hat{\Sigma}(p) + \frac{\log n}{n}pd^2$$

with

$$\hat{\Sigma}(p) = \frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t\hat{\epsilon}_t'.$$

Identification of the GDP Growth rates

Using the R package vars

```
> VARselect(GDP)

$selection
AIC(p)  HQ(p)  SC(p) FPE(p)
1       1      1     1

$criteria
                1          2          3          4          5          6
7
AIC(p)   2.901209   3.045274   3.192804   3.399094   3.583011   3.723848   3.954257
HQ(p)    3.084403   3.365863   3.650788   3.994473   4.315786   4.594018   4.961822
SC(p)    3.407873   3.931936   4.459463   5.045752   5.609667   6.130501   6.740908
FPE(p) 18.232783  21.246540  25.160632  32.215872  41.467292  53.138023  78.613632
                8          9          10
AIC(p)   3.943299   3.519457    3.698421
HQ(p)    5.088259   4.801812    5.118172
SC(p)    7.109948   7.066104    7.625066
```

Estimation of the VAR paramater

Estimation is obtained by OLS, equation-by-equation

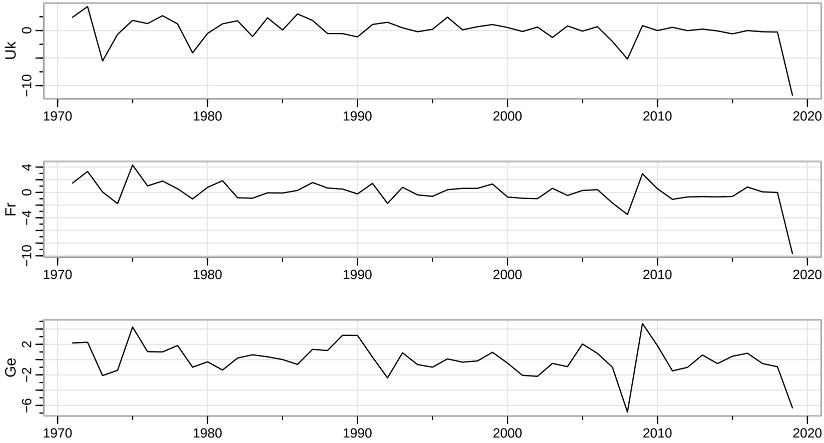
```
VAR Estimation Results:
=========================
Endogenous variables: Uk, Fr, Ge
Deterministic variables: const
Sample size: 49
Log Likelihood: -275.412
Roots of the characteristic polynomial:
0.4407 0.4407 0.2521   |.|<2
Call:
VAR(y = GDP, lag.max = 1)


Estimation results for equation Uk:
===================================
Uk = Uk.l1 + Fr.l1 + Ge.l1 + const

        Estimate Std. Error t value Pr(>|t|)
Uk.l1    0.5801     0.2166   2.678   0.0103 *
Fr.l1   -0.1453     0.3499  -0.415   0.6799
Ge.l1   -0.2836     0.2742  -1.034   0.3065
const    1.5410     0.6254   2.464   0.0176 *
```

Estimation of the VAR paramater

```
Fr = Uk.l1 + Fr.l1 + Ge.l1 + const

       Estimate Std. Error t value Pr(>|t|)
Uk.l1   0.36107    0.16993   2.125   0.0391 *
Fr.l1   0.24168    0.27449   0.880   0.3833
Ge.l1  -0.08016    0.21507  -0.373   0.7111
const   0.72993    0.49061   1.488   0.1438

Estimation results for equation Ge:
==========================================
Ge = Uk.l1 + Fr.l1 + Ge.l1 + const

       Estimate Std. Error t value Pr(>|t|)
Uk.l1    0.2137     0.1825   1.171   0.2477
Fr.l1   -0.0389     0.2947  -0.132   0.8956
Ge.l1    0.1538     0.2309   0.666   0.5088
const    1.1000     0.5268   2.088   0.0425 *

Covariance matrix of residuals:
      Uk    Fr    Ge
Uk 6.605 4.000 3.375
Fr 4.000 4.064 3.412
Ge 3.375 3.412 4.686
```

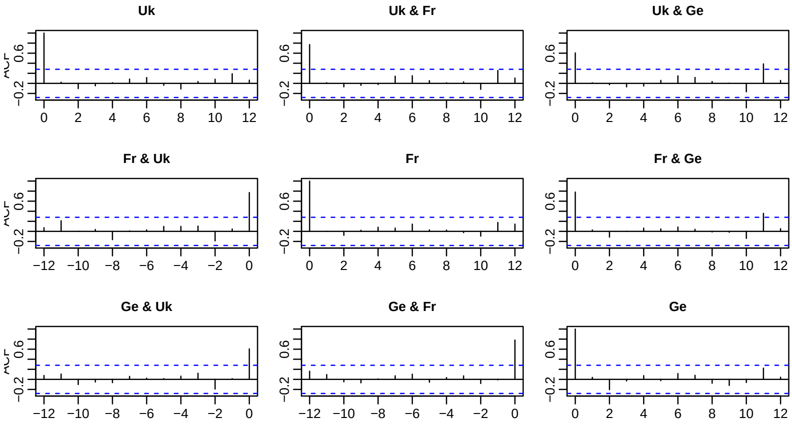
Model adequacy

The residuals should resemble a white noise (2019 is an outlier)



Model adequacy

The residuals should resemble a white noise

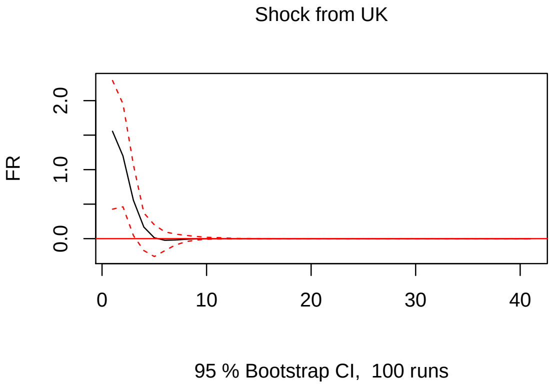


Model adequacy

For model adequacy checking we can perform some diagnostic tests on residuals:

```
var<-VAR(GDP, lag=1)
LjungBox(var)

 lags  statistic  df   p-value
    5   25.71514  36 0.8981230
   10   65.25647  81 0.8987188
   15   99.95438 126 0.9579114
   20  147.09508 171 0.9068847
   25  174.13662 216 0.9833597
   30  245.23738 261 0.7501546
```

Impulse response functions



The effect of a shock from Uk_t (Uk_t replaced by $\mathrm{Uk}_t + \tau$) on Fr_{t+h} disappears after 5 lags ($h \geq 5$).

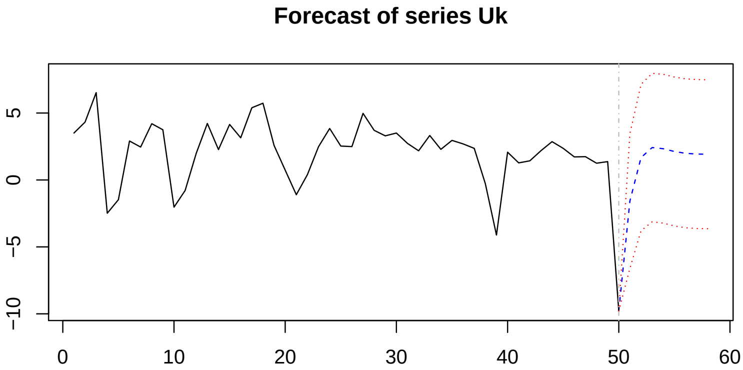
Point prediction

Suppose the AR is known, and the WN (ϵ_t) is iid with variance Σ.
Let $\hat{X}_{t+h|t}$ the optimal prediction at horizon $h > 0$ of X_{t+h} given X_t and its past values.
Since ϵ_{t+h} is independent from $\{X_{t-i}, i \geq 0\}$, we have

$$\hat{X}_{t+h|t} = \nu + \sum_{i=1}^{p} A_i \hat{X}_{t+h-i|t}, \quad \text{with } \hat{X}_{t+h-i|t} = X_{t+h-i} \text{ when } i \geq h.$$

Prediction region

Prediction



Totally wrong predictions! (at least for 2020)

Integrated vectorial time series

A stationary or trend-stationary time series is said to be $I(0)$.

A nonstationary series that can be made $I(0)$ by differencing once is said to be $I(1)$, **integrated of order 1**.
For $X_t \in \mathbb{R}^d$, we have

$$(X_t) \sim I(1) \text{ if and only if } (X_t - X_{t-1}) \sim I(0)$$

If a series needs to be differenced k times to be stationary, it is said to be $I(k)$.

Examples:

- $X_t = \epsilon_t \sim I(0)$, $X_t = \epsilon_t + at + b \sim I(0)$, $X_t \sim \text{ARMA}(p,q) \sim I(0)$;
- $X_t = \sum_{i=1}^{t} \epsilon_i + at + b \sim I(1)$, $X_t \sim \text{ARMA}(p,d,q) \sim I(d)$.

Note that $I(1)$ series contain **stochastic trends**, for example $\sum_{i=1}^{t} \epsilon_i$, and $I(0)$ series do not.

Cointegration

An important property of I(1) series is that there can exist **linear combinations** of these series that are **I(0)**. If it is so, these variables are called cointegrated.

In other words, when series contain a common stochastic trend, they are cointegrated.

Then there exist long-run relationships between these variables (*i.e.* stationary combinations of these variables).

Cointegration was introduced by Granger (1981) and Engle and Granger (1987). [Nobel price in 2003]

Example

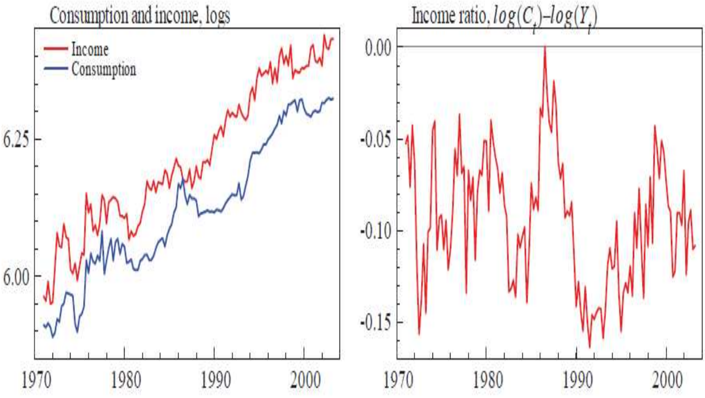


Figure: Relation income-consumption

Notation $I(k)$

$X_t \sim I(0)$ means (X_t) stationary

$\nabla X_t = (1-B)X_t = X_t - X_{t-1}, \quad \nabla^2 X_t = (1-B)^2 X_t = X_t - 2X_{t-1} + X_{t-2}, \ldots$

Integrated process of order $k \geq 1$

$$X_t \sim I(k) \quad \text{iff} \quad \nabla^{k-1} X_t \nsim I(0) \quad \text{and} \quad \nabla^k X_t \sim I(0).$$

Integrated process of order k in the Granger sense

$$X_t \sim I(k) \quad \text{iff} \quad \nabla^k X_t \sim I(0)$$

where $X_t \sim I(0)$ iff $X_t = EX_t + C(B)\epsilon_t$ with ϵ_t a WN and $C(z) = \sum_{i=1}^{\infty} C_i z^i$ such that $C(1) \neq 0$.

With this latter definition a trend-stationary process is $I(0)$.

Cointegration relationships

Cointegrated series

Let (X_t) a **vector** $I(1)$ process.

The components of X_t are cointegrated if there exists a vector $v \neq 0$, called **cointegration vector**, such that $v'X_t$ is $I(0)$.

$v'X_t$ is called a cointegration (or long-run) relationship.

Cointegration rank

The **cointegration rank** is the number of linearly independent cointegration vectors.

Economic examples

- (C_t) and (R_t): consumption and income
  The ratio C_t/R_t is often stationary. Thus $\log(C_t) - \log(R_t)$ is stationary, i.e. $v = (1, -1)$ is a cointegration vector for $X_t = (\log C_t, \log R_t)'$.
- P_t and P_t^*: prices of the same product in 2 countries
  C_t: exchange rate

  Thus P_t should be approximately equal to $C_t P_t^*$, that is,

  $$\log P_t \approx \log C_t + \log P_t^*.$$

  A more realistic assumption is

  $$\log P_t = \log C_t + \log P_t^* + I(0),$$

  $\Rightarrow v' X_t$ stationary, with $v = (1, -1, -1)$ and $X_t = (\log P_t, \log C_t, \log P_t^*)'$.

Example with a common stochastic trend

Let $(\epsilon_t, \epsilon_{1\,t}, \epsilon_{2\,t})'$ a WN. The variables

$$\begin{cases} X_{1\,t} &= a\sum_{i=1}^{t} \epsilon_i + \epsilon_{1\,t} \\ X_{2\,t} &= b\sum_{i=1}^{t} \epsilon_i + \epsilon_{2\,t} \end{cases}$$

are cointegrated when $ab \neq 0$. The long-run relationships are multiple of $bX_{1\,t} - aX_{2\,t}$.

Example with cointegration rank $r = 2$

$$\begin{cases} X_{1\,t} &= aX_{2\,t} + \epsilon_{1\,t} \\ X_{2\,t} &= bX_{3\,t} + \epsilon_{2\,t} \\ X_{3\,t} &= X_{3\,t-1} + \epsilon_{3\,t} \end{cases}$$

Non stationary (3rd component is a RW).
The first 2 equations are cointegration relationships. The cointegration rank is 2.

Writing $bX_{3\,t} = bX_{3\,t-1} + b\epsilon_{3\,t}$ in the 2nd eq., and $aX_{2\,t} = abX_{3\,t-1} + ab\epsilon_{3\,t} + a\epsilon_{2\,t}$ in the 1st eq., we get the AR(1) representation

$$X_t = \begin{pmatrix} 0 & 0 & ab \\ 0 & 0 & b \\ 0 & 0 & 1 \end{pmatrix} X_{t-1} + \begin{pmatrix} \epsilon_{1\,t} + a\epsilon_{2\,t} + ab\epsilon_{3\,t} \\ \epsilon_{2\,t} + b\epsilon_{3\,t} \\ \epsilon_{3\,t} \end{pmatrix}.$$

Assumption on the VAR polynomial

Let a $I(1)$ process satisfying an AR(p) representation

$$A(B)X_t = \epsilon_t \quad t \geq 1, \quad A(z) = I_d - \sum_{i=1}^{p} A_i z^i$$

(with initial values independent of the WN) such that $\det A(1) = 0$ and

$$\det A(z) = 0 \quad \Rightarrow \quad |z| > 1 \text{ or } z = 1.$$

In the VECM form,

$$\nabla X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Pi_i \nabla X_{t-i} + \epsilon_t$$

the matrix $\Pi = -(I_d - \sum_{i=1}^{p} A_i) = -A(1)$ is singular.

Rank of Π

$r =$ **rank of Π**

- If $r = 0$, i.e. $\Pi = 0$, then ∇X_t is an AR$(p-1)$ and the variables are not cointegrated.

  Indeed, if $Z_t = a'X_t$ with $a \neq 0$ then $\nabla Z_t = a'\Phi^{-1}(B)\epsilon_t \sim I(0)$ and it can be shown that $Z_t \nsim I(0)$.

Rank of Π

- If $r \in \{1, \ldots, d-1\}$ then there exists matrices α and β of size $d \times r$ and full rank r such that

$$\Pi = \alpha\beta'.$$

The vector ΠX_{t-1} is stationary, because

$$\Pi X_{t-1} = \nabla X_t - \sum_{i=1}^{p-1} \Pi_i \nabla X_{t-i} + \epsilon_t,$$

is a function of the stationary process $(\nabla X_t, \epsilon_t)$.

Rank of Π and cointegration rank

The cointegration relations are in the columns of β since

$$\beta' X_t \text{ is } I(0).$$

Indeed

$$\beta' X_t = (\alpha'\alpha)^{-1}\alpha'\alpha\beta' X_t = (\alpha'\alpha)^{-1}\alpha'\Pi X_t$$

and ΠX_t is stationary.

α is called the loading matrix, the Π_i are the short-run parameters, and ΠX_{t-1} is the long-run component.
VECM form:

$$\nabla X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{p-1} \Pi_i \nabla X_{t-i} + \epsilon_t$$

Example with $r = 1$

If $r = 1$, α and β are vectors.

VECM form:

$$\nabla X_t = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix} \underbrace{(\beta_1, \ldots, \beta_d) X_{t-1}}_{\text{stationary}} + \sum_{i=1}^{p-1} \Pi_i \nabla X_{t-i} + \epsilon_t$$

There is one cointegration relationship between the components of X_t given by

$$\beta_1 X_{1t} + \cdots + \beta_d X_{dt} \sim I(0)$$

Non uniqueness of α and β

The matrices α and β are not unique in the decomposition $\Pi = \alpha\beta'$. To define them in a unique way, one can for instance choose β of the form

$$\beta' = \left[I_r \; : \; \tilde{\beta}'_{r\times(d-r)} \right].$$

A cointegration matrix of this form exists provided the order of the variables is appropriately chosen. Example

Estimation of the VECM

The estimators of the short-run parameters converge at the usual rate $\sqrt{n}$

For known orders p and r, the VECM can be estimated by Gaussian Maximum Likelihood.

Under general assumptions, it can be shown that

$$\sqrt{n}\left\{\text{vec}\left(\left[\hat{\Pi}_1 - \Pi_1, \ldots, \hat{\Pi}_{p-1} - \Pi_{p-1}\right]\right)\right\} \xrightarrow{\mathscr{L}} \mathscr{N}(0, \Sigma)$$

and

$$\sqrt{n}\left\{\text{vec}\left(\hat{\Pi} - \Pi\right)\right\} \xrightarrow{\mathscr{L}} \mathscr{N}(0, \Sigma_\Pi)$$

where Σ_Π is singular.

Estimation of the cointegration relationships

Estimation at rate n

The ML estimator of the parameter $\tilde{\beta}_{r\times(d-r)}$ is super-consistent:

$$n\left\{\text{vec}\left(\widehat{\tilde{\beta}}_{r\times(d-r)} - \tilde{\beta}_{r\times(d-r)}\right)\right\} \xrightarrow{\mathscr{L}} \mathscr{N}(0, \tilde{\Sigma}).$$

The ML estimator of α has the classical rate of convergence $\sqrt{n}$.

Johansen test

The Johansen test allows to test successively the hypotheses

$$H_0 : \text{rank}\,(\Pi) = r_0 \quad \text{against} \quad H_0 : \text{rank}\,(\Pi) > r_0$$

for $r_0 = 0, 1, \ldots, d-1$.
The first value r_0 which cannot be rejected is selected.

Example

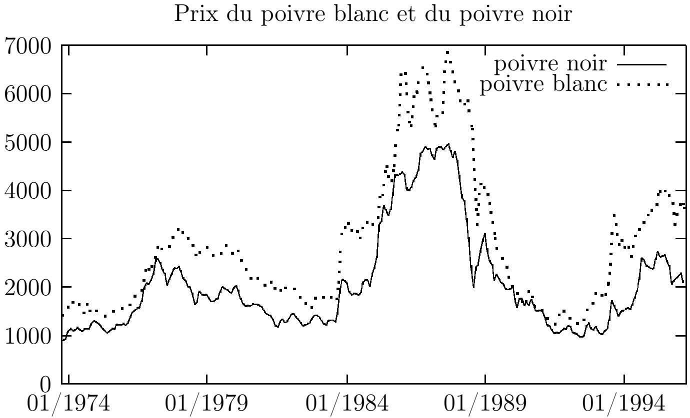


FIGURE 1.5 – Prix mensuel européen du poivre blanc et du poivre noir de 1973.10 à 1996.04 en dollars US par million de tonnes

The components are I(1)

The ADF test does not reject the UR hypothesis on the black pepper series. The UR hypothesis is rejected on the differenced series.
Same conclusion for the white pepper series: I(1).

```
ADF Test for series:       Blackpepper
sample range:              [1974 M1, 1996 M4], T = 268
lagged differences:        2
intercept, no time trend
asymptotic critical values
1%          5%          10%
-3.43       -2.86       -2.57
value of test statistic: -1.8635
```

Cointegration is not rejected

The hypothesis $H_0 : \text{rank}(\Pi) = 0$ (absence of cointegration) can be rejected, but not the hypothesis $H_0 : \text{rank}(\Pi) = 1$.

```
Johansen Trace Test for:  Blackpepper Whitepepper
sample range:             [1973 M12, 1996 M4], T = 269
included lags (levels):   2
dimension of the process: 2
intercept included
response surface computed:
-----------------------------------------------
r0  LR        pval      90%       95%       99%
-----------------------------------------------
0   21.88     0.0278    17.98     20.16     24.69
1   3.59      0.4873    7.60      9.14      12.53
```

Estimated VEC(1)

The VEC(1) model is

$$\begin{pmatrix} \nabla \mathrm{BP}_t \\ \nabla \mathrm{WP}_t \end{pmatrix} = \begin{pmatrix} -0.017 \\ 0.144 \end{pmatrix} (1 \quad -0.742) \begin{pmatrix} \mathrm{BP}_{t-1} \\ \mathrm{WP}_{t-1} \end{pmatrix} + \begin{pmatrix} 0.387 & 0.012 \\ 0.463 & 0.145 \end{pmatrix} \begin{pmatrix} \nabla \mathrm{BP}_{t-1} \\ \nabla \mathrm{WP}_{t-1} \end{pmatrix} + \begin{pmatrix} 1.316 \\ 11.966 \end{pmatrix} + \begin{pmatrix} \epsilon_{1\,t} \\ \epsilon_{2\,t} \end{pmatrix}.$$

In this model, the signs of the loading coefficients can be interpreted.

Long-run relationship: $\mathrm{BP}_t - 0.74\mathrm{WP}_t \sim I(0).$

Regression (= danger! in general ....)

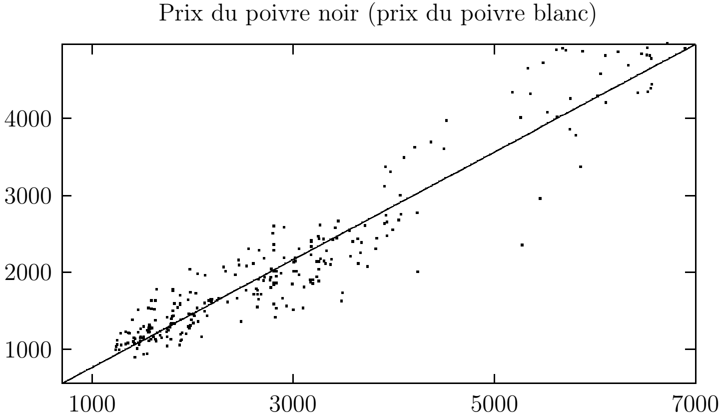


FIGURE 4.1 – Prix mensuel du poivre blanc et du poivre noir de 1973.10 à 1996.04. La droite de régression est PN=0.7003PB+63.5433

Regression on the components of a multivariate time series

- If the multivariate time series is stationary and ergodic the OLS estimator is consistent, but its asymptotic distribution differs from the iid case.
- If a cointegration relationship exits, the OLS estimator of this long-run relationship is super-consistent.
- If the series is I(1) though not cointegrated, we have a **spurious regression** (see the example given in Chapter 1).

Informal definition

The concept of causality was introduced by Granger (1969).

A variable X_t is said to **cause** another variable Y_t in the Granger sense if X_t is useful to predict Y_t.

Should not be interpreted as a causality in the usual sense (not a "cause"!).

For instance, the variable "consumption" may well cause in the Granger sense the variable "income".

Causality in the Granger sense

Let $\hat{Y}_{|Z_1,Z_2,\dots}$ the linear optimal prediction (L^2 sense) of a variable Y given variables $Z_1, Z_2, \dots$.

(X_t) is said to **cause (Y_t) in the Granger sense** if and only if

$$\hat{Y}_{t+h|\{X_u, Y_u,\, u \leq t\}} \neq \hat{Y}_{t+h|\{Y_u,\, u \leq t\}}$$

for at least some $h > 0$ and some $t \in \mathbb{Z}$.

Thus (X_t) does not cause (Y_t) if the past of X_t is useless to predict Y_{t+h} at any horizon.

Stationary series

When (X_t, Y_t) is stationary, (X_t) causes (Y_t) at horizon h iff

$$\hat{Y}_{t+h|\{X_u, Y_u, u \leq t\}} \neq \hat{Y}_{t+h|\{Y_u, u \leq t\}}.$$

Notation: $C^{(h)}_{X \to Y}$.

(X_t) causes (Y_t) if $C^{(h)}_{X \to Y}$ for at least one horizon $h > 0$.

Notation: $C_{X \to Y}$

Instantaneous causality in Granger sense

If X_{t+1} is useful to predict Y_{t+1} at time t, then (X_t) is said to cause (Y_t) instantaneously.

Let $Z_t = (X_t', Y_t')'$ a stationary series. (X_t) is said to cause (Y_t) instantaneously iff

$$\hat{Y}_{t+1|\{X_u, Y_u, u \leq t\} \cup \{X_{t+1}\}} \neq \hat{Y}_{t+1|\{X_u, Y_u, u \leq t\}}.$$

Definition based on the correlations of the errors

Let the innovation

$$\epsilon_t = Z_t - \hat{Z}_{t|\{Z_u,\, u<t\}} = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = \begin{pmatrix} X_t - \hat{X}_{t|\{X_u, Y_u,\, u \leq t-1\}} \\ Y_t - \hat{Y}_{t|\{X_u, Y_u,\, u \leq t-1\}} \end{pmatrix}.$$

Using the fact that $\hat{X}_{|\{Y,Z\}} = \hat{X}_{|Z}$ iff X and Y are uncorrelated conditionally on Z (*i.e.* $X - \hat{X}|Z$ and $Y - \hat{Y}|Z$ are non correlated), we get:

(X_t) does not cause (Y_t) instantaneously iff ϵ_{1t} and ϵ_{2t} are non correlated.

The concept of instantaneous causality is thus symmetric.

Notation: C_{X-Y}

Example

Let (ϵ_t) and (η_t) 2 independent non degenerate WN, and let

$$\begin{cases} X_t = \epsilon_t + a\eta_t + b\eta_{t-1} \\ Y_t = \eta_t \end{cases}$$

We have $\hat{X}_{t|\{X_u, Y_u, u<t\}} = bY_{t-1}$ (which is not a function of $\{X_u, u < t\}$) hence

$$C_{Y \to X} \quad \text{iff} \quad b \neq 0.$$

But $C_{X \not\to Y}$ and $C^{(h)}_{Y \not\to X}$ for $h > 1$.
We have $\hat{X}_{t|\{X_u, Y_u, u<t\}, Y_t} = aY_t + bY_{t-1}$, thus

$$C_{Y-X} \quad \text{iff} \quad a \neq 0.$$

Prediction of the components of a VAR

Let

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{p} \begin{pmatrix} A_i^{XX} & A_i^{XY} \\ A_i^{YX} & A_i^{YY} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

then

$$\hat{Y}_{t|\{X_u, Y_u, u<t\}} = \sum_{i=1}^{p} A_i^{YY} Y_{t-i} + \sum_{i=1}^{p} A_i^{YX} X_{t-i}.$$

Characterization through the VAR coefficients

If the variance-covariance matrix Σ of the WN is non-singular,

$$C_{X\nrightarrow Y} \quad \text{iff} \quad A_i^{YX} = 0 \quad \text{for } i = 1, \ldots, p$$

$$C_{Y\nrightarrow X} \quad \text{iff} \quad A_i^{XY} = 0 \quad \text{for } i = 1, \ldots, p$$

$$C_{Y\neq X} \quad \text{iff} \quad \Sigma^{XY} := \mathrm{Cov}(\epsilon_{1t}, \epsilon_{2t}) = 0$$

Example: Trivariate AR(1)

$$\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} 1/2 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1/2 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 1/4 & 0 \\ 1/4 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus

$$C_{(Y,Z)\not\to X}, \quad C_{X\to(Y,Z)}, \quad C_{X-(Y,Z)},$$
$$C_{Z\not\to(X,Y)}, \quad C_{(X,Y)\to Z}, \quad C_{Z\not-(X,Y)}.$$

Example (continued)

Writing

$$\begin{pmatrix} Y_t \\ X_t \\ Z_t \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 0 \\ 1 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \\ Z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{2t} \\ \epsilon_{1t} \\ \epsilon_{3t} \end{pmatrix}, \quad \tilde{\Sigma} = \begin{pmatrix} 1 & 1/4 & 0 \\ 1/4 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

we see that

$$C_{(X,Z)\to Y}, \quad C_{Y\to(X,Z)}, \quad C_{Y-(X,Z)}.$$

Example (end)

From the equation

$$Z_t = 1/2 Z_{t-1} + Y_{t-1} + \epsilon_{3t}$$

we see that X is useless to predict Z_t at horizon 1, that is, $C^{(1)}_{X \nrightarrow Z}$.

On the contrary, since $A_1^2 = \begin{pmatrix} 1/4 & 0 & 0 \\ 1/2 & 0 & 0 \\ 1 & 1/2 & 1/4 \end{pmatrix}$, we have

$$Z_t = X_{t-2} + 1/2 Y_{t-2} + 1/4 Z_{t-2} + 1/2 \epsilon_{3t-1} + \epsilon_{2t-1} + \epsilon_{3t}$$

and $C^{(2)}_{X \to Z}$, thus $C_{X \to Z}$. This shows that it is not easy to characterize the non-causality of a component w.r.t. another one, when other components exist.

Null assumption of non causality

Let $X_t \in \mathbb{R}^{d_1}$ and $Y_t \in \mathbb{R}^{d_2}$ and suppose $(X_t', Y_t')'$ follows a VAR(p).

Under regularity assumptions (in particular stationarity) the LS estimator of $\underline{A} = [A_1, \ldots, A_p]$ satisfies

$$\sqrt{n}\left(\mathrm{vec}\,\hat{\underline{A}} - \mathrm{vec}\,\underline{A}\right) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \Sigma_A).$$

The non-causality assumption from X to Y writes

$$H_0 : R\mathrm{vec}\,\underline{A} = 0$$

where R is a matrix of size $pd_1d_2 \times p(d_1+d_2)^2$ which contains only 0 or 1.

▸ VAR(1) example

Wald test

$$\sqrt{n}\left(\mathrm{vec}\underline{\hat{A}} - \mathrm{vec}\underline{A}\right) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \Sigma_A).$$

Under $H_0 : R\mathrm{vec}\underline{A} = 0$ we thus have

$$\sqrt{n}R\mathrm{vec}\underline{\hat{A}} \xrightarrow{\mathscr{L}} \mathscr{N}\left(0, R\Sigma_A R'\right)$$

and, if $\hat{\Sigma}_A$ is any consistent estimator of Σ_A, the quadratic form

$$S = n(\mathrm{vec}\underline{\hat{A}}')R'(R\hat{\Sigma}_A R')^{-1}R\mathrm{vec}\underline{\hat{A}}$$

follows approximately a $\chi^2_{pd_1 d_2}$.

Wald test at level α: the assumption that X does not cause Y is rejected if

$$S > \chi^2_{pd_1 d_2}(1-\alpha),$$

where $\chi^2_k(1-\alpha)$ is the $(1-\alpha)$−quantile of χ^2_k.

Example

On the series of prices of black and white pepper, we fitted an AR(3):

$$\begin{pmatrix} WP_t \\ BP_t \end{pmatrix} = \begin{pmatrix} 33.148 \\ 30.618 \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} + \begin{pmatrix} 1.066 & 0.617 \\ 0.036 & 1.362 \end{pmatrix} \begin{pmatrix} WP_{t-1} \\ BP_{t-1} \end{pmatrix}$$
$$+ \begin{pmatrix} -0.258 & -0.508 \\ -0.070 & -0.377 \end{pmatrix} \begin{pmatrix} WP_{t-2} \\ BP_{t-2} \end{pmatrix} + \begin{pmatrix} 0.085 & 0.024 \\ 0.046 & -0.014 \end{pmatrix} \begin{pmatrix} WP_{t-3} \\ BP_{t-3} \end{pmatrix}$$

Results of the causality tests:

- No significant causality from white to black.

- Causality from black to white and also instantaneous causality.

Causality tests

```
TEST FOR GRANGER-CAUSALITY:
H0: "WHITEPEPPER" do not Granger-cause "BLACKPEPPER"
Test statistic l = 1.1110
pval-F( l; 2, 546) = 0.3300

TEST FOR GRANGER-CAUSALITY:
H0: "BLACKPEPPER" do not Granger-cause "WHITEPEPPER"
Test statistic l = 13.6919
pval-F( l; 2, 546) = 0.0000

TEST FOR INSTANTANEOUS CAUSALITY:
H0: No instantaneous causality between "BLACKPEPPER" and "WHITEP
Test statistic: c = 80.0983
pval-Chi( c; 1) = 0.0000
```

End of the course ☺ !

Example: bivariate AR(2) with $r = 1$

$$\begin{pmatrix} \nabla X_{1\,t} \\ \nabla X_{2\,t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1, \tilde{\beta}) \begin{pmatrix} X_{1\,t-1} \\ X_{2\,t-1} \end{pmatrix} + \begin{pmatrix} \Pi_1(1,1) & \Pi_1(1,2) \\ \Pi_1(2,1) & \Pi_1(2,2) \end{pmatrix} \begin{pmatrix} \nabla X_{1\,t-1} \\ \nabla X_{2\,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1\,t} \\ \epsilon_{2\,t} \end{pmatrix}.$$

Adding a stationary AR(1) component $X_{3\,t}$ independent from X_1 and X_2 yields

$$\nabla X_t = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} 1 & \tilde{\beta} & 0 \\ 0 & 0 & 1 \end{pmatrix} X_{t-1} + \Pi \nabla X_{t-1} + \epsilon_t$$

where $\alpha_{11} = \alpha_1$, $\alpha_{21} = \alpha_2$, $\alpha_{12} = \alpha_{22} = \alpha_{31} = 0$, and $X_t = (X_{1\,t}, X_{2\,t}, X_{3\,t})'$.

By inverting the components 2 and 3, we obtain β under the required form.

Return

Bivariate VAR(1)

If

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix},$$

the non-causality from X to Y writes $a_{21} = 0$, that is

$$H_0 : R\text{vec}\underline{A} = 0, \quad R = (0,1,0,0), \quad \text{vec}\,\underline{A} = (a_{11}, a_{21}, a_{12}, a_{22})'.$$

Return

Proof of the VECM representation

Assume $\nu = 0$ for simplicity (otherwise replace ϵ_t by $\epsilon_t + \nu$).

$$X_t = A_1 X_{t-1} + \sum_{i=2}^{p} A_i X_{t-i} + \epsilon_t, \qquad \Pi = \sum_{i=1}^{p} A_i - I_d, \qquad \Pi_i = -\sum_{j=i+1}^{p} A_j$$

$$\begin{aligned}
\nabla X_t &= (A_1 - I_d) X_{t-1} + \sum_{i=2}^{p} A_i X_{t-i} + \epsilon_t \\
&= \Big(\sum_{i=1}^{p} A_i - I_d\Big) X_{t-1} + \sum_{i=2}^{p} A_i (X_{t-i} - X_{t-1}) + \epsilon_t \\
&= \Pi X_{t-1} + \sum_{i=2}^{p} A_i (\underbrace{X_{t-2} - X_{t-1}}_{-\nabla X_{t-1}}) + \sum_{i=3}^{p} A_i (X_{t-i} - X_{t-2}) + \epsilon_t \\
&= \Pi X_{t-1} + \Pi_1 \nabla X_{t-1} + \cdots + \Pi_{p-2} \nabla X_{t-p-2} + \sum_{i=p}^{p} A_i (X_{t-i} - X_{t-p+1}) + \epsilon_t
\end{aligned}$$

Return

Mean square error (MSE) of prediction

From the MA(∞) representation

$$X_t = \mu + \sum_{i=0}^{\infty} C_i \epsilon_{t-i}, \quad C_0 = I_d,$$

we get

$$\hat{X}_{t+h|t} = \mu + \sum_{i=h}^{\infty} C_i \epsilon_{t+h-i}$$

and the MSE of prediction is

$$\begin{aligned} \text{MSE}(h) &:= E\left(X_{t+h} - \hat{X}_{t+h|t}\right)\left(X_{t+h} - \hat{X}_{t+h|t}\right)' \\ &= \text{Var} \sum_{i=0}^{h-1} C_i \epsilon_{t+h-i} = \sum_{i=0}^{h-1} C_i \Sigma C_i'. \end{aligned}$$

Thus the recursion

$$\text{MSE}(h) = \text{MSE}(h-1) + C_{h-1} \Sigma C_{h-1}', \quad h > 0$$

with $\text{MSE}(0) = 0$.

Prediction interval

Assume in addition that

$$\epsilon_t \sim \mathcal{N}(0, \Sigma)$$

and denote by $X_{i,t}$ the i^{th} component of $X_t = (X_{1,t}, \ldots, X_{d,t})'$.

A prediction interval at level 95% and horizon h for $X_{i,t+h}$ is

$$\left[\hat{X}_{i,t+h|t} - 1.96\sqrt{\mathrm{MSE}_{ii}(h)}, \hat{X}_{i,t+h|t} + 1.96\sqrt{\mathrm{MSE}_{ii}(h)}\right]$$

where $\mathrm{MSE}_{ij}(h)$ is the entry ij of the matrix $\mathrm{MSE}(h)$.

Prediction region

Bivariate AR(1):

$$X_t = \begin{pmatrix} 1/2 & 2 \\ 0 & 1/2 \end{pmatrix} X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}\left\{0, \Sigma := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right\}.$$

We have

$$\hat{X}_{t+1|t} = \begin{pmatrix} 1/2X_{1t} + 2X_{2t} \\ 1/2X_{2t} \end{pmatrix}, \qquad X_{t+1} - \hat{X}_{t+1|t} = \epsilon_{t+1}.$$

The inverse of Σ has a square root $\Sigma^{-1/2}$. Thus, given $\{X_u, u \leq t\}$,

$$\Sigma^{-1/2}\epsilon_{t+1} \sim \mathcal{N}(0, I_2) \text{ and } \left(X_{t+1} - \hat{X}_{t+1|t}\right)' \Sigma^{-1} \left(X_{t+1} - \hat{X}_{t+1|t}\right) \sim \chi^2_2.$$

Allows to define a prediction region (interior of an ellipse) for the vector X_{t+1}.

Prediction ellipse at 95%

Return

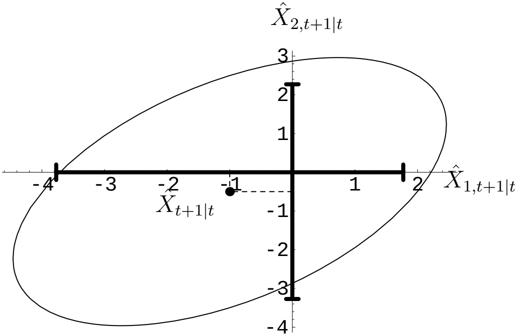


FIGURE 3.1 – Région de confiance pour la valeur X_{t+1} du processus AR(1) bivarié.