

2011
2012

LICENCE 3 – SCIENCES ECONOMIQUES

COURS DE M. MICHEL TERRAZA



Econométrie



Cours magistral d'économétrie

Écrit pour les étudiants de troisième année de licence en sciences économiques

Pour toutes incompréhensions, imperfections ou erreurs éventuelles,

Merci de les signaler sur le forum de la faculté de sciences économiques de l'UM1, à cette adresse :

<http://www.forum-sceco.fr> (Connexion à partir de <http://gide-eco.fr/forum>), à défaut de ne pouvoir me contacter directement...

Bibliographie :

Cours de Stat 2-3 (L2S3 & L2S4) + Intro à l'économétrie (Cours de Françoise Seyte)

Johnston & Dinardo, 1999, Méthodes économétrique, Economica (voir les annexes)

Bourbonnais, Econométrie, Dunod

Labrousse, 1978, Introduction à l'économétrie, Dunod

Dor, 2004, Econométrie, Pierson Education

Giraud & Shaiks, 1989, Econométrie, Puf

Raffin, 1993, Econométrie, Flash

Cohen & Pradel, 1993, Econométrie, Litec Economie

Dormont, 1998, Introduction à l'économétrie, Monchrétien

Densart, Rys, Vaneecloo, 2009, Econométrie, Presse universitaire de France (Livre de gestion sous SAS)

Greene, 2005, Econométrie, Pierson Education (5^{ème} édition)

Stock & Watson, 2007, Introduction to econometrics, Pierson Education

Ericour & Reynaud, 2007, TD d'Econométrie, Dunod

?, 2008, Exercice pédagogique d'économétrie, Economica

Mignon, 2008, Econométrie, Economica

Introduction

L'économétrie signifie la mesure des phénomènes économiques, c'est la science qui utilise la théorie économique, la mathématique, la statistique, l'information et, plus récemment, l'informatique pour analyser, vérifier et parfois prévoir les phénomènes économiques. L'économétrie a été créée dans les années 1930 aux Etats Unis par Frisch et la société d'économétrie avec Fischer et Ross, qui publieront *Econometrica*.

L'économétrie se divise ensuite plus tard en développement des méthodes et développement de la théorie économique. Le développement des méthodes passe par la méthode des moindres carrés, avec l'introduction des probabilités grâce à Haavelmo et les doubles moindres carrés. On appelle à l'heure actuelle l'économétrie, qui associe la théorie avec la méthodologie, pour vérifier les hypothèses économiques, on a appelé cette économétrie, Econométrie structurelle. Cette économétrie s'est beaucoup développé jusqu'en 1980 et ces aboutissements sont des modèles complexes qui ont été créés par différents organismes, comme l'INSEE, des universités. Elle a connu ensuite une concurrence du à une rivalité entre 2 grands organismes américains, la Cowles Commission (approche structurel) et le NBER. En effet, le NBER utilise une approche économétrique, plus empirique, qui ne fait pas particulièrement appel à la théorie économique, elle utilise la théorie des processus aléatoires, utilisé par Yule et aussi par Bochner, Box & Jenkins, ... Le NBER utilise l'analyse multidimensionnel des processus aléatoire, initié par Granger de l'université de Princeton et Sims, qui propose une nouvelle modélisation du modèle, la modélisation VAR (Vecteur Auto Régressif).

Ces deux approches sont devenues complémentaires à l'heure actuelle. Dans ce cours de licence, il est présenté l'économétrie structurelle, et l'économétrie non-structurelle est présentée en M1 puis M2.

Le modèle est l'instrument privilégié de l'économétrie structurelle, c'est une représentation simplifié mais la plus exhaustive possible d'une entité économique quelconque. Un phénomène économique peut concerner un modèle, un secteur d'activité ou géographique. Le modèle est construit à partir de présupposés théoriques qui vont en indiquer le contenu. Le modèle se présente sous la forme de système d'équations, la plupart du temps linéaire, mais pas obligatoirement, reliant entre elles, 2 grands types de variables, des variables endogènes, à expliquer, et des variables exogènes explicatives. Formellement, si on appelle Y , le vecteur des variables endogènes, B la matrice des coefficients de ces variables, X le vecteur des variables exogènes, A la matrice des coefficients de ces exogènes, si on suppose que le modèle est parfaitement déterminé, c'est-à-dire, qu'il possède autant d'équations G que de variables endogènes, et si ε est un vecteur contenant des variables aléatoires, appelé vecteur de l'erreur, le modèle s'écrit comme ça :

$$\begin{matrix} A & X \\ (G, K) & (K, 1) \end{matrix} + \begin{matrix} B & Y \\ (K, G) & (G, 1) \end{matrix} = \begin{matrix} \varepsilon \\ (G, 1) \end{matrix}$$

Ecrit sous cette forme, le modèle porte le nom de forme structurelle, c'est cette forme qui découle de la théorie économique, ce qui explique l'appellation d'économie structurelle. Dans un tel modèle, les variables endogènes et exogènes jouent un rôle identique

On utilise parfois la forme réduite du modèle et un modèle où la variable endogène s'exprime en fonction des seuls exogènes :

$$B \text{ (inv.)} = \begin{bmatrix} Y = CX + \eta \\ C = -B^{-1}A \\ \eta = B^{-1}\varepsilon \end{bmatrix}$$

On constate que le passage de la forme structurelle à la forme réduite est pratiquement toujours possible à condition que la matrice B soit inversible. Dans certains cas, on n'a que la forme réduite du modèle. Dans certains cas, on ne dispose que de la forme réduite. Or ce modèle intéressant est celui de la forme structurelle qui possède la matrice des coefficients des endogènes qui va leur indiquer la manière dont ils peuvent orienter leurs politiques économiques. Le passage de la forme réduite à la forme structurelle porte le nom d'identification du modèle. On considère alors 3 cas, le modèle juste identifié à une forme réduite correspond une forme structurelle, le modèle sous-identifié, que l'on ne considère pas en économétrie, et le modèle sur-identifié, où à une forme réduite correspond plusieurs formes structurelles. En économie, la plupart des modèles sont sur-identifiés, ce sont des modèles qui ont comme base la théorie économique de Keynes. Il existe des théorèmes permettant de vérifier l'identification des théorèmes économiques.

→ Arrangements B

Une fois le modèle construit, il peut se présenter sous différentes formes. Ces formes sont obtenues après avoir analysé un arrangement des variables endogènes dans les équations de la matrice B. Lorsque l'on a affaire à la représentation Bloc Diagonale, cela signifie que le modèle de départ est en fait constitué de plusieurs sous modèles indépendants les uns et des autres, il est préférable de traiter chacun d'entre eux séparément. La deuxième représentation, le Bloc Triangulaire, est la construction de modèles récurrents, c'est-à-dire que certaines équations dépendent des résultats d'équations précédentes. Et la dernière matrice est un modèle à équations simultanées, ce sont des modèles à interactions générales entre les variables entre les diverses équations : une variable est endogène mais peut être exogène dans une autre équation. Les modèles économétriques sont généralement sur-identifiés et à équations simultanées.

Première méthode d'estimation : On applique la méthode des moindres carrés ordinaires et du maximum de vraisemblance ; c'est la plus mauvaise des méthodes car les variables exogènes des modèles sont des variables certaines, indépendantes de l'aléa, incompatible avec le modèle à équations simultanées.

Deuxième méthode d'estimation : Utilisé seulement en cas de modèle JUSTE identifié, c'est la méthode des moindres carrés indirects. On estime les paramètres de la forme réduite, équations par équations, par la méthode des MCO puis on calcule les paramètres du modèle structurelle.

Troisième méthode d'estimation : Méthode des doubles moindres carrés, construite par Haavelmo, est une méthode pour modèle à équations simultanées sur-identifiée. C'est une méthode en deux étapes. La première étape consiste à estimer les paramètres de la forme réduite, ce qui permet d'obtenir des variables endogènes calculées. Deuxième étape : on reporte dans la forme structurelle les valeurs des endogènes en position d'exogènes, par leur valeurs calculées dans l'étape précédente ; on estime ensuite les paramètres de la forme structurelle par les MCO, équations par équations.

Quatrième méthode d'estimation : Méthode des triples moindres carrés. Les deux premières étapes sont identiques à la méthode précédente, et la troisième étape est en fait d'estimer simultanément par un algorithme les paramètres de la forme structurelle, et dans cet algorithme, les résultats des DMC servent à démarrer cet algorithme.

Modèle exemple :

$$\begin{cases} Y = C + I + G \\ C = a(Y - T) + b \end{cases}$$

Y, C : Variables endogènes

Modèle parfaitement déterminé

I, G, T : Variables exogènes

$$\underbrace{\begin{cases} 1 * Y - 1 * C \\ -aY + 1 * C \end{cases}}_{\text{Endogènes}} \rightarrow \underbrace{\begin{cases} -1 * Y - 1 * G + 0 * T + 0 * K = 0 \\ 0 * I + 0 * G + aT - bK = 0 \end{cases}}_{\text{Exogènes}}$$

$$\begin{bmatrix} 1 & -1 \\ -a & 1 \end{bmatrix} \begin{bmatrix} Y \\ C \end{bmatrix} + \begin{bmatrix} -1 & -1 & 0 & 0 \\ 0 & 0 & a & -b \end{bmatrix} \begin{bmatrix} I \\ G \\ T \\ K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\text{Forme réduite: } \begin{bmatrix} Y \\ C \end{bmatrix} = - \begin{bmatrix} 1 & -1 \\ -a & 1 \end{bmatrix}^{-1} * \begin{bmatrix} -1 & -1 & 0 & 0 \\ 0 & 0 & a & -b \end{bmatrix} \begin{bmatrix} I \\ G \\ T \\ K \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 \\ a & 1 \end{bmatrix} = 1 - a \neq 0, \quad a \neq 1$$

$$\begin{cases} Y = \frac{1}{1-a} (I + S - aT + b) \\ C = \frac{1}{1-a} (aI + aG - aT + b) \end{cases}, \quad a = \text{propension marginale à consommer}$$

Dans le cadre de ce cours, on étudie essentiellement le modèle structurel le plus simple, c'est celui qui possède une équation et plusieurs variables explicatives, on l'appelle le modèle linéaire générale simple (MLGS). Le premier chapitre de ce cours est consacré à ce modèle et à ses extensions pratiques. Le second chapitre porte sur le modèle d'autres types mais où certaines hypothèses ne sont plus vérifiées. Le 3^{ème} chapitre concerne des modèles à une équation particulières que l'on appelle, modèle à retard échelonné, qui sont des représentations

de la théorie des anticipations économiques, enfin le dernier chapitre de ce cours est consacré aux modèles à plusieurs équations.

Partie I

Modèle linéaire générale simple

Section A. Rappels et compléments en algèbres matricielles

A. Définitions :

Propriétés des transposés :

- La matrice symétrique : $B' = B$
- Le produit matriciel : $(AB)' = B'A'$

Propriété de la matrice inverse, propriétés du rang

Propriétés des Matrices orthogonal

Diagonalisation et produits scalaires

Trace d'une matrice carrée : opérateur linéaire

Idempotence : Carré d'une matrice est égale à elle-même

Formes quadratiques et matrices définies positive

Dérivée matricielle, dérivé d'une forme quadratique

Section B. Estimations par les MCO et par le MV

I. Spécification du MLGS

Le MLGS examine le cas où une variable endogène Y est liée à $k - 1$ variables explicatives, x_2, x_3, \dots, x_k est à un aléa ε de façon linéaire. Les différentes variables d'un modèle peuvent être de nature différente. Il peut s'agir de variables observés à un moment donné pour des échantillons, c'est un modèle en coupe instantané, ou au cours du temps en série temporel, d'où le nom de modèle en série temporel, ou les variables peuvent être observé pour des échantillons au cours du temps, c'est un modèle de Panel.

$$G = 1$$

Coupe instantané : $Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i, \forall i \in \mathbb{R}$ (Econométrie urbaine)

Série temporelle : $Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t$

Panel: $Y_{ti} = \beta_1 + \beta_2 X_{2ti} + \dots + \beta_k X_{kti} + \varepsilon_{ti}$

Dans ce cours, on ne va considérer que la seconde écriture...

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_t \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} = 1 & \dots & X_{t1} & \dots & X_{n1} \\ \vdots & \ddots & \vdots & \dots & \vdots \\ X_{1t} = 1 & \dots & X_{tt} & \dots & X_{nt} \\ \vdots & \dots & \vdots & \ddots & \vdots \\ X_{1n} = 1 & \dots & X_{tn} & \dots & X_{nn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_t \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{matrix} Y \\ (n, 1) \end{matrix} = \begin{matrix} X & \beta \\ (n, k) & (k, 1) \end{matrix} + \begin{matrix} \varepsilon \\ (k, 1) \end{matrix}$$

II. Hypothèse :

1) X_j constantes, $\forall j \in \mathbb{R}$

$$E[Y/X] = E[Y]$$

X et ε indépendantes

2) ε , vecteur de variables aléatoires

$$Y = X\beta + \varepsilon \rightarrow Y \text{ vecteur de variables aléatoires}$$

3) $Y_i, X_{ji}, \sigma_{X_{ji}}, \sigma_{X_{ji}} = X_{j'i'}$ Valeur finies quand $n \rightarrow +\infty$

4) $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow \varepsilon_t, t = 1, \dots, n$: Processus aléatoire stationnaire de type Bruit Blanc

Formule de VA

→ Processus aléatoire stationnaire d'ordre 2 : variables caractérisées que par les 2 premiers moments (espérance et variance)

→ Bruit blanc : Covariance = 0

$\forall t, \varepsilon_t \sim c_t(0, \sigma_\varepsilon) \Rightarrow \varepsilon_t$ IID (Indépendamment et identiquement distribué)

BB : $E[\varepsilon_t] = 0, \forall t$

$$V[\varepsilon_t] = E[\varepsilon_t^2] = \sigma_\varepsilon^2, \forall t$$

$Cov[\varepsilon_t, \varepsilon_t'] = E[\varepsilon_t \varepsilon_t'] = 0$, homoscédactique $\forall t, \forall t' \rightarrow tXt'$

IID: $\varepsilon_t \sim \mathcal{N}(0, \sigma_t)$

$$\underbrace{E \begin{bmatrix} \varepsilon & \varepsilon' \\ (n, 1) & (1, n) \end{bmatrix}}_{(n, n)} = \Omega_t = E \left[\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} [\varepsilon_1 \dots \varepsilon_n] \right]$$

$$= \begin{bmatrix} \varepsilon_1^2 & \dots & \dots & \dots & \varepsilon_1 \varepsilon_n \\ \varepsilon_2 \varepsilon_1 & \ddots & \vdots & \text{Symétrie} & \vdots \\ \vdots & \dots & \ddots & \dots & \vdots \\ \vdots & \text{Symétrie} & \vdots & \ddots & \vdots \\ \varepsilon_n \varepsilon_1 & \dots & \dots & \dots & \varepsilon_n^2 \end{bmatrix}$$

$$\begin{bmatrix} V[\varepsilon_1] & \dots & \text{Cov}(\varepsilon_{11}, \varepsilon_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(\varepsilon_n, \varepsilon_{11}) & \dots & V[\varepsilon_n] \end{bmatrix} = \sigma_\varepsilon^2 I_n = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{bmatrix}$$

$\forall t: V[\varepsilon_t] = E[\varepsilon_t^2] = \sigma_\varepsilon^2$
 $\text{Cov}(\varepsilon, \varepsilon') = E[\varepsilon_t \varepsilon_t'] = 0$

Ce qu'il y a à retenir de l'hypothèse 4 :

$$E[\varepsilon \varepsilon'] = \sigma_\varepsilon^2 I_n$$

$$E[\varepsilon \varepsilon] = 0$$

$$\varepsilon \sim \mathcal{N}\left(0, \sqrt{\sigma_\varepsilon^2 I_n}\right)$$

$$E[\varepsilon] = 0$$

$$E[Y/X] = E[X\beta + \varepsilon]$$

$$= E[X\beta] + \underbrace{E[\varepsilon]}_{=0}$$

$$= E[X\beta]$$

$$= XE[\beta]$$

$$= X\beta$$

$$\Rightarrow E[Y_t/X_t] = \beta' X_t$$

5) Pour n grand, $n \gg k$

6) $\text{Rang}_{(n,k)}^X = k$

Absence de colinéarité ou de multi-colinéarité entre les variables explicatives

Application des MCO (Moindres carrés ordinaires) et du MV (Maximum de vraisemblance)

⇒ Même résultat

III. Estimation du modèle

a) Les estimateurs des MCO

$$\text{Soit } \begin{matrix} Y \\ (n, k) \end{matrix} = \begin{matrix} X \\ (n, k) \end{matrix} \begin{matrix} \beta \\ (k, 1) \end{matrix} + \begin{matrix} \varepsilon \\ (n, 1) \end{matrix}$$

$$\begin{matrix} e \\ (n, 1) \end{matrix} = \begin{matrix} Y \\ (n, 1) \end{matrix} - \begin{matrix} \hat{Y} \\ (n, 1) \end{matrix}$$

e = Résidu, c'est un échantillon de ε donc il doit posséder les caractéristiques de l'aléa, c'est pourquoi $E(e) = 0$ & $V(e) = \text{constante}$ et les résidus sont non-corrélés

→ Permet de vérifier les propriétés de l'aléa

La méthode des MCO consiste à trouver $\min \sum_t e_t^2 = \min e'e$

$$e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y' - \hat{\beta}'X')(Y - X\hat{\beta})$$

$$= Y'Y - \underbrace{Y'X\hat{\beta}}_{\substack{\text{scalaire} \\ = \gamma \in \mathbb{R}}} - \underbrace{\hat{\beta}'X'Y}_{\substack{\text{scalaire} \\ = \tilde{\gamma} \in \mathbb{R}}} + \hat{\beta}'X'X\hat{\beta}$$

Nous voyons que $(\hat{\beta}'X'Y)' = Y'X\hat{\beta}$

$$\text{Soit } \gamma' = \tilde{\gamma}$$

Soit $\gamma = \tilde{\gamma}$ (la transposé d'un scalaire est lui-même)

$$\text{Donc } e'e = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$\text{Cherchons } \frac{\partial e'e}{\partial \hat{\beta}} = 0$$

$$\frac{\partial e'e}{\partial \hat{\beta}} = 0 - \frac{2\partial(\hat{\beta}'X'Y)}{\partial \hat{\beta}} + \frac{\partial(\hat{\beta}'(X'X)\hat{\beta})}{\partial \hat{\beta}}$$

$$\rightarrow \hat{\beta}'X'Y = \gamma = \alpha_1\hat{\beta}_1 + \alpha_2\hat{\beta}_2 + \dots + \alpha_k\hat{\beta}_k$$

$$\text{donc } \frac{\partial(\hat{\beta}'X'Y)}{\partial \hat{\beta}} = [\alpha_1; \alpha_2; \alpha_3; \dots; \alpha_k] = X'Y$$

Le produit $\hat{\beta}'(X'X)\hat{\beta}$ est une formule quadratique en $\hat{\beta}$ avec $X'X$ une matrice de dimension $k \times k$ qui est symétrique. C'est une matrice définie positive puisque le rang de $\tilde{\beta}$ est égal, au minimum, $(k, 1)$ c'est-à-dire 1.

$$\hat{\beta}'X'X\hat{\beta} \text{ est un scalaire et } \frac{\partial(\hat{\beta}'(X'X)\hat{\beta})}{\partial \tilde{\beta}} = 2(X'X)\hat{\beta}$$

$$\text{Donc } \frac{\partial e'e}{\partial \hat{\beta}} = -2X'Y + 2(X'X)\hat{\beta} = 0$$

$$(X'X)\hat{\beta} = X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Cela est possible si le déterminant de $X'X \neq 0$

Or par hypothèse, $\text{rg}(X) = k$ donc $\text{rg}(X'X) = k$ et donc l'inverse de $X'X$ existe.

On constate l'importance de l'hypothèse de $\text{rg}(X)$

b) Les caractéristiques des estimateurs

✚ $\hat{\beta}$ est un estimateur linéaire de β

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= \underbrace{(X'X)^{-1}X'X}_{=I} \beta + (X'X)^{-1}X' \varepsilon\end{aligned}$$

$$\rightarrow \hat{\beta} = \beta + (X'X)^{-1}X' \varepsilon$$

Donc $\hat{\beta}$ dépend de ε . La variable ε est l'aléa, donc $\hat{\beta}$ est un vecteur aléa (sans avoir à savoir la loi de probabilité)

✚ L'espérance mathématique de $\hat{\beta}$

$$E[\hat{\beta}] = \beta + E[(X'X)^{-1}X' \varepsilon]$$

Sachant que X est une valeur certaine \rightarrow

$$E[\hat{\beta}] = \beta + (X'X)^{-1}X' \underbrace{E[\varepsilon]}_{=0}$$

Donc $E[X] = X$ et $E[X'] = X'$

$E[\tilde{\beta}] = \beta \rightarrow$ Donc $\tilde{\beta}$ est un estimateur sans biais de β

✚ Construction de la matrice des variances-covariances de $\tilde{\beta}$

$$\Omega_{\tilde{\beta}} = \begin{bmatrix} V[\hat{\beta}_1] & \cdots & \text{Cov}(\hat{\beta}_i; \hat{\beta}_j) \\ \vdots & \ddots & \vdots \\ \text{Cov}(\hat{\beta}_j; \hat{\beta}_i) & \cdots & V[\hat{\beta}_k] \end{bmatrix}$$

$$\text{Cov}(\hat{\beta}_i; \hat{\beta}_j) = E[(\hat{\beta}_i - E[\hat{\beta}_i])(\hat{\beta}_j - E[\hat{\beta}_j])]$$

$$= E[(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)]$$

$$\text{si } i = j \rightarrow V[\hat{\beta}]$$

$$\Omega_{\tilde{\beta}} = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

$$= E[(\beta + (X'X)^{-1}X' \varepsilon - \beta)(\beta + (X'X)^{-1}X' \varepsilon - \beta)']$$

$$= E[(X'X)^{-1}X'\varepsilon](\varepsilon'X(X'X)^{-1})$$

Le transposé de $(X'X)^{-1}$ est égal à : $(X'X)^{-1}$ car $X'X$ est une matrice symétrique

$$= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}]$$

$$= (X'X)^{-1}X' \underbrace{E[\varepsilon\varepsilon']}_{\sigma_\varepsilon^2 I_n} X(X'X)^{-1}$$

$$= \sigma_\varepsilon^2 (X'X)^{-1}X'X(X'X)^{-1}$$

$$\Omega_{\hat{\beta}} = \sigma_\varepsilon^2 (X'X)^{-1}$$

→ Les variances des $\hat{\beta}$ qu'on utilise pour effectuer les tests de signification de paramètres se calculent en multipliant σ_ε^2 par les éléments de la diagonale de $(X'X)^{-1}$

✚ $\hat{\beta}$ est de variance minimale

Pour démontrer cette propriété ; considérons un autre estimateur $\hat{\hat{\beta}}$ et qu'on suppose linéaire

$$\hat{\hat{\beta}} = M Y = M X \beta + M \varepsilon$$

M étant de $\dim(k, n)$

On constate que si $\hat{\hat{\beta}}$ est l'estimateur recherché, alors $M = (X'X)^{-1}X' + N$, en supposant $N = 0$

→ Démonstration : $N = 0$

Pour que $\hat{\hat{\beta}}$ soit bon, il doit être sans biais, c'est-à-dire $E[\hat{\hat{\beta}}] = \beta$

$$\Leftrightarrow E(MX\beta + M\varepsilon) = \beta$$

$$\Leftrightarrow MX\beta + ME[\varepsilon] = \beta$$

$$\Leftrightarrow MX\beta = \beta$$

Soit $MX = I$

$$\text{Soit } [(X'X)^{-1}X' + N]X = I$$

$$\Leftrightarrow (X'X)^{-1}X'X + NX = I$$

$$\Leftrightarrow I + \underbrace{NX}_{=0} = I$$

$$\Omega_{\hat{\hat{\beta}}} = E\left[\left(\hat{\hat{\beta}} - E[\hat{\hat{\beta}}]\right)\left(\hat{\hat{\beta}} - E[\hat{\hat{\beta}}]\right)'\right]$$

$$\hat{\beta} = \frac{MX}{I} \beta + M\varepsilon$$

$$\hat{\beta} - E[\hat{\beta}] \equiv \beta + M\varepsilon = E[\hat{\beta}] + M\varepsilon$$

$$\Omega_{\hat{\beta}} = E[M\varepsilon][M\varepsilon']$$

$$= E[M\varepsilon\varepsilon'M']$$

$$= ME[\varepsilon\varepsilon']M'$$

$$= M\sigma_{\varepsilon}^2 M'$$

$$\text{Avec } M = (X'X)^{-1}X' + N \rightarrow \Omega_{\hat{\beta}} = \sigma_{\varepsilon}^2 MM'$$

$$= \sigma_{\varepsilon}^2 [(X'X)^{-1}X' + N][(X'X)^{-1}X' + N']$$

$$= \sigma_{\varepsilon}^2 [(X'X)^{-1}X' + N][(X'X)^{-1}X' + N']$$

$$= \sigma_{\varepsilon}^2 [(X'X)^{-1}(X'X)(X'X)^{-1} + (X'X)^{-1}X'N' + NX(X'X)^{-1} + NN']$$

$$= \sigma_{\varepsilon}^2 \left[(X'X)^{-1} + (X'X)^{-1} \underbrace{(NX)'}_{=0} + \underbrace{NX}_{=0} (X'X)^{-1} + NN' \right]$$

$$= \sigma_{\varepsilon}^2 [(X'X)^{-1} + NN'] = \Omega_{\hat{\beta}} + \sigma_{\varepsilon}^2 NN'$$

Nous cherchons $\min \sum n_{ij}^2$ (Termes de la diagonale de NN')

Entraine : $N = 0$ car on doit alors avoir $n_{ij} = 0$ pour $\forall i, j$

Donc $\hat{\beta} = \beta$ et $\hat{\beta}$ est bien estimateur MCO de variance minimale

Conclusion :

L'estimateur $\hat{\beta}$ est linéaire, sans biais et de variance minimale (convergent). On dit que c'est un estimateur BLUE (Best Linear Unbiased Estimator), c'est à dire qu'il s'agit du meilleur estimateur linéaire, celui qui a la plus petite variance parmi tous les estimateurs possibles de β .

Enfin $\hat{\beta}$ est convergent, il faut pour cela que $(X'X) \rightarrow 0$ lorsque $n \rightarrow \infty$

On considère que dans la pratique cette hypothèse est vérifiée.

c) Estimateur de la variance des erreurs : σ_ε^2

La seule info dont nous disposons sur l'aléa est le résidu e de la régression.

$$\begin{aligned} e &= Y - X\hat{\beta} \\ &= X\beta + \varepsilon - X[(X'X)^{-1}X'Y] \\ &= X\beta + \varepsilon - X(X'X)^{-1}X'(X\beta + \varepsilon) \\ &= X\beta + \varepsilon - X \underbrace{(X'X)^{-1}X'X}_{=I} \beta - X(X'X)^{-1}X'\varepsilon \\ &= X\beta + \varepsilon - X\beta - X(X'X)^{-1}X'\varepsilon \end{aligned}$$

$$= \underbrace{[I_n - X(X'X)^{-1}X']}_M \varepsilon$$

Par soucis de simplification
nous appellerons cet
encadrement :
 M
Toutefois, il est
différent du M
de tout à l'heure

$$e = M\varepsilon$$

$$\text{Donc } \sum_t e_t^2 = e'e = [M\varepsilon]'[M\varepsilon] = \varepsilon'M'M\varepsilon, \quad \text{or } M = I_n - X(X'X)^{-1}X'$$

$$M' = I_n - (X(X'X)^{-1}X')'$$

$$= I_n - X(X'X)^{-1}X' = M'$$

$$\text{Donc } M'M = M^2$$

→ M est une matrice idempotente

$$\text{Et donc } e'e = \varepsilon'M\varepsilon$$

Calculons $E[e'e]$

$$\varepsilon'M\varepsilon = \begin{matrix} [\varepsilon_1; \dots; \varepsilon_n] \\ (1, n) \end{matrix} \begin{matrix} \begin{bmatrix} n_{11} & n_{1j} \\ n_{ji} & n_{nn} \end{bmatrix} \\ (n, n) \end{matrix} \begin{matrix} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\ (n, 1) \end{matrix}$$

$$= \sum_i^n n_{ji} \varepsilon_i^2 + 2 \sum_i^n \sum_j^n n_{ij} \varepsilon_i \varepsilon_j$$

$$\text{Donc } E[\varepsilon'M\varepsilon] = \sum_i^n n_{ij} \underbrace{E[\varepsilon_i^2]}_{=\sigma_\varepsilon^2} + 2 \sum_i^n \sum_j^n n_{ij} \underbrace{E[\varepsilon_i \varepsilon_j]}_{=0}$$

$$E[e'e] = \sum n_{ij} \sigma_\varepsilon^2$$

$$= \text{trace}(M) \sigma_\varepsilon^2$$

$$= \sigma_\varepsilon^2 - \text{trace} \left[I_n - \frac{X(X'X)^{-1}X'}{\text{trace} [(X'X)^{-1}(X'X)]} \right]$$

$$E[e'e] = \sigma_\varepsilon^2 * (n - k)$$

$$E \left[\frac{e'e}{n - k} \right] = \sigma_\varepsilon^2 = \frac{e'e}{n - k}$$

On va donc choisir :

$$\sigma_\varepsilon^2 = \frac{\sum e_t^2}{n - k}$$

Or nous savons : $\Omega_{\hat{\beta}} = \sigma_\varepsilon^2 (X'X)^{-1}$

Donc les variances sont égales à $\sigma_\varepsilon^2 X$ (Les éléments de la diagonale de $(X'X)^{-1}$)

Cette diagonale devient un vecteur appelé $\text{diag}(X'X)^{-1}$

$$\text{Soit } V[\hat{\beta}] = \hat{\sigma}_\varepsilon^2 * \text{diag}(X'X)^{-1}$$

IV. Les Test des estimateurs

A. Les estimateurs et la maximum de vraisemblance

L'autre méthode d'estimation des paramètres est le maximum de vraisemblance. Il faut rajouter une hypothèse concernant l'aléa pour l'utiliser chacune des variables aléatoires obéit à une loi normale centrée 0 et d'écart type ε .

→ MV

$$\varepsilon_t \equiv \mathcal{N}(0, \sigma_\varepsilon)$$

$$\varepsilon \equiv \mathcal{N} \left(0, \sqrt{\sigma_\varepsilon^2 I_n} \right)$$

La fonction de MV sous forme matricielle s'écrit :

$$\mathcal{L} = \frac{1}{\sigma_\varepsilon^n} * \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} (Y - X\beta)'(Y - X\beta)\right)$$

$$\text{Ou } \frac{\partial \log \mathcal{L}}{\partial \beta} = 0$$

$$\hat{\beta} \rightarrow \frac{\partial \left[(Y - X\hat{\beta})' (Y - X\hat{\beta}) \right]}{\partial \hat{\beta}} = 0 \text{ (Méthode des MCO)}$$

$$\Rightarrow \underbrace{\hat{\beta}}_{\text{MV}} = \underbrace{\hat{\beta}}_{\text{MCO}} = (X'X)^{-1}X'Y$$

L'estimateur des MCO, c'est aussi l'estimateur du maximum de vraisemblance, l'important dans cette méthode, c'est l'hypothèse de normalité des aléas qui va permettre d'introduire les lois des paramètres inconnus de définir des intervalles de confiance et d'effectuer des tests de significations de ces paramètres.

B. Le test du coefficient de régression

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$

Epsilon obéit à une loi normale à n dimensions, d'après cette relation $\hat{\beta}$ est un estimateur linéaire de ε et obéit à une loi normale multi dimensionnelle mais la dimension de cette loi est cette fois k .

En supposant $\hat{\beta}$ sans biais et que l'on connaît sa variance :

$$\hat{\beta} \equiv \mathcal{N} \left(\beta, \sqrt{\sigma_\varepsilon^2 \text{diag}(X'X)^{-1}} \right)$$

Dans cette expression, σ_ε^2 est inconnu mais on pourrait le remplacer par :

$$\hat{\sigma}_\varepsilon^2 = \frac{e'e}{n - k}$$

Mais cette méthode ne suffit pas pour trouver $\hat{\beta}$

$$e'e = \underbrace{\varepsilon' M \varepsilon}_{\substack{\text{Forme} \\ \text{Quadratique}}} = \varepsilon' [I_n - X(X'X)^{-1}X'] \varepsilon$$

Dans cette expression, M est une matrice symétrique (idempotente) de trace $(n - k)$ et donc son rang est aussi $(n - k)$. Les valeurs propres sont donc soit nul soit égale à 1. Il est donc possible de trouver une matrice orthogonale, noté P qui diagonalise M .

Par cet intermédiaire, il est alors possible de déterminer la loi de $e'e$ à partir de ε en faisant subir à $\varepsilon'M\varepsilon$ une transformation, on utilise pour cela la matrice P . La matrice orthogonale P transforme le vecteur ε en un vecteur v de telle sorte que ε soit égale à Pv

$\varepsilon = Pv$, où ε : aléa et v : vecteurs propres

$$e'e = \varepsilon'M\varepsilon = v'P'MPv$$

Dans cette expression P est la matrice qui diagonalise M et c' est une matrice orthogonal, alors dans ce cas, $P'MP$ est une matrice diagonale qui est formé de $n - k$ chiffres 1 sur la diagonale des premières colonnes et de 0 partout ailleurs.

De ce fait :

$$e'e = \sum_{i=1}^{n-k} v_i^2$$

Avec v_i = composantes non-nulles précédentes, celles qui sont égales à 1 du produit $P'MP$

Comme ε est une variable normale à n dimensions, v qui est obtenu de ε par une transformation linéaire est aussi une variable aléatoire normale. Par ailleurs, les v_i sont indépendants puisque les vecteurs propres, associés aux valeurs propres d'une matrice symétrique forment un système indépendant (groupe orthogonal).

$$\Rightarrow \frac{e'e}{\sigma_\varepsilon^2} \equiv \chi^2(n - k)$$

Il est alors possible d'effectuer un test pour σ_ε^2 . Par ailleurs, on peut vérifier quelle est la relation entre le résidu e et le vecteur $\hat{\beta}$.

$$\Rightarrow \text{IC et test de } \sigma_\varepsilon^2$$

$$\rightarrow \text{Cov}[e, \hat{\beta}] = E \left[[e - E[e]] [\hat{\beta} - \beta]' \right]$$

$$E(e) = 0$$

$$\hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$$

$$\rightarrow \text{Cov}[e, \hat{\beta}] = E[(I_n - X(X'X)^{-1}X')\varepsilon]$$

$$= E[M\varepsilon] = ME[\varepsilon] = 0$$

Dans ces conditions, il va être possible de considérer un test concernant $\hat{\beta}$

Conséquence :

$$\frac{(\hat{\beta} - \beta)\sqrt{n-k}}{\sqrt{\sum e_t^2} \sqrt{\text{diag}(X'X)^{-1}}} \sim T(n-k)$$

Si $a_{jj} \rightarrow \text{Matrices } (X'X)^{-1}$

$$a_{jj} \rightarrow \text{diag}(X'X)^{-1}$$

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{a_{jj} \sum e_t^2}} \sqrt{n-k} \sim T(n-k)$$

Intervalle d'acceptation :

$$\text{Prob} \left[\beta_0 - \frac{t_{\alpha/2} \sqrt{a_{jj} \sum e_t^2}}{\sqrt{n-k}} < \hat{\beta}_j < \beta_0 + \frac{t_{\alpha/2} \sqrt{a_{jj} \sum e_t^2}}{\sqrt{n-k}} \right] = 1 - p$$

Si $\hat{\beta}_j \in$ à l'intervalle, H_0 accepté ($\beta_0 = \beta_j$)

Soit $H_0: \beta_j = 0$

$$t_c = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}}$$

$$s_{\hat{\beta}_j} = \hat{\sigma}_\varepsilon \sqrt{a_{jj}}$$

$|t_c| < t_{lu}, H_0$ accepté

$|t_c| > t_{lu}, H_0$ rejeté

C. Principe du test d'un ensemble de coefficient de régression

La signification d'un seul coefficient de régression peut être testée en utilisant la loi de Student

Lorsque l'on veut tester tous les coefficients de régression, il faut pouvoir déterminer la contribution des variables explicatives correspondantes dans l'explication de la variance totale de Y . Il faut donc décomposer la variance de Y , c'est-à-dire réaliser une analyse de la variance. Pour cela on recourt au modèle écrit avec variables centrées.

1. Ecriture du MLGS avec variables centrées

Soit x, y les variables centrées.

$$x_{jt} = X_{jt} - \bar{X}_j$$

$$y_t = Y_t - \bar{Y}$$

MLGS à l'instant t

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t \quad (1)$$

$$\left(\frac{\sum_t Y_t}{n} = \bar{Y} \right)$$

$$\frac{Y}{n} = \beta_1 + \beta_2 \bar{X}_2 + \dots + \beta_k \bar{X}_k + \bar{\varepsilon} \quad (2)$$

$$(1) - (2) = Y_t - \bar{Y} = y_t = \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \varepsilon_t - \bar{\varepsilon}$$

avec $t = 1, \dots, n$

$$\rightarrow y = x\beta + \varepsilon - \bar{\varepsilon}$$

$$y = x\beta + \eta \quad \eta = \varepsilon - \bar{\varepsilon}$$

Dans cette écriture, il est évident que β ne comporte plus β_1 et x et n'a plus de 1^{ère} colonne ne contenant que des 1. Sur ce nouveau modèle, on peut appliquer la méthode des MCO :

$$\text{Même résultat MLGS (Y)} \begin{cases} \hat{\beta} = (x'x)^{-1}x'y \\ E[\hat{\beta}] = \beta, \dots \\ V[\hat{\beta}] = \sigma_\varepsilon^2 \text{diag}(x'x)^{-1} \end{cases}$$

2. Décomposition de la variance totale de la variable endogène

$$y = x\beta + \eta$$

$$\hat{y} = x\hat{\beta}, \quad e = y - \hat{y}$$

$$\Rightarrow y = \hat{y} + e$$

$$= x\hat{\beta} + e$$

$$\Rightarrow e'e = (y - x\hat{\beta})'(y - x\hat{\beta})$$

$$\text{MCO} : \frac{\partial e'e}{\partial \hat{\beta}} = 0$$

$$\frac{\partial e'e}{\partial \hat{\beta}} = 0 - 2x'y + 2x'x\hat{\beta} = 0$$

$$\Rightarrow x'y = x'x\hat{\beta}$$

$$\Rightarrow e'e = y'y - 2\hat{\beta}'x'y + \hat{\beta}'x'y$$

$$= y'y - \hat{\beta}'x'y$$

$$y'y = \hat{\beta}'x'y + e'e$$

$$y'y = \sum y_t^2 = \sum (Y_t - \bar{Y})^2$$

$$e'e = \sum e_t^2$$

$$\hat{\beta}'x'y = y'y - e'e$$

Variance totale = Variance expliquée + Variance résiduelle

3. Coefficients de régression sur variables orthogonales

Pour pouvoir tester un groupe de coefficients de régression, on peut utiliser comme précédemment, une transformation qui exprime y la variable endogène en fonction de variables orthogonales. On va appeler z_2, \dots, z_k ($k - 1$) nouvelles variables que l'on va définir de la façon suivante :

$$\begin{cases} z_{2t} = w_{22}x_{2t} \\ z_{3t} = x_{32}x_{2t} + w_{33}x_{3t} \\ \vdots \\ z_{kt} = w_{k2}x_{2t} + \dots + w_{kk}x_{kt} \end{cases} \rightarrow \sum_t z_{jt}^2 = 1, \quad \forall j = 2, \dots, k$$

$$\sum_t z_{jt}z_{lt} = 0, \quad j \neq l, \quad l = 2, \dots, k$$

$$\rightarrow z'z = I_{k-1}$$

$$z = xw$$

$$x = zw^{-1}$$

$$y = x\beta + \varepsilon - \bar{\varepsilon} = x\hat{\beta} + e$$

$$\begin{cases} y = zw^{-1}\beta + \varepsilon - \bar{\varepsilon} \\ y = zw^{-1}\hat{\beta} + e \end{cases}$$

$$= \hat{\beta}^* = \frac{\beta}{w} \text{ MCO de } \beta^*$$

$$\hat{\beta}^* = (z'z)^{-1}z'y$$

$$= \text{Var}[\hat{\beta}^*] = \sigma_\varepsilon^2 I_{k-1}$$

$$E[\hat{\beta}^*] = \beta^*$$

4. Test de signification ou de l'ensemble ou d'un groupe de coefficients de régression

$\hat{\beta}_j^*$ est une combi linéaire de VA normale y_j , ils obéissent donc par construction à la loi normale :

$$\hat{\beta}_j^* \equiv \mathcal{N}(\beta_j^*, \sigma_\varepsilon)$$

$$\Leftrightarrow \frac{\hat{\beta}_j^* - \beta_j^*}{\sigma_\varepsilon} \equiv \mathcal{N}(0,1)$$

$$\begin{cases} \frac{\sum_{j=1}^k (\hat{\beta}_j^* - \beta_j^*)^2}{\sigma_\varepsilon^2} \equiv \chi^2(k-1) \\ \frac{\sum_t e_t^2}{\sigma_\varepsilon^2} \equiv \chi^2(n-k) \end{cases}$$

$$\Rightarrow F_c = \frac{\sum_j (\hat{\beta}_j^* - \beta_j^*)^2 / (k-1)}{\sum_t e_t^2 / (n-k)} \equiv \mathcal{F}(k-1, n-k)$$

$$H_0^* : \beta_2^* = \dots = \beta_k^* = 0$$

$$\beta = w\beta^*$$

$$\Leftrightarrow H_0 : \beta_2 = \dots = \beta_k = 0$$

$$\text{Si } H_0 \text{ vraie : } F_c = \frac{\sum(\hat{\beta}_j^* - \beta_j^*)^2 / (k - 1)}{\sum e_t^2 / (n - k)}$$

→ Forme matricielle :

$$F_c = \frac{\hat{\beta}' x' y / (k - 1)}{e' e / (n - k)} = \mathcal{F}(k - 1, n - k)$$

$\frac{\hat{\beta}' x' y}{k-1}$: Variance de Y expliquée par X_2, \dots, X_k

$\frac{e' e}{n-k}$: Variance résiduelle

Si $F_c < \mathcal{F}_{\alpha, k-1, n-k}$, H_0 acceptée

→ $H_0^* : \beta_k^* = 0 \leftrightarrow H_0 : \beta_k = 0$

$$\Rightarrow F_c = \frac{\hat{\beta}_k^{*2}}{\sum e_t^2 / (n - k)} \equiv \mathcal{F}(1, n - k)$$

$$t_c = \frac{\hat{\beta}_k}{\sqrt{w \sum e_t^2 / \sqrt{n - k}}} \equiv t(n - k)$$

V. Le coefficient de détermination

A. ...

Considérons l'équation de la variance précédente :

$$y'y = e'e + \hat{\beta}' x' y$$

Par définition, le coefficient de détermination, c'est la part de la variance expliquée dans la variance totale de Y , il s'écrit donc :

$$R^2 = \frac{\hat{\beta}' x' y}{y'y} = 1 - \frac{e'e}{y'y}$$

Par construction, ce coefficient de détermination n'a pas d'unité, compris entre 0 et 1.

$$e'e = \sum_t e_t^2$$

$$y'y = \sum_t y_t^2 = \sum_t (Y_t - \bar{Y})^2 = Y'Y - n\bar{Y}^2$$

$$(k-1, k-1) = \begin{matrix} R \\ x_2 \\ \vdots \\ x_k \end{matrix} \begin{bmatrix} x_2 & \cdots & x_k \\ & & \end{bmatrix} = \text{Matrices des coefficients des variables exogènes}$$

B. Test du coefficient de détermination

On a vu que $\frac{\hat{\beta}'x'y/(k-1)}{e'e/(n-k)} \sim \mathcal{F}(k-1, n-k)$

$$R^2 = \frac{\hat{\beta}'x'y}{y'y} \Rightarrow \hat{\beta}'x'y = R^2 y'y$$

$$\text{et } R^2 = 1 - \frac{e'e}{y'y}; \text{ donc}$$

$$\frac{R^2}{1-R^2} \frac{n-k}{k-1} \sim \mathcal{F}(k-1; n-k)$$

Test :

$$H_0 : \rho^2 = 0 ; H_1 : \rho^2 \neq 0$$

Si $F_c < F_{lu}$, alors H_0 acceptée

VI. La prévision (prédiction)

N'a de sens que si le modèle de régression précédent est correct. On réalise deux types de prédictions :

a, prédicteur de la valeur moyenne de la variable expliquée :

On cherche un prédicteur de la valeur moyenne de la variable endogène pour des variables explicatives données.

$$Y_0 / X_{20}, \dots, X_{k0}$$

o : Un individu quelconque

$$Y = X\beta + \varepsilon$$

$$Y_0 = \beta_1 + \beta_2 X_{20} + \cdots + \beta_k X_{k0} + \varepsilon_0$$

$$c = \begin{bmatrix} 1 \\ X_{20} \\ \vdots \\ X_{k0} \end{bmatrix}$$

$$\hat{Y}_0 = c' \hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_k X_{k0}$$

$$\begin{cases} E[c' \hat{\beta}] = c' \beta \\ V[c' \hat{\beta}] = \sigma_\varepsilon^2 c' (X'X)^{-1} c \end{cases}$$

$$\begin{cases} c' \hat{\beta} \sim \mathcal{N} \left(c' \beta, \sqrt{\sigma_\varepsilon^2 c' (X'X)^{-1} c} \right) \\ \frac{\sum e_t^2}{\sigma_\varepsilon^2} \sim \chi^2(n-k) \end{cases}$$

$$\frac{(c' \hat{\beta} - c' \beta)(n-k)}{\sqrt{\sum e_t^2 c' (X'X)^{-1} c}}$$

Test:

$$a = 1 - p = P \left[\hat{Y}_0 \pm \frac{t_{p/2} \sqrt{\sum e_t^2} \sqrt{c' (X'X)^{-1} c}}{\sqrt{n-k}} \geq c' \beta \right]$$

✚ Comparaison d'une prévision ponctuelle à l'équation des moindres carrés

$(Y_0, X_{20}, \dots, X_{k0})$

La prévision est-elle compatible avec l'estimation par les moindres carrés ?

$$Y_0 = c' \beta + \varepsilon_0$$

$$\begin{cases} MCO \\ MV \end{cases} \hat{Y}_0 = c' \hat{\beta}$$

$$\alpha = \hat{Y}_0 - \underbrace{Y_0}_{\substack{\text{Valeur} \\ \text{fixée} \\ \text{recherchée}}} = c' \hat{\beta} - c' \beta - \varepsilon_0$$

$$= c' (\hat{\beta} - \beta) - \varepsilon_0$$

$$E[\alpha] = E[c' \hat{\beta}] - E[c' \beta] - E[\varepsilon_0]$$

$$= c' \hat{\beta} - c' \beta - 0 = 0$$

$$\begin{aligned}
 V[\alpha] &= V[c'\hat{\beta}] + V[c'\beta] - 2Cov[c'\hat{\beta}, c'\beta] + V[\varepsilon_0] \\
 &= V[c'\hat{\beta}] + \sigma_\varepsilon^2 \\
 &= \sigma_\varepsilon^2 [1 + c'(X'X)^{-1}c]
 \end{aligned}$$

$$\begin{cases} \alpha \sim \mathcal{N}(0, \sqrt{\sigma_\varepsilon^2 [1 + c'(X'X)^{-1}c]}) \\ \frac{\sum e_t^2}{n-k} \sim \chi^2(n-k) \end{cases}$$

$$H_0 = \hat{Y}_0 \in I.C$$

$$Prob \left[\hat{Y}_0 \pm \frac{t_{p/2} \sqrt{\sum e_t^2} \sqrt{1 + c'(X'X)^{-1}c}}{\sqrt{n-k}} \geq Y_0 \right] = 1 - p$$

Si $\hat{Y}_0 \in I.C.$, H_0 acceptée
 Si $\hat{Y}_0 \notin I.C.$, H_0 rejetée

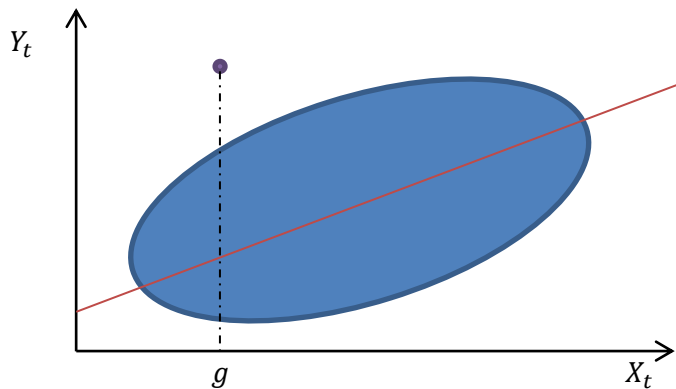
Le vecteur $(Y_0; X_{20} \dots X_{k0})$ appartient (H_0) ou non (H_1) au modèle de régression

Section C. Extension

I. Les variables indicatrices

A. Définitions et spécifications

On appelle variables indicatrices (ou variables tampons / auxiliaires / binaires / dichotomiques / « dummy variables »), traduisant la présence ou l'absence d'un phénomène explicative. On leur attribue la valeur 1 ou 0, selon que l'évènement est présent ou absent dans l'observation considérée. La spécification du modèle économétrique ne pose, en général, aucun problème lorsque l'on prend en compte ce type de variable. Pour illustrer cette spécification, expliquons un modèle simple à 3 variables dont une variable exogène est indicatrice mais uniquement pour une observation de 1, ..., n. C'est, par exemple, le cas de la régression linéaire simple où pour laquelle la variable explicative, présente pour une raison quelconque, un point, une valeur aberrante (exceptionnel).



$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t, \quad \forall t = 1, \dots, n$$

On va supposer qu'à un moment donnée t , une variable exceptionnel modifie la régression de tel sorte que la valeur soit considérée comme aberrante dans la régression. On va considérer que cette perturbation ce produit au moment $t = g$

$$\begin{cases} Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + \varepsilon_t \\ \text{avec } \begin{cases} Z_t = 0, & \forall t, t \neq g \\ Z_t = 1, & t = g \end{cases} \end{cases}$$

$$\text{Si } \begin{cases} t \neq g, & Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \\ t = g, & Y_t = (\beta_1 + \beta_3) + \beta_2 X_t + \varepsilon_t \end{cases}$$

On constate que ces deux modèles ne se différencient que par leur ordonnée à l'origine qui est β_1 dans le premier et $\beta_1 + \beta_3$ dans le second. B3 est la modification de l'ordonnée à l'origine, dû à la perturbation. On constate que dans cette exemple, l'influence linéaire de X sur Y n'est pas modifier par cette perturbation, seul l'ordonnée à l'origine va changer. On peut aussi envisager un modèle où la perturbation va modifier la relation. C'est par exemple le modèle suivante :

$$\begin{cases} Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + \beta_4 Z_t X_t + \varepsilon_t \\ \text{avec } \begin{cases} Z_t = 0, & \forall t, t \neq g \\ Z_t = 1, & t = g \end{cases} \end{cases}$$

$$\text{Si } \begin{cases} t \neq g, & Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \\ t = g, & Y_t = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) X_t + \varepsilon_t \end{cases}$$

Ces deux modèles se généralisent au MLGS avec une variable dichotomique mais aussi au MLGS avec plusieurs variables dichotomiques. Le problème statistique ne présente aucune difficulté, il faut simplement vérifier que la matrice X du MLGS ne présente pas de colinéarité

B. Quelques applications courantes

Les variables indicatrices peuvent traduire plusieurs origines outre celles des valeurs aberrantes : effets temporels, spatiaux, dus aux variables qualitatives.

- L'effet temporel est l'un des plus courants en économétrie. Il concerne les modifications entre 2 périodes séparées par des événements exceptionnels (guerre, etc...)

L'exemple le plus cité dans les cours d'économétrie est celui emprunté à l'ouvrage de Johnston : c'est le cas de la fonction de consommation en état de guerre et en état de paix.

$$C = \alpha_1 + \beta Y + \varepsilon, \text{ en état de guerre}$$

$$C = \alpha_2 + \beta Y + \varepsilon, \text{ en état de paix}$$

$$\begin{cases} C = \alpha_1 X_1 + \alpha_2 X_2 + \beta Y + \varepsilon \\ X_1 = 1, & X_2 = 0, \text{ en état de guerre} \\ X_1 = 0, & X_2 = 1, \text{ en état de paix} \end{cases}$$

Dans ces équations de consommations ne se différencient que par l'ordonnée à l'origine. Si ce n'était pas le cas, on ne procéderait pas avec des variables dichotomiques, on ferait des ajustements séparés sur les deux périodes.

Les effets spatiaux concernent, par exemple, la modification des ordonnées à l'origine de relations construites à partir de régions ou de pays différents.

Les variables qualitatives jouent un rôle de plus en plus important en économie et peuvent être prise en compte par des variables de types dichotomiques.

Il convient de distinguer 2 cas dans les variables qualitatives :

- Celui où la variable qualitative est exogène :
On procède comme précédemment. Soit une variable exogène, le genre. 0 pour le genre féminin et 1 pour le genre masculin.
- Celui où la variable qualitative est endogène :
On fait alors appel à des modèles plus complexes qui portent des noms comme LOGYT ou PROBYT. Il s'agit d'une branche particulière de l'économétrie, l'économétrie des variables qualitatives. (Pas traité en L3)

C. Utilisation des variables dichotomiques

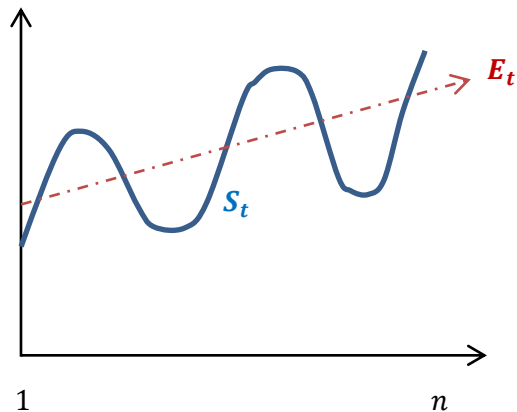
Rappel : Soit x_t une chronique, l'observation d'un phénomène au cours du temps

$t = 1, \dots, n$ avec un pas de temps constant $\Delta t = 1$

On considère, à un instant t de cette série chronologique, l'extra saisonnier E_t , le saisonnier S_t et le résiduel R_t

Ces 3 composantes sont orthogonales ou liées entre elles pour former la valeur x_t . Si elles sont orthogonales, le modèle est additif, si elles sont liées entre elles, le modèle est multiplicatif :

$$\begin{cases} x_t = E_t + S_t + R_t & \text{Additif} \\ x_t = E_t S_t R_t & \text{Multiplicatif} \end{cases}$$



Pour désaisonnaliser, on utilise deux méthodes de régression :

- Les fonctions trigonométriques
- Les variables dichotomiques

$$\begin{cases} x_t, \text{ annuelle} \\ E_t \text{ est linéaire} \\ \text{fonction trigonométrique, modèle additif} \end{cases}$$

$$x_t = E_t + S_t + R_t$$

$$= \alpha + \beta * t + \underbrace{\sum_{j=1} A_j \cos(\omega + \Phi_j)}_{+\sum_j c_j \cos \omega_{jt} + \sum_j D_j \sin \omega_{jt}} + \varepsilon_t$$

$$\omega = \frac{2\pi}{T} = 2\pi\lambda$$

La saisonnalité est modélisée par une somme de cosinusoides avec :

- A : l'amplitude
- ω = Radian (pulsation) = $\frac{2\pi}{T} = 2\pi\lambda$
 - o T = Unité de temps (période)
 - o λ = Hertz (fréquence)
- Φ = Déphasage

La cosinusoïde n'est pas une forme linéaire, il faut utiliser la formule $\cos(a + b) = \cos(a) \cos(b) - \sin(a) \sin(b)$, donc :

$$S_t = \sum_j A_j \cos \omega_{jT} \cos \Phi_j - \sum_j A_j \sin \omega_{jT} \sin \Phi_j$$

$$= \sum_{j=1}^{\frac{D}{2}-1} c_j \cos \omega_{jT} + \sum_{j=1}^{\frac{D}{2}-1} D_j \sin \omega_{jT}$$

avec $c_j = A_j \cos \Phi_j$ et $D_j = -A_j \sin \Phi_j$

Il y a 5 harmoniques car ensuite nous passons à $\frac{2\pi}{2}$ soit 2 points, ce qui est insuffisant pour représenter un cycle.

$$S_t = c_1 \cos \frac{2\pi}{12} T + c_2 \cos \frac{2\pi}{6} T + c_3 \cos \frac{2\pi}{4} T + c_4 \cos \frac{2\pi}{3} T + c_5 \cos \frac{2\pi}{2.4} T + D_1 \sin \frac{2\pi}{12} t + \dots$$

$$\frac{T}{2} - 1 = \frac{12}{2} - 1 = 5$$

10 + 2 = 12 coefficients (variables explicatives)

$$\frac{2\pi}{12} * j \left\{ \begin{array}{l} \rightarrow \frac{2\pi}{12} \\ \rightarrow \frac{2\pi}{6} \\ \rightarrow \frac{2\pi}{4} \\ \rightarrow \frac{2\pi}{3} \\ \rightarrow \frac{2\pi}{2.4} \end{array} \right. \Rightarrow j = 1, \dots, 5$$

Via les MCO : $\hat{x}_t = \hat{\alpha} + \hat{\beta} * t + \hat{c}_1 \cos$ etc ...

$$x_t^{CVS} = x_t - (\hat{S}_t)$$

$$= x_t - \left[\hat{c}_1 \cos \frac{2\pi}{12} T + \dots + \hat{D}_1 \sin \frac{2\pi}{12} T + \dots \right]$$

✚ Dessaisonnalisation par variables dichotomiques

x_t : Chronique Trim. modèle additif
 $E_t \sim$ modèle linéaire

$$x_t = E_t + S_t + R_t$$

$$= \alpha + \beta * t + \left[\sum_{i=1} \delta_i * IND_{ti} \right] + \varepsilon_t, \text{ avec } IND : \text{ variables dichotomiques}$$

$$\begin{array}{cccc}
 & IND_1 & IND_2 & IND_3 & IND_4 \\
 \begin{array}{c} 1 \\ 2 \\ 3 \\ S_t = 4 \\ 5 \\ \vdots \\ n \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ \vdots & & & \\ \text{etc} & \dots & & \end{bmatrix}
 \end{array}$$

$$\text{Donc } x_t = \alpha + \beta * t + \sum \delta_i IND_{ti} + \varepsilon_t$$

$$\left(\text{Rappel : } \alpha + \beta * t = E_t, \quad \sum \delta_i IND_{ti} = S_t, \quad \varepsilon_t = R_t \right)$$

$$\begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ 1 & 5 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & n & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

On ne peut pas utiliser directement la méthode des MCO parce que le rang de la matrice A n'est plus maximal. Il y a colinéarité entre les valeurs.

2 Solutions :

- On utilise les variables centrées
- On utilise le principe de conservation des aires (Somme des coefficients = 0)
donc $\delta_4 = -(\delta_1 + \delta_2 + \delta_3)$ (Par exemple)

Solution :

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 1 \\ 1 & 4 & -1 & -1 & -1 \\ 1 & 5 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & n & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$$

$$x_t^{CVS} = x_t - (\hat{S}_t)$$

II. La sélection de sous ensemble de variables

A. L'évaluation de la qualité d'une régression

L'évaluation d'une qualité d'une régression est entendue dans ce paragraphe comme étant le choix pertinent de la combinaison linéaire de variables explicatives en nombre le plus restreint possible qui va contribuer le mieux à expliquer la variable endogène. Considérons par exemple une régression effectuée à partir de k variables explicatives puis une autre régression, effectuée à partir d'un sous ensemble de cardinal M de ces k variables explicatives.

N'y a-t-il pas, parmi les k variables de départ, $k - M$ variables qui n'auraient qu'une explication négligeables pour Y ?

Pour répondre, on dispose, en dehors des tests du MLGS, de plusieurs critères d'évaluation de la qualité d'une régression.

- Le carré moyen résiduel

$$S_M^2 = \frac{SCE_M}{n - M}, \text{ SCE} = \text{Somme des carrés des écarts} \left(e'e \text{ ou } \sum e_t^2 \right)$$

- Coefficient de détermination

$$R_M^2 = \frac{ST - SCE_M}{ST} = 1 - \frac{SCE_M}{ST}$$

$$ST = y'y \text{ ou } \sum \hat{y}_t^2$$

$$R_M^2 = 1 - \frac{(n - M)S_M^2}{ST}$$

On peut, à partir du R^2 précédent, calculer un autre R^2 pour un gain de variance que l'on appelle coefficient de détermination ajusté qui a pour formule :

$$\bar{R}_M^2 = 1 - \frac{(n - 1)S_M^2}{ST}$$

- Statistique de Mallows

$$C_M : \text{Mesure du biais} = \frac{SCE_M}{S_k^2} + (2M - n)$$

$$E[C_M] = M \Leftrightarrow C_M - M$$

Quand on augmente le nombre de variables, la SCE n'augmente pas :

$$SCE_k \leq SCE_M$$

Par contre comme SCE_M est divisé par $n - M$ cette inégalité n'est pas vraie entre S_k^2 et S_M^2 , c'est la raison pour laquelle on préfère utiliser le second critère qui induit le résultat suivante :

$$R_k^2 \geq R_M^2$$

Dans un modèle de régression, le coefficient de détermination s'accroît avec le nombre de variables explicatives, c'est souvent pour cette raison que l'on construit des modèles non parcimonieux.

$$1 - \bar{R}_M^2 = \frac{n-1}{n-M} (1 - R_M^2)$$

On peut aussi calculer la statistique suivante F_M , qui s'écrit :

$$F_M = \frac{(SCE_M - SCE_k)/(k - M)}{SCE_k/(n - k)}$$

Cette statistique permet de sélectionner le sous ensemble de cardinal M en maximisant la valeur de F_M . On peut montrer :

$$M / \max F_M$$

$$C_M - M = (k - M)(F_M - 1)$$

De façon plus récente, les économètres utilisent des statistiques différentes pour évaluer la qualité de leur régression, ces statistiques sont issues de l'analyse des séries temporelles et seront revus en M1. Elles portent le nom de critères d'informations, et le premier à en avoir Akaike en 1970 et son critère prend le nom de Akaike Information Criterion, le meilleur modèle est celui qui minimise ce critère.

B. La méthode d'évaluation de toutes les régressions possibles

On considère l'ensemble de toutes les régressions possibles entre la variable endogène et les combinaisons des variables exogènes. La méthode consiste :

- A calculer les 2^{k-1} régressions possibles
- De répartir ces régressions en $k - 1$ groupes correspondant chacun à un nombre différent de variables explicatives, et pour chacun de ces groupes, de retenir comme critère, le meilleur R^2 , on obtient alors la répartition des R^2 pour chacun des groupes et la méthode consiste à retenir le R^2 le meilleur de tel sorte que les autres ne soit pas significativement différent de lui

Avec 10 variables explicatives, on obtient 1023 ($2^{k-1} - 1$) régressions

Les statisticiens préfèrent les méthodes « pas à pas »

 La régression pas à pas descendante

Cette procédure considère d'abord la régression complète puis élimine la variable qui correspond au t_c le plus petit ou la variable qui accroît le moins le carré des résidus. Cette méthode et les suivantes possèdent des critères d'arrêts, sinon, elles se poursuivent jusqu'à élimination ou récupération de toutes les variables. Cette méthode est une méthode en 4 étapes :

- Estimation de la régression complète
- Calcul des t_c , pour chaque variables explicatives
- On retient la valeur la plus petit du t_c , on la note t_M et on la compare à un t_c pour un seuil choisi. Si t_{\min} correspond au t_{lu} on conserve la variable et a fortiori toutes les autres. Si t_{\min} est inférieur au t_{lu} , alors on élimine la variable
- On recalcule la régression avec les $k - 2$ variables explicatives restantes, et on réitère l'algorithme à partir de la deuxième étape.

S'il n'y a pas de critère d'arrêt, la méthode peut éliminer toutes les variables.

Sélection ascendante des variables explicatives

Méthode symétrique à la précédente, elle consiste à introduire pas à pas les variables explicatives. Les étapes de calculs sont donc les suivantes.

- Calcul des coefficients de corrélation simple entre Y et toutes les variables exogènes. On retient l'exogène qui a le plus fort coefficient de corrélation
- On calcule le résidu de la régression précédente, puis les coefficients de corrélation avec les variables explicatives restantes. On sélectionne celle qui possède le plus fort coefficient de corrélation
- A chaque pas de l'algorithme, on peut calculer par exemple le coefficient de détermination du modèle correspondant. On peut aussi le t_c . Cela permet de construire un critère d'arrêt, par exemple, la nouvelle variable introduite à un t_c non significatif, ou encore, considérer un critère (par exemple : 10^{-6}) et si le R^2 du modèle n'est pas supérieur dans son accroissement à ce critère, l'algorithme s'arrête.

Cette procédure est souvent utilisée, car, avec un critère d'arrêt, elle peut être très rapide. Mais cette procédure n'indique pas l'effet que peut produire l'introduction d'une nouvelle variable sur les variables antérieurement introduites.

La sélection progressive avec possibilité d'élimination

Cette procédure fonctionne comme la précédente mais à chaque pas elle s'interroge sur la signification de la variable précédente introduite. C'est donc une version améliorée de la régression pas à pas, ascendante ou progressive. Les étapes de calcul sont donc entièrement identiques au précédente et possède aussi un critère d'arrêt. La différence vient du fait qu'à

partir du second pas (et au suivant ensuite) on teste la signification le coefficient de régression de la seconde et de la première variable.

Si le coefficient de la seconde variable n'est pas significatif, on s'arrête à la première. Si le coefficient de la première variable n'est pas significatif alors que la seconde l'est, on élimine la première variable et la procédure se poursuit avec une troisième variable.

Ces procédures sont commodes lors des études qui comportent beaucoup de variances explicatives mais elles peuvent être décevantes sur le plan du résultat. Cela tient de l'automatisme de la technique qui en effet sélectionne un ensemble de variable mais qui ne garantit pas qu'ils soient le meilleur, surtout sur un plan économique. Parmi toutes ces méthodes de sélection, y compris d'autres, rien ne permet d'affirmer que l'une d'entre elles est meilleure qu'une autre.

III. Les moindres carrés contraints

A. Le MLGS contraint

Dans le MLGS, le vecteur β est constitué d'élément constant inconnu. Il peut arriver des cas où l'on souhaite que ce vecteur appartienne à un sous espace du corps des réels. Le cas le plus simple traité ici étant celui d'une variété linéaire. Ce cas est celui de l'introduction au problème, plus difficile, de l'estimation de paramètres sous contrainte.

Supposons que les éléments de β sous soumis à p contraintes compatibles.

$$R\beta = r$$

Avec R , une matrice (p, k)

et r , un vecteur donnée à p éléments

Le modèle contraint s'écrit comme ceci

$$\begin{cases} Y = X\beta + \varepsilon \\ (n, 1) = (n, k)(k, 1) + (n, 1) \\ R\beta = r \\ (p, k)(k, 1) = (p, 1) \end{cases}, \text{ avec } \text{rg}(R) = p < k$$

Ecrire que le rang de R soit égale à p revient à supposer que les p contraintes sont linéairement indépendants. Si le rang de R était égale à k , on pourrait trouver β en résolvant le système d'équation linéaire, la méthode de Cramer : $\beta = R^{-1}r$

B. Les estimateurs des MCC

La condition sur le rang de R assure que le système des contraintes admet une infinité de solutions en β . Les MCC choisissent parmi l'ensemble des solutions, celle qui minimise la somme des carrés des résidus. Le problème mathématique s'écrit donc de la façon suivante :

$$\begin{cases} \min e'e / R\hat{\beta}_0 = r \\ \min(Y - X\hat{\beta}_0)'(Y - X\hat{\beta}_0) / R\hat{\beta}_0 = r \end{cases}$$

$$\phi = (Y - X\hat{\beta}_0)'(Y - X\hat{\beta}_0) - 2\lambda'(R\hat{\beta}_0 - r)$$

$$\begin{cases} \frac{\partial \phi}{\partial \hat{\beta}_0} = -2X'Y + 2X'X\hat{\beta}_0 - 2\lambda'R = 0 \\ \frac{\partial \phi}{\partial \lambda} = -2(R\hat{\beta}_0 - r) = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} X'X\hat{\beta}_0 - X'Y - R'\lambda = 0 \\ R\hat{\beta}_0 - r = 0 \end{cases}$$

$$\hat{\beta}_0 = (X'X)^{-1}X'Y - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R(X'X)^{-1}X'Y - r]$$

$$Y = X\beta_0 + \varepsilon$$

$$\hat{\beta}_0 = \beta_0 + \{(X'X)^{-1} - (X'X)^{-1}R'[\dots]^{-1}R(X'X)^{-1}\}X'\varepsilon$$

$$E[\hat{\beta}_0] = \beta_0$$

$$V[\hat{\beta}_0] = \sigma_\varepsilon^2 \{(X'X)^{-1} - (X'X)^{-1}R'[\dots]^{-1}R(X'X)^{-1}\}$$

On constate que $V[\hat{\beta}] - V[\hat{\beta}_0]$ est une matrice définie positive qui montre que l'introduction d'informations a priori sur β réduit nécessairement la variance des estimateurs qui prennent en compte cette info.

Exemple :

$$\begin{cases} Y = b_0 + b_1X_1 + b_2X_2 + \varepsilon \\ b_1 + b_2 = 1 \end{cases}$$

$$b_2 = 1 - b_1$$

$$Y = b_0 + b_1X_1 + X_2 - b_1X_2 + \varepsilon$$

$$Y - X_2 = b_0 + b_1(X_1 - X_2) + \varepsilon$$

$$\text{MCO} : ((Y - X_2); (X_1 - X_2))$$

$$\hat{b}_0, \hat{b}_1, \quad b_2 = 1 - b_1$$

$$\hat{b}_2 = 1 - \hat{b}_1$$

Section D. Robustesse du MLGS

Un modèle est robuste lorsqu'il est valide dans des circonstances différentes, par exemple, l'estimation de la fonction de consommation avant et après la seconde guerre mondiale, la relation prix/récolte de vin : cette relation s'est-elle modifiée après l'introduction en 1970 de ce marché dans la CEE, ... Dans ces cas, on les appelle des cas de « robustesse structurelle », l'étude porte sur des temps consécutifs, mais elle peut concerner des périodes de temps qui se chevauchent, par exemple, la courbe de Philips au cours de différentes phases de conjoncture. Elle peut être liée à des problèmes d'homogénéité spatiale, comme la loi rang-dimension, tel qu'annoncée par Zipf. Cette relation peut aussi concerner des problèmes sectoriels, comme la fonction d'investissement dans la sidérurgie. Enfin, l'étude de la robustesse peut concerner le sens de la causalité entre les variables, par exemple, les prix sont-ils toujours la cause des salaires, ou le contraire ?

On va considérer que l'ordre causal n'est pas remis en question.

Considérons le MLGS suivant :

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

On dira que ce modèle est robuste si quelque soit les sous-ensembles constitué à partir d'observations consécutives sur la période $1, n$, les estimations du même modèle sur chacun de ces sous-ensembles sont :

- Valides
La régression « passe » les tests
- Stables
Les paramètres estimés ne sont pas significativement différents des paramètres $\hat{\alpha}$ & $\hat{\beta}$ de la régression complète. Si on considère deux sous-ensembles disjoints, on dispose de 3 estimations :

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t, \quad t \in [1, \dots, n]$$

$$\hat{Y}_t = \hat{\alpha}^1 + \hat{\beta}^1 X_t, \quad t \in [1, \dots, n_1]$$

$$\hat{Y}_t = \hat{\alpha}^2 + \hat{\beta}^2 X_t, \quad t \in [n_1 + 1, \dots, n]$$

On dit que le modèle estimé sur la période $1, n$ si les 3 modèles précédents passent les tests et si $\hat{\alpha}^1$ et $\hat{\beta}^1$ ne sont pas significativement différents de $\hat{\alpha}^2$ & $\hat{\beta}^2$ et de $\hat{\alpha}$ & $\hat{\beta}$

Cette définition de la robustesse amène l'étude de la comparaison statistique de 2 coefficients de régression et par là à celle de la comparaison entre 2 coefficients de corrélation linéaire simple.

✚ Test de 2 coefficients de corrélation

Supposons que l'on dispose de 2 sous-ensembles indépendants de taille n_1 et n_2 . On va appeler r_1 et r_2 les coefficients de corrélations linéaires simple de chacun des deux sous-échantillons. On peut montrer pour de petits échantillons que le coefficient de corrélation n'obéit pas à une distribution d'expression simple autour de son espérance mathématique ρ . On constate que la distribution est fortement asymétrique pour les valeurs de ρ qui sont éloignées de 0. Pour cette raison, Fisher a proposé une transformation non-linéaire du coefficient de corrélation : cette transformation est la suivante.

$$z = \operatorname{arctan} \rho = \frac{1}{2} \log \frac{1 + \rho}{1 - \rho}$$

En utilisant cette transformation, on considère que la loi de probabilité obéit à une loi normale, même pour des valeurs éloignées de 0, les caractéristiques, les moments, de cette loi normale sont :

$$E[z] = \operatorname{arctan} \rho$$

$$s^2[z] = \frac{1}{n - 3}$$

$$n_1 \rightarrow \rho_1 \Rightarrow z_1 = \operatorname{arctan} \rho_1$$

$$n_2 \rightarrow \rho_2 \Rightarrow z_2 = \operatorname{arctan} \rho_2$$

$$H_0 : d = z_1 - z_2 = 0$$

$$H_1 : d \neq 0$$

$$E[d] = E[z_1 - z_2] = E[z_1] - E[z_2]$$

$$s^2[d] = \frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}$$

$$d_c = \frac{d}{s[d]} \sim \mathcal{N}(0,1)$$

RDD : $t_c < 1,96, H_0$ acceptée

✚ Test de comparaison de 2 coefficients de régression

On dispose des mêmes sous-ensembles.

$$H_0 : d = \beta^1 - \beta^2 = 0$$

$$H_1 : d \neq 0$$

$$E[\hat{\beta}^1] = \beta^1 \text{ \& } E[\hat{\beta}^2] = \beta^2$$

$$\Rightarrow E[\hat{\beta}^1] - E[\hat{\beta}^2] = E[\hat{\beta}^1 - \hat{\beta}^2] = 0$$

$$s^2[d] = s^2[\hat{\beta}^1] + s^2[\hat{\beta}^2]$$

$$\text{RDD} : d_c = \frac{\hat{\beta}^1 - \hat{\beta}^2}{s[d]} < t_{lu}(n_1 + n_2 - 4), H_0 \text{ acceptée}$$

✚ Tests généraux de stabilité du modèle de régression

Il existe de nombreux tests de stabilité des coefficients d'un modèle de régression. Certains de ces tests sont de conceptions simples ou intuitives, ce sont les régressions « roulantes » ou des régressions « CUSUM ». D'autres sont plus élaborés sur un plan purement stat : test « Chow » et le test de la covariance. Enfin il existe des tests de constructions plus complexes, comme la régression à paramètres variables, bien que l'objectif recherché soit différents, ou les modèles « Espace d'états », associés au filtre de Kalman.

Les régressions roulantes :

Les régressions roulantes consistent à estimer les paramètres de modèles successifs au pas de 1 ou de plusieurs observations sur la période de l'étude dans le sens début vers la fin (régression forward) soit dans le sens contraire, de la fin vers le début (régression backward).

Les paramètres ainsi que les caractéristiques de ces régressions (t de Student, R^2 , Durbin-Watson) sont alors représentées graphiquement. Dans le cas de fortes instabilités, ces graphiques font alors apparaître des ruptures.

Méthode du CUSUM :

Cette méthode du CUSUM est un moyen de détecter les changements de tendance, la forme analytique d'une tendance, dans une chronique. Pour cela considérons une chronique X_t quelconque prenant ces valeurs dans la période 1, n . La fonction CUSUM (j) associé au temps $t = j$, s'écrit de la manière suivant e :

$$\text{CUSUM}(j) = \sum_{t=1}^j X_t - \left(\frac{j}{m}\right) \sum_{t=1}^{\frac{n}{2}} X_t$$

$$\bar{X}_j(X_1, \dots, X_j), \quad j \leq n, \quad m \leq j$$

m est une valeur compris en 2 et n que l'on choisit le plus souvent à $\frac{n}{2}$ car l'on peut vérifier que les CUSUM pour les différentes valeurs de n ne sont pas très différentes.

Si la moyenne \bar{X}_j des premières j valeurs est constante, quelque soit $j \leq n$ et $m \leq j$ alors on peut montrer que le $\text{CUSUM}(j) = 0$. Sur ce constat, si la moyenne augmente (respectivement diminue) brusquement pour $j = j_0$ alors on peut vérifier que la fonction CUSUM devient linéairement croissante (respectivement décroissante) pour $j > j_0$. De même si la moyenne augmente (respectivement diminue) progressivement après $j = j_0$, la fonction CUSUM devient quadratique croissante (respectivement décroissante) pour $j > j_0$. On utilise donc ce principe sur les résidus du modèle de régression, qui par définition, ne possède pas de tendance. Si le CUSUM en détecte une [de tendance], cela signifie effectivement que le modèle est instable au cours du temps, et le CUSUM peut être un instrument pour détecter les ruptures dans les résidus, et, de ce fait, les ruptures dans les régressions.

Test de Chow :

Considérons les 3 modèles suivants :

$$\text{Modèle (1)} : Y_t = \alpha + \hat{\beta} X_t, \quad t \in [1, \dots, n]$$

$$\text{Modèle (2)} : Y_t = \alpha^1 + \hat{\beta}^1 X_t, \quad t \in [1, \dots, n_1]$$

$$\text{Modèle (3)} : Y_t = \alpha^2 + \hat{\beta}^2 X_t, \quad t \in [n_1 + 1, \dots, n]$$

L'auteur montre que ces 3 modèles peuvent être réécrites sous la forme unique suivante :

$$Y_t = \alpha^i + \beta^i X_t + \varepsilon_t, \quad t \in [1, \dots, n]$$

$$i = 1, \quad t = 1, \dots, n$$

$$i = 2, \quad t = n_1 + 1, \dots, n$$

$$\text{Modèle (4)} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ Y_{n_1+1} \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n_1} & 0 & 0 \\ 0 & 0 & 1 & X_{n_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & X_n \end{bmatrix} \begin{bmatrix} \alpha^1 \\ \alpha^2 \\ \beta^1 \\ \beta^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$H_0: \begin{cases} \beta^1 = \beta^2 \\ \alpha^1 = \alpha^2 \end{cases}$$

$$H_1: \begin{cases} \beta^1 \neq \beta^2 \\ \alpha^1 \neq \alpha^2 \end{cases}$$

$$F_c = \frac{n-4}{2} * \frac{SCR_0 - SCR_a}{SCR_a} \sim \mathcal{F}(2, n-4)$$

SCR_0 = Somme des carrés des résidus du modèle complet

SCR_a = Somme des carrés des résidus du modèle 4, le a signifie alternatif

On peut d'ailleurs vérifier que cette $SCR_a = SCR$ du modèle 2 + SCR du modèle 3

Le test de Chow ne s'applique que dans le cas où le modèle 3 ne concerne qu'un petit nombre d'observation, inférieur au nombre de variables explicatives du modèle : dans le cas SCR_a est la somme des carrés des résidus du modèle 2. Il porte souvent le nom de Test de la covariance et la quantité F est écrite dans la littérature ANACOVA.

$$\text{ANACOVA} = \frac{SCR_0 - SCR_a}{SCR_a} * \frac{n-2k}{k}$$

Sous H_0 ANACOVA $\sim \mathcal{F}(k, n-2k)$

Le test de Chow est une présentation un peu différente du test de comparaison de 2 coefficient de régression, il est souvent en économétrie car ces formules s'intègre facilement dans un logiciel de régression. Ce test ne s'applique pas uniquement à des périodes de temps mais aussi à des groupes d'individus. Dans les 2 cas, l'objet est souvent la recherche des points de rupture et, pour cela, on peut utiliser les procédures suivantes :

- Calculer les statistiques ANACOVA, par des régressions roulantes
La plus grande valeur trouvée d'ANACOVA peut être la valeur recherchée
- Calculer les statistiques ANACOVA, par la méthode du Stepwise Chow
Méthode qui permet de déceler l'existence de plusieurs sous-périodes, ce qui ne permet pas l'ANACOVA roulant. On calcule une statistique ANACOVA(r), où $r \rightarrow n_1 + 1, \dots, n$

Si ANACOVA(r) n'est pas significatif, on conclue qu'il n'y a pas de changement de structure important sur la période. Il est recommandé dans cette première procédure de choisir le n_1 le plus petit possible pour que la première estimation est un sens.

Si ANACOVA(r) est significatif quelque soit r , dans ce cas on conserve, pour r, r_0 qui maximise ANACOVA. La période 1, n est alors découpée en 2 sous périodes, de 1 à r_0 et de $r_0 + 1$ à n , et sur ces 2 sous périodes, on recalcule ANACOVA(r). On réitère tous ces calculs jusqu'à ce qu'ANACOVA(r) ne soit plus significatif.

Méthode de la régression à paramètres variables RPV de Cooley et Prescott :

Il s'agit d'une méthode d'estimation d'un modèle linéaire pour lequel les paramètres ne sont plus des constantes, comme dans le MLGS, mais soumis à des variations de type aléatoire.

$$Y_t = \beta_t K_t, \quad t = 1, \dots, n$$

$$\text{C-P} : \beta_t = \beta_t^p + u_t$$

$$\beta_t^p = \beta_{t-1}^p + v_t$$

p pour permanent : Composante permanente

Ce modèle RPV peut être utilisée ici, car il fournit après estimation l'évolution (et leur stabilité) des paramètres β_t au cours du temps.

Il est également possible d'utiliser des méthodes plus récentes, mais aussi beaucoup plus complètes, et qui sont rassemblé dans l'appellation modèle d'Espaces d'état. Ces modèles sont estimés par un algorithme informatique, complexes, du type algorithme non-linéaire. Ce filtre est utilisé actuellement dans les recherches récentes pour vérifier la stabilité des modèles qui calculent les ratios permettant de gérer les portefeuilles.

Chapitre II

La discussion des hypothèses du MLGS

Section 1 : Autocorrélation et Homoscédasticité de l'aléa

Dans l'estimation du MLGS, $Y = X\beta + \varepsilon$ par la méthode des MCO ou du MV, un certain nombre d'hypothèses au départ ont été formulées, en particulier, l'hypothèse concernant l'aléa qui est homoscédastique et non auto-corrélé. Dans ce chapitre nous détruisons cette hypothèse et fournissons des tests pour les détecter mais aussi des solutions pour les contourner. Une méthode générale d'estimation peut être proposée dans le cas de l'hétéroscédasticité et/ou d'autocorrélation des estimateurs. Cette méthode a été proposée par Aitken et la méthode porte le nom de Moindres Carrés Généralisés et le modèle de régression dans ce contexte est appelé Modèle Linéaire Générale. En fait, dans ce cours, nous mettrons l'accent sur 2 grandes hypothèses, celle qui porte sur l'aléa ε mais aussi celle qui concerne le rang de la matrice X , c'est-à-dire la colinéarité.

A. La méthode des Moindres Carrés Généralisés, la Méthode d'Aitken et MLG.

Cette méthode est une solution à l'estimation des paramètres du modèle de régression dans le cas où l'aléa est hétéroscédastique et/ou auto-corrélé. Considérons le modèle $Y = X\beta + \varepsilon$, de même dimension que dans le MLGS. Toutes les hypothèses du MLGS concernent ce modèle sauf pour l'aléa.

$$E[\varepsilon] = 0$$

$$E[\varepsilon\varepsilon'] = \Omega_\varepsilon \neq \sigma_\varepsilon^2 I_n$$

Les aléas peuvent être hétéroscédastique et/ou auto-corrélé.

Considérons un nouvel estimateur $\tilde{\beta}$ linéaire égale à MY

$$= M(X\beta + \varepsilon)$$

$\tilde{\beta}$ doit être sans biais :

$$E[\tilde{\beta}] = \beta$$

$$E[\tilde{\beta}] = E[MX\beta + M\varepsilon]$$

$$= E[MX\beta] + E[M\varepsilon]$$

$$= E[MX\beta] + ME[\varepsilon]$$

$$= E[MX\beta] = \beta$$

$$\Rightarrow MX = I$$

$$\Omega_{\tilde{\beta}} = E \left\{ [\tilde{\beta} - E[\tilde{\beta}]] [\tilde{\beta} - E[\tilde{\beta}]]' \right\}$$

$$E[\tilde{\beta}] = \beta \text{ si } MX = I$$

L'estimateur doit être de variance minimale

$$\Rightarrow \tilde{\beta} = MX\beta + M\varepsilon$$

$$= \tilde{\beta} + M\varepsilon = E[\tilde{\beta}] + M\varepsilon$$

$$\tilde{\beta} - E[\tilde{\beta}] = M\varepsilon$$

$$\Omega_{\tilde{\beta}} = E\{[M\varepsilon][M\varepsilon]'\}$$

$$= E\{M\varepsilon\varepsilon'M'\} = ME[\varepsilon\varepsilon']M'$$

$$\Omega_{\tilde{\beta}} = M\Omega_{\varepsilon}M'$$

Il faut maintenant diagonaliser les éléments sur $\Omega_{\tilde{\beta}}$

$$\Sigma = E \left\{ [\tilde{\beta} - E[\tilde{\beta}]]' [\tilde{\beta} - E[\tilde{\beta}]] \right\}$$

$$= E\{[M\varepsilon]'[M\varepsilon]\}$$

$$= E[\varepsilon'M'M\varepsilon]$$

$M'M$ symétrique

$$\varepsilon' \underbrace{(M'M)}_{\substack{\text{Matrice carré} \\ \text{symétrique}}} \varepsilon = \text{scalaire}$$

$$\text{tr}(M'M\varepsilon\varepsilon') = \sum_i m_{ii} \varepsilon_i^2 + 2 \sum \sum m_{ij} \varepsilon_i \varepsilon_j$$

$$\Rightarrow \varepsilon'(M'M)\varepsilon = \text{tr}(M'M\varepsilon\varepsilon')$$

$$E[\varepsilon'(M'M)\varepsilon] = E[\text{tr}(M'M\varepsilon\varepsilon')]$$

$$= \text{tr}[E(M'M\varepsilon\varepsilon')]$$

$$= \text{tr}[M'ME(\varepsilon\varepsilon')]$$

$$= \text{tr}[M' M \Omega_\varepsilon]$$

$$\Sigma = \text{tr}[M' M \Omega_\varepsilon]$$

Lagrangien :

$$\Phi = \text{tr}[M' M \Omega_\varepsilon] - \text{tr}[\Lambda' (MX - I)]$$

$$\Rightarrow \Phi = \text{tr}[M' M \Omega_\varepsilon] - \text{tr}[\Lambda' MX] + \text{tr}[\Lambda' I]$$

$$\text{Or : } \text{tr}[\Lambda' MX] = \text{tr}[(\Lambda' MX)'] = \text{tr}[X' M' \Lambda] = \text{tr}[M' \Lambda X']$$

$$\text{Donc } \Phi = \text{tr}[M' M \Omega_\varepsilon] - \text{tr}[M' \Lambda X'] + \text{tr}[\Lambda' I_n]$$

$$\frac{\partial \Phi}{\partial M} = \text{tr}[2M - \Omega_\varepsilon - \Lambda X' - 0] = 0$$

$$\Rightarrow 2M \Omega_\varepsilon = \Lambda X'$$

Donc :

$$2M \Omega_\varepsilon \Omega_\varepsilon^{-1} X = \Lambda X' \Omega_\varepsilon^{-1} X$$

$$\Rightarrow 2 = \Lambda (X' \Omega_\varepsilon^{-1} X)$$

$$\Lambda = 2(X' \Omega_\varepsilon^{-1} X)$$

$$\text{Donc } 2M \Omega_\varepsilon = 2(X' \Omega_\varepsilon^{-1} X) X'$$

$$M = (X' \Omega_\varepsilon^{-1} X)^{-1} X' \Omega_\varepsilon^{-1} \text{ et } \tilde{\beta} = [(X' \Omega_\varepsilon^{-1} X)^{-1} X' \Omega_\varepsilon^{-1}] Y$$

Est-ce un estimateur de variance minimale ?

$$\text{Soit : } \tilde{\beta} = (M + N)Y, \text{ avec } E[\tilde{\beta}] = \beta, MX = I, NX = 0$$

$$\text{On a : } \Sigma' = E \left[\left[\tilde{\beta} - E[\tilde{\beta}] \right] \left[\tilde{\beta} - E[\tilde{\beta}] \right]' \right]$$

$$\text{Nous devons montrer : } \Sigma' = \Sigma + \underbrace{E[\varepsilon' N' N \varepsilon]}_{\text{Fonction quadratique définie positive}}$$

Donc Σ est toujours supérieur à Σ' , donc $\tilde{\beta}$ correspond au minimum de la variance.

L'estimateur d'Aitken s'écrit donc de la façon suivante :

$$\begin{aligned}\tilde{\beta} &= (X' \Omega_\varepsilon^{-1} X)^{-1} X' \Omega_\varepsilon^{-1} Y \\ \Omega_{\tilde{\beta}} &= M \Omega_\varepsilon M' \\ &= [(X' \Omega_\varepsilon^{-1} X)^{-1} X' \Omega_\varepsilon^{-1}] \Omega_\varepsilon [(X' \Omega_\varepsilon^{-1} X)^{-1} X' \Omega_\varepsilon^{-1}]^{-1} \\ &= (X' \Omega_\varepsilon^{-1} X)^{-1} X' \Omega_\varepsilon^{-1} \Omega_\varepsilon \Omega_\varepsilon^{-1} X (X' \Omega_\varepsilon^{-1} X)^{-1} \\ &= \underbrace{(X' \Omega_\varepsilon^{-1} X)^{-1} (X' \Omega_\varepsilon^{-1} X)}_I (X' \Omega_\varepsilon^{-1} X)^{-1} \\ &= (X' \Omega_\varepsilon^{-1} X)^{-1}\end{aligned}$$

$$= \tilde{\beta} = (X' \Omega_\varepsilon^{-1} X)^{-1} X' \Omega_\varepsilon^{-1} Y$$

$$\Omega_\varepsilon = \sigma_\varepsilon^2 I_n \text{ (MLGS)}$$

$$\Omega_\varepsilon^{-1} = \sigma_\varepsilon^{-2} I_n$$

$$\tilde{\beta} = (X' \sigma_\varepsilon^{-1} X)^{-1} X' \sigma_\varepsilon^2 Y$$

$$\tilde{\beta} = (X' X)^{-1} X' Y \underbrace{(\sigma_\varepsilon^2 * \sigma_\varepsilon^{-2})}_{=1} = \hat{\beta}$$

$$\Omega_{\tilde{\beta}} = \sigma_\varepsilon^2 (X' X)^{-1} = \Omega_{\hat{\beta}}$$

B. Le problème de l'Autocorrélation

Pour vérifier les hypothèses concernant l'aléa, on utilise le résidu du modèle estimé. L'aléa ε_t est un processus aléatoire particulier dans la régression puisqu'à chacun des instants t , il est centré d'espérance nulle, homoscédastique, et non-auto corrélé, la covariance entre deux variables aléatoires est égale à 0.

Ce processus aléatoire est théorique, et pour les économistes, sauf si on réalise des simulations, il est inconnu. On fait alors l'hypothèse que le résidu du modèle de régression $e_t = Y_t - \hat{Y}_t$ est un échantillon (trajectoire/réalisation) du processus aléatoire. En d'autres termes, à un instant t où on réalise une expérience qui est réitéré, pour obtenir une variable aléatoire, on fait l'hypothèse que le résidu à l'instant t est un état, un des résultats de l'expérience.

Avec cette hypothèse, $E[\varepsilon_t] = 0$ se traduira par $\bar{e} = 0$. Cette hypothèse n'est pas, en général, testée, puisque par construction, on démontre que la moyenne des résidus du MLG est égale à 0. On dit que c'est une hypothèse ad hoc. $E[\varepsilon_t^2] = V[\varepsilon_t] = \sigma_\varepsilon^2$ & $E[\varepsilon_t \varepsilon_t'] = 0$, c'est deux hypothèse seront vérifiés et testés à partir de l'échantillon e_t , il faut donc définir par la suite, ce que l'on entend par variance, et covariance du résidu e_t , sachant que e_t c'est l'observation du résidu du modèle pour chacun des instants t , c'est-à-dire une série temporelle.

Dans ce paragraphe, on s'intéresse au problème de l'autocorrélation dans les résidus. La présence d'une autocorrélation se manifeste par des évolutions des résidus qui présentent plus de régularités qu'une série qui serait purement aléatoire. Ainsi on dira qu'il y a autocorrélation positive si une valeur e_t est fréquemment suivie par une valeur plus élevée, et réciproquement. On dira qu'il y a autocorrélation négative si une valeur élevée est fréquemment suivi par une valeur plus faible et réciproquement. Pour mettre en évidence cette autocorrélation, on fait appel au coefficient de corrélation linéaire simple que l'on appelle, le coefficient d'autocorrélation. C'est le coefficient de corrélation linéaire simple entre la chronique e_t et elle-même, décalée sur l'axe du temps. Si le décalage est d'un pas de temps, on calcule le coefficient de corrélation d'ordre 1, s'il est de deux pas de temps, il sera d'ordre 2, etc...

Il est possible de calculer ces autocorrélations avec différents décalages sur l'axe du temps, si on choisit des décalages à un pas de temps croissant, on pourra donc calculer, construire, une succession de coefficient d'autocorrélation $r_1, r_2, r_3, \dots, r_K$. K est le retard maximal entre les deux chronique des résidus décalés qui fournit une intersection suffisante pour calculer le coefficient de corrélation. Lorsque ces calculs sont réalisés, on peut alors mettre en correspondance les décalages $1, 2 \dots K$ avec les coefficients $r_1, r_2 \dots r_K$. Cette correspondance est une fonction des décalages $k \in [1, K]$, dans $r_k \in [-1; +1]$, cette correspondance est la fonction d'autocorrélation noté FAC, sa représentation graphique est un diagramme en bâton répondant au nom de corrélogramme.

Lorsque l'on peut constater une autocorrélation des résidus, on peut vérifier les points suivants qui peuvent être à l'origine de cette autocorrélation :

- Un facteur, une variable qui influence la variable endogène, de façon systématique, a été oublié dans la régression. Il faut dans ce cas repenser la théorie et réécrire le modèle.
- La tendance dans les variables ou encore la saisonnalité dans les variables n'ont pas été correctement éliminé, c'est bien une source d'autocorrélation.
- Le profil saisonnier varie au cours du temps et produit après son élimination des fluctuations parasites, c'est aussi un cas classique menant à de l'autocorrélation.
- Des perturbations, telles que les guerres, les grèves, les accidents climatiques, qui se produisent à plusieurs périodes, peuvent aussi créer des autocorrélations. Leurs éliminations (leurs prises en compte) par des méthodes de types « extrapolation linéaire » peuvent aussi être sources d'autocorrélation.
- Les erreurs de mesures sont aussi sources d'autocorrélation.

Il peut cependant aussi exister de l'autocorrélation dans les résidus pour des raisons non évoquées précédemment, ou encore, on souhaite conserver cette autocorrélation parce que l'on considère que les raisons précédentes doivent être conservées. Tous ces cas reviennent à considérer l'autocorrélation des résidus, dans l'estimation du MLGS, et donc, de faire appel d'utiliser le MLG à sa place.

1. Les effets et la détection de l'autocorrélation

Pour montrer les effets de l'autocorrélation des erreurs sur l'estimation des paramètres, on va utiliser le modèle de régression linéaire

$$y_t = \beta x_t + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t, \quad \eta_t : \text{Hypothèse}$$

$$\text{AR}(1) : |\rho| < 1$$

On va donc appliquer la méthode des moindres carrés ordinaires :

$$\begin{aligned} \hat{\beta} &= \frac{\sum x_t y_t}{\sum x_t^2} \\ &= \sum \omega_t y_t, \quad \omega_t = \frac{x_t}{\sum x_t^2} \\ &= \sum \omega_t (Y_t - \bar{Y}) \\ &= \sum \omega_t Y_t - \bar{Y} \underbrace{\sum \omega_t}_{=0} \end{aligned}$$

Rappel :

$$\begin{aligned} &= \sum \omega_t (\alpha + \beta X_t + \varepsilon_t) \\ &= \alpha \underbrace{\sum \omega_t}_{=0} + \beta \underbrace{\sum \omega_t X_t}_{=1} + \sum \omega_t \varepsilon_t \end{aligned}$$

$$\hat{\beta} = \beta + \sum \omega_t \varepsilon_t$$

$$\begin{aligned} E[\hat{\beta}] &= \beta + \sum \omega_t E[\varepsilon_t] \\ &= \beta \left\{ \begin{array}{l} \text{si } E[\varepsilon_t] = 0 \\ \text{ou avec } \sum \omega_t = 0 \end{array} \right. \end{aligned}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t$$

$$\varepsilon_{t-1} = \rho \varepsilon_{t-2} + \eta_{t-1}$$

$$\rightarrow \varepsilon_t = \rho[\rho\varepsilon_{t-2} + \eta_{t-1}] + \eta_t$$

Etc...

$$= \eta_t + \rho\eta_{t-1} + \rho^2\eta_{t-2} + \dots$$

AR(1) \leftrightarrow MA(∞) (*Mobile Average, Moyenne mobile*)

$$\varepsilon_t = \sum_{\theta=0}^{\infty} \rho^\theta \eta_{t-\theta}$$

$$E[\varepsilon_t] = \sum_{\theta} \rho^\theta E[\eta_{t-\theta}] = 0$$

$\hat{\beta}$ reste donc sans biais

$$\forall \rho : \text{Var}[\hat{\beta}] = \frac{\sigma_\varepsilon^2}{\sum x_t^2}$$

$$V[\varepsilon_t] = E[\varepsilon_t^2]$$

$$\varepsilon_t = \sum_{\theta} \rho^\theta \eta_{t-\theta}$$

$$= E[\eta_t^2] + \rho^2 E[\eta_{(t-1)}^2] + \rho^4 E[\eta_{(t-2)}^2] + \dots + 2\rho E[\eta_t \eta_{t-1}] + \dots$$

$$= \sigma_\eta^2 + \sigma_\eta^2 \rho^2 + \sigma_\eta^2 \rho^4 \dots$$

$$= \sigma_\eta^2 (1 + \rho^2 + \rho^4 + \dots) \text{ (Suite géométrique d'ordre inférieur à 1)}$$

$$E[\varepsilon_t^2] = \text{Var}[\varepsilon_t]$$

$$E[\varepsilon_t^2] = \frac{1}{1 - \rho^2} \sigma_\eta^2 = \sigma_\varepsilon^2$$

$$\Rightarrow V[\hat{\beta}] = \frac{\sigma_\eta^2}{1 - \rho^2} * \frac{1}{\sum x_t^2}$$

Comme le paramètre ρ en valeur absolue est inférieur à 1, on constate que si ρ tend vers 0 (non-autocorrélation) alors on retrouve les résultats de la régression. Donc si ρ tend vers 1 (forte autocorrélation), alors $V[\hat{\beta}] \rightarrow \infty$

En définitive, quand il y a autocorrélation

- Les estimateurs des paramètres du modèle restent sans biais
- Les variances d'échantillons des coefficients de régression ne sont plus minimales, la méthode des MCO n'est plus la meilleure parmi toutes les méthodes d'estimations linéaires. Cette méthode sous-estime les variances vraies dans le cas d'autocorrélation positive, par exemple, ce qui a pour conséquence une

surestimation de la précision de l'estimation. Dans le cas d'une prévision, on obtient alors des valeurs de l'endogène qui ne sont pas les meilleures possibles.

L'existence d'une autocorrélation des résidus remet en cause l'ajustement du modèle par les moindres carrés. Il est donc indispensable de disposer des tests permettant de détecter l'autocorrélation des résidus.

✚ La détection des tests d'autocorrélation des résidus

On peut les regrouper selon 2 catégories : ceux qui sont basés sur une distribution quelconque des résidus et ceux qui sont basés sur une distribution théorique des résidus.

- Le test Geary

Il fait partie de ces tests construits à partir d'une distribution quelconque des résidus. Il consiste à comptabiliser les changements de signes des résidus. Le nombre de résultats trouvés R est comparé au nombre de valeurs maximales R' et minimales R'' d' Y dans des tables de valeurs critiques élaborées pour des seuils de 1 et de 5% en fonction du nombre n d'observations.

Si $R < R''$ il y a autocorrélation positive, parce qu'il y a un petit nombre de changement de signe, par contre, si $R > R'$ il y a autocorrélation négative, parce qu'il y a un grand nombre de changement de signe. Ce test n'est pas très fiable pour de petits échantillons ($n < 20$)

- Le test de Durbin-Watson

Texte manquant

$$D-W = \frac{\sum_{t=1}^{n-1} (e_{t+1} - e_t)^2}{\sum e_t^2}$$

On peut supposer : $\sum_{t=1}^{n-1} e_{t+1}^2 \sim \sum_{t=1}^{n-1} e_t^2 \sim \sum_{t=1}^n e_t^2$

$$\text{Alors : } D-W = \frac{2 \sum e_t^2}{\sum e_t^2} - \frac{2 \sum e_{t+1} e_t}{\sum e_t^2}$$

$$e_{t+1} = \rho e_t + \eta_t \Leftrightarrow e_t = \rho e_{t-1} + \eta_{t-1}$$

Avec les MCO, on trouve : $D-W = 2(1 - \hat{\rho})$

On constate avec cette expression que lorsque les erreurs ne sont pas auto corrélées ($\hat{\rho} \sim 0$) alors le Durbin-Watson approche les 2. Au contraire, si le $\hat{\rho} \sim 1$, le $D-W \sim 0$ et si $\hat{\rho} \sim -1$, le $D-W \sim 4$.

On constate donc que le Durbin-Watson est compris entre 0 et 4 et que la non-autocorrélation passe par la valeur 2 de la statistique.

La statistique de Durbin-Watson dépend de e_t . Or, $e_t = Y_t - \hat{Y}_t$, donc :

$$e_t = \alpha + \beta X_t + \varepsilon_t - \hat{\alpha} - \hat{\beta} X_t = (\alpha - \hat{\alpha}) + X_t(\beta - \hat{\beta}) + \varepsilon_t$$

✚ Test H_0 : Non-autocorrélation d'ordre 1

e_t dépend aussi de X_t , dans ces conditions, il est difficile de trouver le niveau de signification exactes pour ω_t .

Pour k' étant le nombre de variables exogènes, la statistique de Durbin-Watson est comprises entre d_1 et d_2 indépendantes des X_t . d_1 et d_2 sont des variables aléatoires fonctions de ε_t et sous les hypothèses classiques de régressions, les auteurs ont proposés les tables.

Le test se déroule ainsi : On calcule le Durbin-Watson puis on construit le tableau suivant :

0	d_1	d_2	2	$4 - d_2$	$4 - d_1$	4
Autocorrélation positive	Doute tendant vers l'AC > 0	H_0 acceptée Absence d'autocorrélation		Doute tendant vers l'AC < 0		Autocorrélation négative

Les tables de Durbin-Watson permettent de trouver d_1 et d_2 aux seuils de 1 et 5%, pour $n > 15$ (si $n < 15$, on interpole en arrière)

Pour $n > 100 \Rightarrow \frac{DW_c - \frac{2(n-1)}{n-k}}{2\sqrt{n+2}}$

✚ Utilisation de la fonction d'autocorrélation :

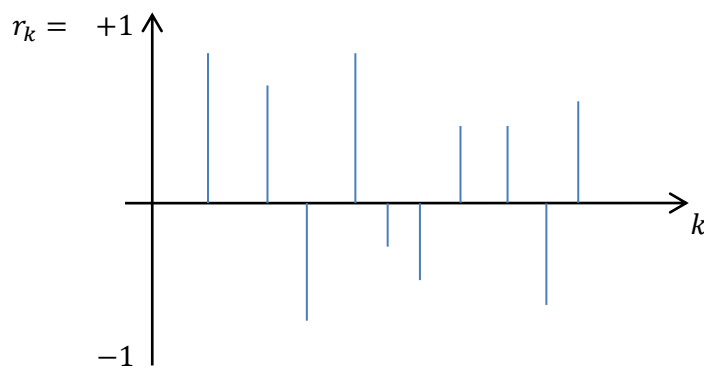
$k \in \mathbb{N} \rightarrow r_k[-1; 1]$

k est le coefficient de corrélation linéaire simple, calculée entre une variable et elle-même, décalée sur l'axe du temps. Le décalage s'effectue à un pas de temps. Si on effectue un décalage de 1 sur la partie commune des deux résidus, on calcule un coefficient de corrélation que l'on appelle r_1 , avec un décalage de 2, un coefficient r_2 , etc...

k	r_k
0	1
1	r_1
\vdots	\vdots
K	r_K

K = Décalage maximal

$$\frac{n}{6} \leq K \leq \frac{n}{3}$$



On peut tester le corrélogramme :

On peut effectuer des tests individuels de r_0 ,

Test de Student $\sim \mathcal{N}$

$$H_0 : \rho_k = 0$$

$$H_1 : \rho_k \neq 0$$

Ou des tests joints (tests du porte-manteau)

Les tests ont été conçus par un statisticien du nom de Quenouille, l'auteur démontre que la somme des carrés cumulés des autocorrélations successives obéit à un χ^2 à K degré de liberté. Sur ce constat, un premier test a été construit par 2 auteurs Box-Pierce, puis plus tard par Ljung-Box, en montrant que la statistique de Box-Pierce était insuffisante et propose une nouvelle statistique, et le test est alors le suivant :

$$H_0 : \rho_1 = \dots = \rho_K = 0$$

$$H_1 : \rho_k \neq 0, \forall k \in [1, \dots, K]$$

$$Q = n(n+1) \sum_{k=1}^K \frac{r_k^2(\rho_k)}{n-k}, \text{ en supposant l'homoscédasticité}$$

Si il y hétérosécédasticité, on utilise la statistique de Box-Pierce corrigée

$$Q < \chi^2(K), \quad H_0 \text{ acceptée}$$

2. Solutions

Lorsqu'il y a autocorrélation des résidus, on essaie d'abord de détecter quel est l'origine de cette autocorrélation. C'est ce qui a été indiqué dans l'introduction de ce paragraphe. Souvent, l'autocorrélation est le résultat d'une forme inadaptée du modèle de régression. Il est alors conseillé de représenter graphiquement la régression entre le résidu e_t et les différentes variables explicatives du modèle. Si ce graphe, le nuage de régression révèle pour la plupart des

formes non linéaires, cela peut être une indication pour transformer le MLGS de départ (on utilisera par exemple le logarithme des variables).

Dans certains cas, pour des raisons de théories économiques, on souhaite conserver dans le modèle de régression, avec l'autocorrélation des résidus. Pour cela, plusieurs techniques ont été proposées :

- On peut utiliser, par exemple, le modèle autorégressif des résidus pour transformer les variables du modèle.

Si le modèle autorégressif des résidus est de retard, on transformera le modèle en utilisant l'estimation de $\hat{\rho}$ du modèle autorégressif, et les variables transformées seront donc : $Y_t - \hat{\rho}Y_{t-1}, X_{2t} - \hat{\rho}X_{2t-1}$, pour toutes les variables régressives. Certains auteurs, comme Stone, proposent d'utiliser pour $\hat{\rho}$, la valeur 1, la transformation porte alors le nom de différences premières, et s'écrit $Y_t - Y_{t-1}, X_{2t} - X_{2t-1}$, etc...

En réalité cette transformation est inadéquate a priori puisqu'elle est contraire à l'une des hypothèses du MLGS, celle qui concerne la stationnarité des variables (cf. Cours du M1). Cette façon de procéder à partir de l'estimation du modèle autorégressif des résidus, porte le nom de méthodes des pseudo-moindres carrés généralisés. Il existe dans la littérature deux versions plus complètes de ces transformations que l'on appelle la méthode de Hildreth-Lu et Cochrane-Orcutt

La méthode de Hildreth-Lu considère les paramètres a, b et ρ du modèle de régression suivant :

$$Y_t - Y_{t-1} = a(1 - \rho) + b(X_t - \rho X_{t-1}) + \varepsilon_t$$

$$\text{Si } \rho = 1 \begin{cases} Y_t - Y_{t-1} \\ X_t - X_{t-1} \end{cases}$$

$$\rho \in]-1; +1[\rightarrow]-1; 0[;]0; 1[$$

Hildreth-Lu utilise une méthode dichotomique. Il prend une valeur de ρ au milieu des deux intervalles précédemment découpés (-0,5 et 0,5), et pour chacune des deux valeurs de ρ , il calcule \hat{Y} du modèle de régression, et en déduit, la somme des carrés des résidus et conserve l'intervalle où la somme est la plus petite. Il réitère l'opération jusqu'à ce que l'intervalle trouvé soit inférieure à une distance défini au préalable.

La méthode de Cochrane-Orcutt, qui est plus populaire que la précédente, procède de la façon suivante :

On estime la valeur de ρ au départ par les MCO, sur le modèle autorégressif des résidus. On transforme alors les variables en $Y'_t = Y_t - \hat{\rho}Y_{t-1}$ et $X'_t = X_t - \hat{\rho}X_{t-1}$, et on réitère la régression de départ sur les variables transformées Y'_t et X'_t . Ce que Cochrane-Orcutt, c'est de recommencer toute l'opération en estimant à nouveau le ρ à partir des résidus de la nouvelle régression $\hat{\rho}$, et de retransformer les variable précédente en $Y''_t = Y'_t - \hat{\rho}Y'_{t-1}$ et $X''_t = X'_t - \hat{\rho}X'_{t-1}$ et les auteurs

proposent de recommencer cette procédure jusqu'à ce que les estimateurs de la régression convergent, c'est-à-dire, jusqu'à ce qu'il ne se modifie plus. On peut vérifier que les 2 méthodes donnent à peu près le même résultat.

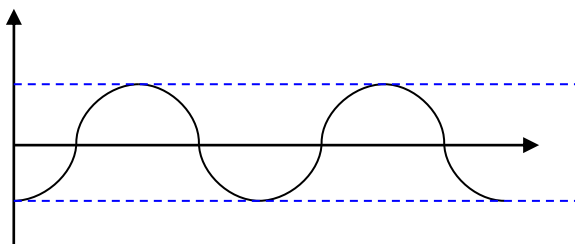
C. Hétéroscédasticité

1. Détection-effets

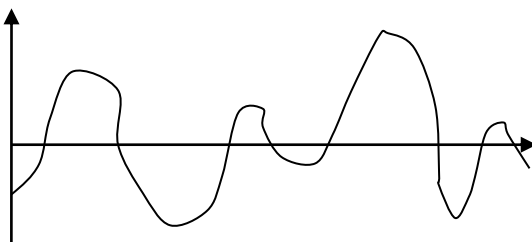
On considère à présent que Ω_ε s'écrit :

$$\Omega_\varepsilon = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$

Nous pouvons comparer le graphique des individus avec celui du processus de Bruit Blanc, c'est-à-dire avec un échantillon de Bruit Blanc centré, homoscédastique et non auto corrélé.



Cas d'homoscédasticité



Cas d'hétéroscédasticité

Il existe des tests plus précis qui permettent de détecter cette hétéroscédasticité.

✚ Test de Goldfeld-Quandt

Ce test s'appuie sur l'hypothèse suivante : il y a hétéroscédasticité parce que l'écart type du modèle de l'erreur augmente proportionnellement avec l'une des variables explicatives du modèle : par exemple la variable X_j . Cette hypothèse s'écrit donc :

$$H_0 : \begin{cases} \sqrt{E[\varepsilon_t^2]} = aX_{jt} \\ E[\varepsilon_t^2] = a^2X_{jt}^2 \end{cases}$$

Le test se déroule de la façon suivante :

- On ordonne les observations de toutes les variables du MLGS en fonction des valeurs croissantes de la variable X_j .
- On néglige les observations centrales du résultat obtenu et on appelle m le nombre de ces variables négligés. Comme m est différent n , on prend pour $n = 30$, $m = 8$, pour $n = 60$, $m = 16$, etc...

On obtient ainsi 2 sous échantillons de cardinal $\frac{n-m}{2}$, l'un qui correspond aux petites observations de X_j et l'autre aux plus fortes observations de X_j . On réalise alors la régression et on estime le MLGS sur ces deux sous échantillons, et on appelle SCR_1 la somme des carrés des résidus du premier sous échantillon et SCR_2 la somme des carrés des résidus du second sous échantillon.

Les auteurs ont alors montré que s'il y a hétéroscédasticité, alors le rapport obéit à une loi de Fisher avec un degré de liberté précisé. Le test se déroule donc de la façon suivante : l'hypothèse H_0 , c'est qu'il y a homoscedasticité et H_1 , hétéroscédasticité. On calcule le rapport :

$$\frac{SCR_2}{SCR_1} < \mathcal{F}_k \left(\frac{n-m-2k}{2}; \frac{n-m-2k}{2} \right)$$

✚ Test de Glejser

Les hypothèses de ce test sont identiques à la précédente. L'auteur propose de régresser la valeur absolue du résidu en fonction de X_j et propose aussi différents types de relations pour cette relation.

$$|e_t| = a_0 + a_1X_{jt}, \text{ régression linéaire}$$

$$|e_t| = a_0 + \frac{a_1}{X_{jt}}$$

$$|e_t| = a_0 + a_1\sqrt{X_{jt}}$$

On estime l'un de ces modèles par les MCO, et si les paramètres ne sont pas significativement différents de 0 on considère qu'il y a homoscedasticité. Considérons par exemple le premier modèle :

$$|e_t| = \alpha_0 + \alpha_1 X_{jt} + \varepsilon_t$$

$$\varepsilon_t = v_t(\alpha_0 + \alpha_1 X_{jt})$$

$$E[v_t] = 0$$

$$E[v_t^2] = \sigma_v^2$$

$$E[v_t v_{t'}] = 0$$

$$\Rightarrow E[\varepsilon_t^2] = \sigma_t^2 [\alpha_0 + \alpha_1 X_{jt}]^2$$

$$= [\alpha_0 \sigma_v + \alpha_1 X_{jt} \sigma_v]^2$$

$$\alpha_0 \sigma_v = \alpha_0, \quad \alpha_1 \sigma_v = \alpha_1 \Rightarrow \begin{cases} E[\varepsilon_t^2] = (\alpha_0 + \alpha_1 X_{jt})^2 \\ E[\varepsilon_t \varepsilon_{t'}] = 0, \quad t \neq t' \end{cases} \Rightarrow \Omega_\varepsilon = E[\sigma_v^2]$$

Dans la pratique :

$$|e_t| = \alpha_0 + \alpha_1 X_{jt} + \varepsilon_t$$

La matrice des variances-covariances Ω_ε indique une absence d'autocorrélation et une présence d'hétéroscedasticité avec l'hypothèse précédemment décrite. On applique donc la méthode des MCO sur le modèle qui conduit au modèle estimé :

$$|\hat{e}_t| = \hat{\alpha}_0 + \hat{\alpha}_1 X_{jt}$$

Les hypothèses sont les suivantes :

H_0 : homoscedasticité, H_1 : hétéroscedasticité

Ce test de Glejser est un test intéressant parce qu'il permet de résoudre le problème de l'estimation MLG. En effet, puisque l'on connaît $\hat{\alpha}_0$ et $\hat{\alpha}_1$ on peut construire $\hat{\Omega}_\varepsilon$ matrice de variance covariance des erreurs :

$$\hat{\Omega}_\varepsilon = \begin{bmatrix} (\hat{\alpha}_0 + \hat{\alpha}_1 X_{j1})^2 & & 0 \\ & \ddots & \\ 0 & & (\hat{\alpha}_0 + \hat{\alpha}_1 X_{jt})^2 \end{bmatrix}$$

Estimateur d'Aitken : $\tilde{\beta}_\varepsilon^{-1} = (X' \hat{\Omega}_\varepsilon^{-1} X)^{-1} X' \hat{\Omega}_\varepsilon^{-1} Y$

$$\hat{\Omega}_\varepsilon^{-1} = \begin{bmatrix} 1 & & 0 \\ (\hat{\alpha}_0 + \hat{\alpha}_1 X_{j1})^2 & & \\ 0 & \ddots & \\ & & \ddots \end{bmatrix}$$

D'autres tests existent, le premier d'entre eux est le test ARCH (Autorégressif Conditionnellement Hétéroscédastique), créé par Engle, en 82, pour modéliser la volatilité des cours boursiers.

On applique la méthode des MCO au modèle, on estime les paramètres

$$e_t^2 = \text{cst} + \Phi_1 e_{t-1}^2 + \dots + \Phi_p e_{t-p}^2 + \varepsilon_t$$

$$e_t^2 = \text{cst} + \varepsilon_t$$

$$V[e_t] = \frac{\sum e_t^2}{n} = \text{cst} + \underbrace{\bar{\varepsilon}}_{=0} \rightarrow \text{homoscédasticité}$$

$$H_0 : \Phi_1 = \dots = \Phi_p = 0 : \text{homoscédasticité}$$

$$\text{MCO} : nR^2$$

$$\Rightarrow nR^2 \sim \chi^2(p)$$

$$nR^2 < \chi_{lu}^2(p), \quad H_0 \text{ acceptée}$$

Le test de Breusch-Pagan est un test similaire à l'ARCH mais prend comme critère la SCR_2

✚ Le test de White

$$e_t = \text{cst} + \alpha_1 X_{1t} + b_1 X_{1t}^2 + \alpha_2 X_{2t} + b_2 X_{2t}^2 + \dots + \alpha_k X_{kt} + b_k X_{kt}^2 + (\text{éventuellement}) c_1 X_{1t} X_{2t} + \dots + \varepsilon_t$$

$$H_0 : a_1 = b_1 = \dots = a_k = b_k$$

$$H_1 : a_i \neq 0, \quad \forall i \in [1, k]$$

2. Solutions

L'hétéroscédasticité est un cas particulier de l'autocorrélation on peut en conclure que les effets de l'hétéroscédasticité sont identique à ceux de l'autocorrélation, à savoir que les estimateurs du MLGS restent sans biais mais ne sont plus de variance minimale. Comme précédemment, il faut proposer différentes propositions.

✚ Utilisation du MLG des estimateurs d'Aitken

$$Y = X\beta + \varepsilon$$

$$\Omega_\varepsilon = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix} = \text{hétéroscédasticité non auto corrélé} \Rightarrow \text{permet d'estimer } V$$

(V étant en fait $\widehat{\Omega}_\varepsilon$)

$$\text{Aitken : } \tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

$$E[\tilde{\beta}] = \beta$$

$$V[\tilde{\beta}] = D(X'V^{-1}X)^{-1}$$

$$\varepsilon \sim \mathcal{N}(0, \sqrt{V})$$

$$\tilde{\beta} \sim \text{MV}$$

On peut utiliser aussi les MCO sur le modèle transformé au départ :

$$Y = X\beta + \varepsilon \text{ (hétéroscédasticité)}$$

$$TY = TX\beta + T\varepsilon$$

$$Y_T = X_T\beta + \varepsilon_T$$

$$E[\varepsilon_{Tt}^2] = a^2, \quad \forall t \text{ (homoscédasticité)}$$

$$E[\varepsilon_T \varepsilon_T'] = a^2 I_n, \quad \text{(homoscédasticité non auto corrélé)}$$

\Rightarrow MCO au modèle transformé

$$\hat{\beta}_T = (X_T' X_T)^{-1} X_T' Y_T$$

$$V[\hat{\beta}_T] = a^2 (X_T' X_T)^{-1}$$

On fait comme si on connaissait le modèle transformé pour arriver au résultat suivant :

$$\Rightarrow \hat{\beta}_T = (X_T' X_T)^{-1} X_T' Y_T$$

$$X_T = TX \text{ et } Y_T = TY \Rightarrow ((TX)'(TX))^{-1} (TX)' TY$$

$$\hat{\beta}_T = (X'T'TX)^{-1} X'T'TY$$

$$\rightarrow V[\hat{\beta}_T] = a^2 (X'T'TX)^{-1} = \left(X' \frac{(T'T)}{a^2} X \right)^{-1}$$

$$\Rightarrow \text{Nous voulons : } \hat{\beta}_T = \tilde{\beta} \text{ soit } V[\hat{\beta}_T] = V[\tilde{\beta}]$$

$$\text{Or } V[\tilde{\beta}] = (X'V^{-1}X)^{-1}$$

Donc nous devons avoir :

$$V^{-1} = \frac{T'T}{a^2}$$

Si cela est possible, on obtient alors des estimateurs avec $X_T = TX$ et $Y_T = TY$

Exemple :

$$\sigma_t^2 = a^2 \lambda_t, \quad t \in [1, n]$$

$$V = a^2 \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \Rightarrow \text{Hétéroscédasticité car les } \lambda \text{ sont différents}$$

$$V^{-1} = \frac{1}{a^2} \begin{bmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_n} \end{bmatrix}$$

$$\text{Essayons } T = \begin{bmatrix} 1/\sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\lambda_n} \end{bmatrix}$$

$$\text{Donc : } \frac{T'T}{a^2} = V^{-1}$$

Autre exemple :

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

$$\sigma_t = aX_t, \quad (\varepsilon_t \text{ est relié linéairement à } X_t)$$

$$\Rightarrow T = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\lambda_n}} \end{bmatrix} = \begin{bmatrix} \frac{1}{X_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{X_n} \end{bmatrix}$$

$$TY = \begin{bmatrix} Y_1/X_1 \\ \vdots \\ Y_n/X_n \end{bmatrix}; TX = \begin{bmatrix} \frac{1}{X_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{X_n} \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} 1/X_1 & 1 \\ \vdots & \vdots \\ 1/X_n & 1 \end{bmatrix}$$

$$T\varepsilon = \begin{bmatrix} \varepsilon/X_1 \\ \vdots \\ \varepsilon/X_n \end{bmatrix}$$

$$\Leftrightarrow \frac{Y_t}{X_t} = \frac{\beta_0}{X_t} + \beta_1 + \frac{\varepsilon_t}{X_t}$$

Donc par les MCO :

$$\left(\frac{Y_t}{X_t}; \frac{1}{X_t} \right) \Leftrightarrow (\hat{\beta}_0; \hat{\beta}_1)$$

Conclusion (Récapitulatif) :

$$\Omega_\varepsilon = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix} : \text{Matrice diagonale, régulière et définie positive}$$

Nous avons vu les 2 cas :

- Ω est connue (on l'estime via 2 méthodes différentes : soit par le MLG, soit par une transformation, puis par les MCO)
- Ω n'est pas connue, on ne peut pas l'estimer, on ne connaît pas T

Les travaux d'Engle ont montré la solution par la conception du modèle ARCH

Soit un modèle $AR(1)$ centré : $y_t = \Phi y_{t-1} + \varepsilon_t$

$$E[\varepsilon_t] = 0 ; E[\varepsilon_t \varepsilon_t'] = 0$$

$$V[\varepsilon_t] = \sigma^2$$

$$|\Phi| < 1 \text{ (Sinon la variance deviendrait infinie)}$$

$$\text{Explication : } y_t = \sum_{i=0}^{\infty} \Phi^i \varepsilon_{t-i} : \text{Modèle de } Moving \text{ Average}$$

$$E[y_t] = 0 = \text{Espérance non conditionnelle}$$

$$\text{Appelons } \Psi_{t-1} = \{y_{t-1}, \dots\}$$

$$E[y_t / \Psi_{t-1}] = \Phi y_{t-1} = \text{Moyenne conditionnelle}$$

$$\text{Donc si } \Phi \neq 0 \text{ alors } E[y_t] \neq E[y_t / \Psi_{t-1}]$$

Dans le cadre du test ARCH :

$$\varepsilon_t / \varepsilon_{t-1} \sim \mathcal{N}(0; k_t), \quad \text{avec } \begin{cases} k_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \\ \alpha_0 > 0 \text{ et } \alpha_1 \geq 0 \end{cases}$$

D'autres modèles existent tels que le modèle GARCH, ARCH-M, etc...

Section II : La multi-colinéarité et la relation variable expliquée avec le résidu

A. Introduction

Dans cette section, nous allons étudier 2 hypothèses du MLGS, qui ne concernent pas directement l'aléa comme dans la section précédente (le MLG). Il s'agit de l'hypothèse de la relation entre la variable explicative et l'aléa. Jusqu'à présent, le MLG et le MLGS, on fait l'hypothèse que les variables explicatives sont certaines et que la covariance entre l'aléa et ces variables est égale à 0. Cette hypothèse a servi pour effectuer toutes les démonstrations concernant le MLGS et le MLG. Nous étudions aussi dans cette section l'hypothèse concernant le rang de $X = k$, qui est l'hypothèse de la possibilité de calculer le vecteur $\hat{\beta}$ du MLGS et le vecteur $\tilde{\beta}$ du MLG

B. Hypothèse de la relation entre la variable explicative et l'aléa

Dans le cas où il y a dépendance entre les variables explicatives et l'aléa, on peut démontrer que l'estimation par les MCO ne conduit pas à des estimateurs convergents, même si n , le nombre d'observations, devient grand. Pour contourner cette difficulté, on recherche des variables qui sont non corrélées avec ε , que l'on appelle variable instrumentale, mais qui sont corrélées avec les variables explicatives, et cette façon de procéder conduit à la recherche d'un estimateur des variables instrumentales. Il est donc important, comme pour les autres hypothèses de la section 1 de ce chapitre, de pouvoir détecter l'existence d'une corrélation entre les variables explicatives et ε : c'est justement l'objet du test de Hausman, de 1978, qui utilise cet estimateur des variables instrumentales.

a. L'estimateur des variables instrumentales (V.I)

$$Y \begin{matrix} (n, 1) \\ = \\ (n, k) \end{matrix} X \begin{matrix} \beta \\ (k, 1) \end{matrix} + \begin{matrix} \varepsilon \\ (n, 1) \end{matrix}$$

On va considérer k variables, constante comprise :

$$Z_{1t}, Z_{2t}, \dots, Z_{kt} \rightarrow Z$$

Z : Matrice des k variables instrumentales, non corrélées avec ε

$$Z / Cov(Z' \varepsilon) = 0 \text{ ou } E[Z_{it} \varepsilon_t] = 0, \quad \forall i = 1, \dots, k \text{ et } t = 1, \dots, n$$

Pré-multiplions le MLGS par la matrice Z'

$$Z'Y = Z'X\beta + Z'\varepsilon$$

Par hypothèse :

$$Z'\varepsilon = 0 \Rightarrow Z'Y = Z'X\beta$$

$$\text{Si } |Z'X| \neq 0 : \hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$$

$\hat{\beta}_{VI}$ converge vers β

La matrice Z contient les variables instrumentales qui peuvent être des variables contenues dans X . Par construction, ces variables instrumentales doivent être corrélées avec celles de X .

$$\text{Cov}(Z'X) \neq 0$$

Si tel n'est pas le cas, l'estimateur précédent n'existerait pas. Si on pose que \hat{X} est le produit matriciel suivant :

$$\hat{X} = Z(Z'Z)^{-1}Z'X$$

$$\text{Alors } \Rightarrow \hat{\beta}_{VI} = (X'X)^{-1}X'Y = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$$

$$\text{On peut montrer alors : } \Omega_{\hat{\beta}_{VI}} = \sigma_{\varepsilon}^2(\hat{X}'\hat{X})^{-1}$$

De ce fait, on peut à nouveau réaliser des tests. Cette méthode convient par exemple lorsqu'il y a des erreurs de mesures sur les variables (comme dans les données d'enquêtes, comme dans les données agrégées, etc...). Est-ce qu'il y a indépendance entre les variables explicatives et l'aléa ?

b. Le test d'Hausman

Ce test a pour hypothèse :

H_0 : non corrélation entre X et ε

H_1 : corrélation entre X et ε

Sous l'hypothèse H_0 , la méthode des variables instrumentales et la méthode des MCO du chapitre 1, conduisent à des estimateurs convergents. Sous cette hypothèse, on retient la méthode des MCO qui est plus précise. Sous l'hypothèse H_1 , les estimateurs des MCO ne sont plus convergents ; dans ce cas, on retient la méthode des variables instrumentales.

On utilisera l'estimateur de Wald :

$$W = (\hat{\beta}_{VI} - \hat{\beta})' \left[\hat{\sigma}_\varepsilon^2 \left[(\hat{X}'X)^{-1} - (X'X)^{-1} \right] \right]^{-1}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{e'e}{n-k} = \frac{e'e}{n-K-1}, \quad K = k + 1$$

k : Nombre de variables explicatives

K : Nombre de variables explicatives + la constante

$$W \sim \chi^2 = W < \chi^2 \rightarrow H_0$$

$\nu^1 = k + 1$ si X et Z n'ont pas de variables communes

$\nu = k - k'$ si X et Z ont k' variables communes

C. La colinéarité

Dans le MLGS, l'utilisation des MCO ou du maximum de vraisemblance suppose que les variables X_j sont linéairement indépendantes, j variant de 1 à k . On dit qu'il y a colinéarité parfaites si ces variables sont reliées entre elles par des variables de types linéaires. Ces cas de colinéarités strictes sont rares (exemple : la dessaisonalisation sur variables dichotomiques) ; lorsqu'ils se produisent, il faut alors modifier le modèle de départ, comme nous l'avons fait pour la dessaisonalisation. On peut aussi supprimer dans un modèle de régression l'une des variables responsables de cette colinéarité stricte. Dans la majorité des cas, il n'existe pas de colinéarité stricte mais il existe des quasi-colinéarités, c'est à dire des relations dont le coefficient de corrélation est proche de 1 sans être égale à 1. Dans ces cas de colinéarités les calculs des estimateurs peuvent être entièrement erronés.

1. Effets de la quasi-colinéarité

✚ Premier effet : Les estimateurs des MCO tendent à être très grands en valeur absolu.

$$\hat{\beta} = (X'X)^{-1}X'Y$$

A orthogonal = \wedge

¹ ν = Lettre grecque « nu »

On choisit Λ la matrice diagonale des valeurs propres à laquelle on associe la matrice A des vecteurs propres qui est une matrice orthogonale, on peut alors écrire :

$$X'X = A' \Lambda A$$

$$|X'X| = \prod \lambda_i$$

Si $|X'X| = 0 \Rightarrow \prod \lambda_i = 0$

Il y a alors colinéarité entre les variables explicatives de départ.

$$(X'X)^{-1} = \sum_i \lambda_i^{-1} A_i' A_i$$

$$\hat{\beta} = \sum_i \lambda_i^{-1} A_i' A_i (X'Y)$$

Si quasi-colinéarité stricte, une ou plusieurs $\lambda_i \rightarrow 0$, alors la quantité $\lambda_i^{-1} \rightarrow \infty$ et $\hat{\beta} \rightarrow$ grandes valeurs.

✚ Deuxième effet : les variances covariances des estimateurs tendent à être très grandes

Pour montrer ce résultat, considérons :

y, x centrées et réduites

$$\Rightarrow R = \frac{1}{n} x'x$$

$$\begin{cases} \hat{\beta} = (x'x)^{-1} x'y \\ \Omega_{\hat{\beta}} = \sigma_\varepsilon^2 (x'x)^{-1} \Rightarrow V[\hat{\beta}] = c_{ii} \sigma_\varepsilon^2 \end{cases}$$

c_{ii} = Eléments diagonaux. On peut démontrer qu'ils sont égaux à $1/(1 - R_i^2)$

S'il y a colinéarité entre 2 ou plusieurs variables explicatives :

$$R_i^2 \rightarrow 1 \Rightarrow c_{ii} \rightarrow \infty$$

$$\Rightarrow V[\hat{\beta}] : \text{grande}$$

✚ Troisième effet : On peut montrer que la variance totale des $\hat{\beta}$ devient grande.

$$\Omega_{\hat{\beta}} = \sigma_\varepsilon^2 (x'x)^{-1}$$

$$\text{Variance totale des } \hat{\beta} = \text{tr}(\Omega_{\hat{\beta}}) = \sigma_\varepsilon^2 \text{tr}(x'x)^{-1}$$

Si les variables sont orthogonales (donc non-colinéarité), alors :

$$\frac{1}{n}x'x = I_{k-1} \Rightarrow \sigma_{\varepsilon}^2 \text{tr}(x'x)^{-1} = \frac{\sigma_{\varepsilon}^2(k-1)}{n}$$

Exemple :

$$x'x = A' \wedge A$$

$$\text{tr}(x'x) = \text{tr} \wedge = \sum \lambda_i \text{ (Somme des valeurs propres)}$$

$$\text{tr}(x'x)^{-1} = \sum \lambda_i^{-1} = \sum \frac{1}{\lambda_i}$$

$$\Rightarrow \sigma_{\varepsilon}^2 \text{tr}(x'x)^{-1} \Rightarrow \text{Variance Total des } \hat{\beta} \rightarrow \infty$$

Il existe d'autres effets, mais moins récurrents :

- Il est possible que la colinéarité amène qu'un estimateur puisse disposer du mauvais signe (- au lieu de +).
- La sensibilité qu'il peut exister à la variation d'une variable explicative : quand on rentre une autre variable explicative au modèle, il peut y avoir colinéarité.
- La variance/covariance peut être grande avec un test de Student indiquant que l'estimateur est négligeable.

2. Détections de la colinéarité

Cette détection peut être réalisée par des ratios ou par des tests.

- a. Le calcul de R (déjà vu)
- b. Le test de Klein (1962)

$$R = \frac{1}{n}x'x ; R^2 : \text{coefficient de détermination du modèle}$$

On calcul la matrice des coefficients de corrélations entre variables explicatives :

$$\text{Si } r_{x_i x_j}^2 > R^2 : \text{Présomption de colinéarité}$$

c. Le test VIF (Variance Inflation Factor), de Marquardt (1960)

$$\text{VIF}_i = \frac{1}{1 - R_i^2}$$

Soit X_2, X_3, X_4

$$R^2_{\begin{matrix} X_2, X_3 \\ X_2, X_4 \\ X_2, X_3, X_4 \\ X_3, X_4 \end{matrix}} \text{ sont tous des VIF}$$

Il y a colinéarité si $\text{VIF} \rightarrow \infty$

d. Le test de Farrar et Glauber (1967)

H_0 : X_j sont des variables explicatives orthogonales ou $D = 1$

$$\text{ex : } \begin{vmatrix} X_1 & X_2 \\ 1 & 0 \\ 0 & 1 \end{vmatrix} = 1, \quad x_1 \text{ orthogonal } x_2, \text{ non colinéarité}$$

La matrice ici est la matrice des coefficients de corrélation

H_1 : Présomption de colinéarité $|R| \neq 1$

$$\text{FG} = - \left(n - 1 - \frac{1}{6} [2(k + 1) + 5] \right) \ln D$$

k : Nombre de variables explicatives sans la constante

Si $\text{FG} \leq \chi_{1/2}^2 k(k + 1)$, l'hypothèse d'orthogonalité est rejetée

$$\text{FG} = - \left(n - 1 - \frac{1}{6} (2K + 5) \right) \ln D \sim \chi_{1/2}^2 K(K - 1)$$

e. Le test de Haitovsky (1969)

H_0 : $|R| = 0 \Leftrightarrow |x'x| = 0$ (colinéarité)

H_1 : $|R| \neq 0 \Leftrightarrow |x'x| \neq 0$ (orthogonalité)

K toujours égale à $k + 1$

$$H = - \left(n - 1 - \frac{1}{6} (2K + 5) \right) \ln(1 - |R|)$$

Si $H \leq \chi_{1/2}^2 K(K - 1)$, H_0 acceptée

f. Le test de Belsley - Kuh - Welsch (1980)

Colinéarité si λ_i petites ou = 0

$$(|R| = \prod_i \lambda_i)$$

$$V = \frac{\sqrt{\lambda_{\max}}}{\sqrt{\lambda_{\min}}}$$

X normalisée

$V = 1$ si les colonnes (Variances) sont orthogonales

$V > 1$ s'il y a colinéarité

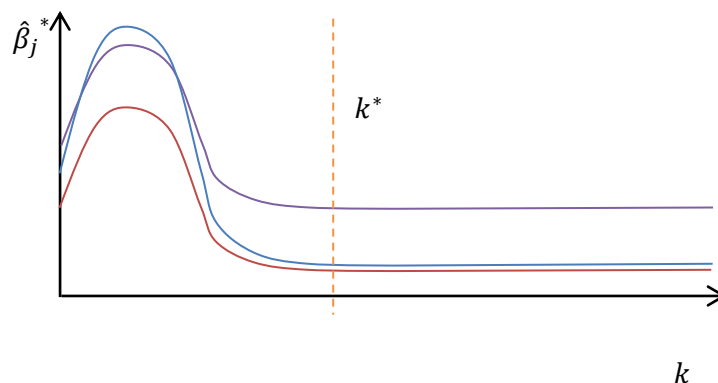
3. Solutions

Ridge Regression (Hoerl, Kennard ; 1962)

$$\hat{\beta}^* = (X'X + kI)^{-1}X'Y$$

$$= [I + k(X'X)^{-1}]^{-1}\hat{\beta}$$

Ridge-Trace



Le but étant de faire varier une valeur k de telle manière que les $\hat{\beta}_j^*$, estimateurs « ridge » se stabilise.

Algorithme d'optimisation

$$k / \min \sum_k e_t^2$$

Ridge-Regression généralisée

$$\hat{\beta}^* = [(X^*)'(X^*) + K]^{-1}(X^*)'Y$$

$$K = ki \geq 0$$

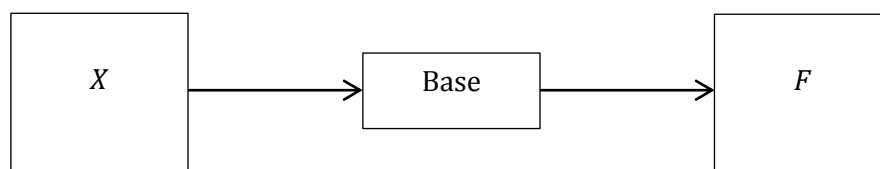
$$X^* = XP' / X'X = P' \wedge P$$

Inverses généralisées de Marquardt

$$\hat{\beta}^* = \tilde{R}^{-1}rX'Y$$

$$\tilde{R}^{-1}r = \sum_{j=1}^n (1/\lambda_j) S_j * S_j'$$

Régression sur composantes principales :



L'ACP est toujours réalisable

Si Objectif : lire l'info dans l'espace donnée par X alors l'ACP réalisable si cette info est lisible à partir des axes de F . F étant constituée de vecteurs colonnes indépendants.

F fonde la méthode de la régression sur composantes principales.

Soit : $Y = X\beta + \varepsilon \rightarrow \text{MCO}$

Par ACP : $Y = Fj + \varepsilon$, Fj orthogonal

Puis par les MCO : $\hat{Y} = F\hat{j}$

Comment passe-t-on de X à F ?

$$\begin{matrix} 1 & X_1 & \dots & X_n & z_j = (x_j - \bar{x}_j)/\sigma x_j \\ \vdots & \left[X \right] & \Rightarrow & \left[Z \right] & \\ n & & & & \end{matrix} \Rightarrow \left[\begin{matrix} \text{Base} \\ \vdots \\ \end{matrix} \right] \Rightarrow \left[F \right]$$

Avec \mathbb{R}^n orthogonal,

$$\begin{matrix} \downarrow & & \mathbb{R}^n \text{ orthogonal} \\ k_1 & \dots & k_n \\ \left[B \right] \end{matrix}$$

Pour un nuage de N points.

$$V[F_1] > \dots > V[F_n]$$

F_j perpendiculaire à $F_{j'}$, $j \neq j'$

$$\max V[F_j], \quad \|b_j\| = 1 \rightarrow \text{Diagonaliser } R$$

Pour B :

- Soit :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

$$\Rightarrow V[F_j] = \lambda_j$$

$$\sum_{j=1,2,3} \lambda_j \simeq 70\% \text{ VT} = n, \quad j = 1,2,3, \text{ alors } \|b_j\| = 1$$

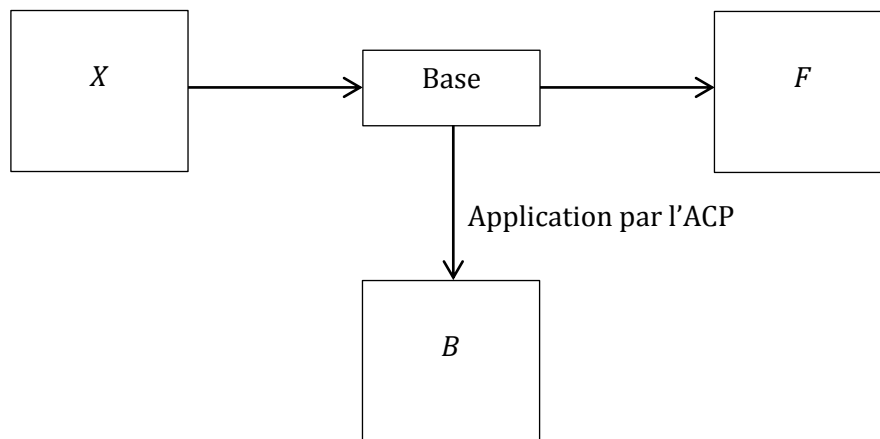
- Soit :

$$B = \begin{bmatrix} \overset{\lambda_1}{\vec{b}_1} & \dots & b_n \\ \vdots & & \vdots \\ \vdots & & \vdots \end{bmatrix}$$

Donc :

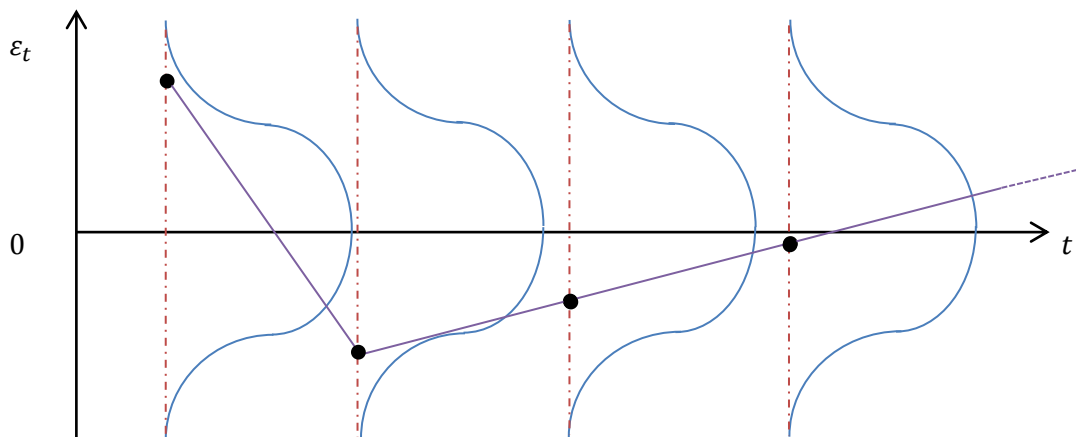
$$F = \begin{bmatrix} F_1 & \dots & F_n \\ \vdots & & \vdots \\ \vdots & & \vdots \end{bmatrix}$$

$$\begin{cases} V[F_1] > \dots > V[F_n] \\ F_j \text{ perpendiculaire à } F_j' \\ \sum_{j=1} V[F_j] \text{ max} \end{cases}$$



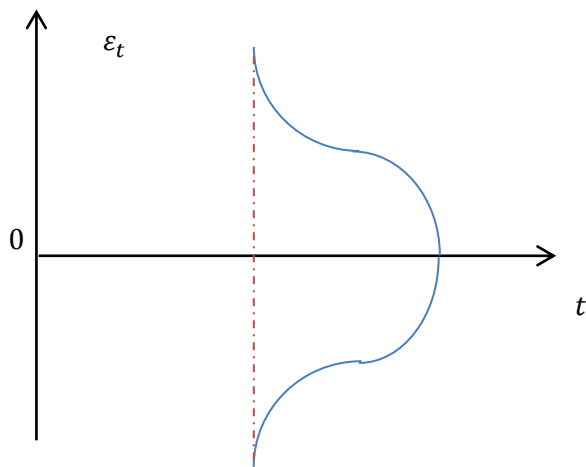
D. Conclusion : Hypothèses de normalité de l'aléa

Soit l'aléa un processus aléatoire, indépendamment et identiquement distribué (IID). Le résidu est considéré comme une trajectoire du processus aléatoire.



Le résidu est une partie du processus aléatoire : il est aussi d'espérance nulle

Si l'aléa est supposé « normal » alors : $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$.



Question : Est-ce que l'histogramme constitué des résidus du modèle suit une loi normale ?

- Test du χ^2

Test de Skewness

→ H_0 : Symétrie normale (*Skewness*)

β_1 (*Skew*) (= coefficient d'asymétrie de Pearson)

$$\beta_1 \equiv N\left(0, \sqrt{\frac{6}{n}}\right)$$

$$\frac{\beta_1 - 0}{\sqrt{\frac{6}{n}}} \equiv N(0,1)$$

RDD :

- Si $\left| \frac{\beta_1}{\sqrt{\frac{6}{n}}} \right| < U_{1-\alpha/2}$, H_0 acceptée au risque p
- Si $\left| \frac{\beta_1}{\sqrt{\frac{6}{n}}} \right| \geq U_{1-\alpha/2}$, H_0 rejetée au risque p

Test de Kurtosis :

H_0 : Aplatissement normal

$$\beta_2 \equiv N(3, \sqrt{24/n})$$

$$\frac{\beta_2 - 3}{\sqrt{24/n}} \equiv N(0,1)$$

RDD :

- Si $\left| \frac{\beta_2 - 3}{\sqrt{24/n}} \right| < U_{1-p/2}$, H_0 accepté au risque p
- Si $\left| \frac{\beta_2 - 3}{\sqrt{24/n}} \right| \geq U_{1-p/2}$, H_0 rejetée au risque p

→ Jarque-Bera

$$JB = n \frac{\beta_1}{6} + n \frac{(\beta_2 - 3)^2}{24} \equiv \chi^2(2)$$

H_0 : Normalité de l'aléa

RDD :

- Si $JB < \chi^2(2)$, H_0 accepté au risque p
- Si $JB \geq \chi^2(2)$, H_0 rejetée au risque p

$$\chi^2(2) = 5,99$$

Fin du cours d'économétrie

Signé par :

(^) (^)
(= ^ - ^ =)
(") (")

POOKIPOOKI

Votre fidèle serviteur ...

2010

2011



Econométrie