

# Économétrie

## Contents

<b>I</b>	<b>Modèle linéaire simple</b>	<b>4</b>
<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Définition . . . . .	4
1.2	Prix Nobels . . . . .	4
<b>2</b>	<b>Histoire de l'économétrie</b>	<b>4</b>
2.1	Des Origines de l'économétrie à l'âge d'or de la modélisation Macro-économique . . .	4
2.1.1	Premières études [17eme 18eme siècle] . . . . .	4
2.1.2	Genèse de l'économétrie [19 <sup>ème</sup> siècle] . . . . .	4
2.1.3	Imbrication de l'économie, des mathématiques et de la statistique [20 <sup>ème</sup> siècle]	5
2.2	De la crise de la modélisation macro-économique à l'âge d'or de l'économétrie . . . .	6
2.2.1	Années 70 et la fin de l'âge d'or de la modélisation marco-économétrique selon la tradition de la Cowles Foundation . . . . .	6
<b>3</b>	<b>Le modèle linéaire simple à deux variables et généralisation a k variables</b>	<b>6</b>
3.1	Les hypothèses de base du modèle . . . . .	7
3.2	Les estimateurs des MCO (moindres carrés ordinaire) . . . . .	8
<b>4</b>	<b>Lois des estimateurs, intervalles de confiance et tests d'hypothèse</b>	<b>9</b>
4.1	Estimation par intervalle de confiance . . . . .	9
4.1.1	Intervalle de confiance de $\beta$ . . . . .	9
4.1.2	Intervalle de confiance de $\hat{\alpha}$ . . . . .	10
4.1.3	Intervalle de confiance de $\sigma_\epsilon^2$ . . . . .	10
4.2	Tests d'hypothèse . . . . .	11
4.2.1	Test de $\beta$ . . . . .	11
4.2.2	Test de $\alpha$ . . . . .	12
4.2.3	Test de $\sigma_\epsilon^2$ . . . . .	13
4.3	Étude de la corrélation . . . . .	13
4.3.1	Propriétés . . . . .	13
4.3.2	Analyse de la Variance . . . . .	14
4.3.3	Calcul du coefficient de détermination $R^2$ . . . . .	15
4.3.4	Test du coefficient de corrélation linéaire r . . . . .	15
4.3.5	Tableau de l'analyse de la variance . . . . .	16
<b>5</b>	<b>Test du coefficient de détermination <math>R^2</math></b>	<b>16</b>

<b>6</b>	<b>Utilisation du modèle de régression en prévision</b>	<b>17</b>
6.1	Intervalle de confiance : de la valeur moyenne de $Y$ connaissant une valeur donnée de $X$ . . . . .	17
6.2	Tests d'hypothèse : comparaison d'une prévision ponctuelle à la droite des moindres carrés . . . . .	18
<b>7</b>	<b>Le modèle linéaire simple à plusieurs variables explicatives</b>	<b>20</b>
7.1	Les estimateurs des moindres carrés ordinaires (MCO) . . . . .	20
7.1.1	Spécifications matricielles du modèle . . . . .	20
7.1.2	Les hypothèses de base du MLGS . . . . .	20
7.1.3	Les estimateurs des moindres carrés ordinaires . . . . .	20
7.1.4	Les propriétés des estimateurs des moindres carrés ordinaires . . . . .	20
7.2	Test des estimateurs . . . . .	20
7.3	L'analyse de la variance et test du coefficient de détermination . . . . .	20
7.4	Utilisation du modèle en prévision . . . . .	20
<b>II</b>	<b>Modèle linéaire général à <math>k</math> paramètres estimés</b>	<b>20</b>
<b>8</b>	<b>Test de normalité</b>	<b>20</b>
8.1	Test de symétrie normale (Skewness) . . . . .	20
8.2	Test d'aplatissement normal (Kurtosis) . . . . .	21
8.3	Test de Jarque-Bera . . . . .	21
<b>9</b>	<b>Le problème de l'autocorrélation des erreurs</b>	<b>22</b>
9.1	Détection de l'autocorrélation . . . . .	22
9.2	Causes de l'autocorrélation . . . . .	22
9.3	Les effets de l'autocorrélation . . . . .	22
9.4	Tests d'autocorrélation . . . . .	23
9.4.1	Test d'autocorrélation d'ordre 1 (Durbin Watson) . . . . .	23
9.4.2	Test d'autocorrélation d'ordre $k$ . . . . .	23
<b>10</b>	<b>Le problème de l'hétéroscédasticité</b>	<b>24</b>
10.1	Définition et conséquences . . . . .	24
10.2	Test de Glejser . . . . .	24
10.3	Test de ARCH . . . . .	25
10.4	Test de Breusch Pagan . . . . .	25
10.5	Test de White . . . . .	25
10.6	Goldfeld - Quandt . . . . .	26
<b>11</b>	<b>La Multi-colinéarité</b>	<b>26</b>
11.1	Définition . . . . .	26
11.2	Les effets de la colinéarité . . . . .	27
11.3	Les tests de colinéarité . . . . .	27
11.3.1	Test d'Haitvosky . . . . .	27
11.3.2	VIF . . . . .	27

11.3.3	Etude des valeurs propres de la matrice des coefficients de corrélation linéaire simple $\tilde{R}$ . . . . .	28
<b>III</b>	<b>Bibliographie</b>	<b>28</b>

## Part I

# Modèle linéaire simple

## 1 Introduction

### 1.1 Définition

L'économétrie est le domaine de l'économie qui s'occupe de l'application de la statistique mathématique et des outils de l'inférence statistique à la mesure empirique des relations postulées par la théorie économique

### 1.2 Prix Nobels

1980, Laurence Klein : Modèles macro-économétriques et leurs applications a l'analyse des fluctuations économiques

1989, Trygve Haavelmo : Approche probabiliste et les modèles a équations simultanées

2000, James Heckman : Travaux sur les théories et méthodes d'analyse des échantillons sélectifs

2000, Mac Fadden : Économétrie des choix discrets

2003, Robert F. Engle : Volatilité des séries temporelles pour les modèles ARCH 2003, Clive Granger : Théorie de la cointégration

## 2 Histoire de l'économétrie

### 2.1 Des Origines de l'économétrie à l'âge d'or de la modélisation Macro-économique

#### 2.1.1 Premières études [17eme 18eme siècle]

L'autorité de la loi naturelle se dégage de celle de la religion et du prince

Apparition des statistiques

#### **Statistique Allemande**

La Statistique est un moyen de classer les savoirs hétéroclites

Conring, Herman

#### **Arithmétique Politique** (Angleterre)

Utilisation de techniques statistiques pour le dépouillement des registres paroissiaux

Petty, William : études natalité/mortalité

King, Gregory : formalisation loi de demande

#### 2.1.2 Genèse de l'économétrie [19<sup>ème</sup> siècle]

La statistique mathématique et l'évolution probabiliste de multiples champs permet de donner une dimension statistique à la représentation de la société

Galton	Statistique mathématique et Analyse mathématique de la régression (corrélation)
Edgeworth	Fonction de densité de la loi normale multivariée. Détermine les expressions de coefficient de régression multiple
Pearson	Coefficient de corrélation multiple : Analyse de la relation entre variables
Yule	Paupérisme, relation avec les mesures d'assistance
Hocker	Utilisation de variables retardées
Lenoir	Première estimation des lois d'offre et demande
Moore	Problème de détermination des salaires, utilisation des corrélations multiples, auto-corrélation et corrélogrammes

### Conjoncturistes Américains :

Juglar, Kitchin, Kondratieff

Analyse des cycles économiques  $\neq$  économie mathématique (Moore)

- Création d'instituts de conjecture (Russie 1920)
- Création du NBER (national bureau of economics research)

### 2.1.3 Imbrication de l'économie, des mathématiques et de la statistique [20<sup>ème</sup> siècle]

A la fin des années 20, l'économétrie gagne de l'intérêt en Europe. Création de la société d'économétrie le 29/12/1930 à Cleveland avec Frish, Timbergen. Elle officialise l'économétrie comme discipline et nomme Fisher président. La société d'économétrie recrute Cowles, qui, en contre-partie, demande la création d'une revue (*Econometrica*, RC : Frish) et la création d'un organisme de recherche Cowles Foundation. Roos est nommé premier président de la Cowles foundation.

- Avant 1930 la théorie des probabilités est pratiquement inutilisée jusqu'à ce que Haavelmo propose d'y apporter une approche stochastique (aléatoire)
- Apparition d'une chaire d'économétrie
- Diversification des sources de financement par l'obtention de subventions (Rockefeller, NBER)
- Le terme d'identification apparaît après les travaux de Haavelmo

Tous ces travaux marquent le début de la modélisation macro-économique qui connaîtra son âge d'or dans les 60's - 70's avec Keynes et les comptes nationaux. Âge d'or de la modélisation macro-économique et essor des modèles dynamiques.

1. Introduction de mécanismes dynamiques dans les modèles économétriques
2. Modèles à retard échelonnés Koyck
3. Développement des prévisions à court terme Box & Jenkins

## 2.2 De la crise de la modélisation macro-économique à l'âge d'or de l'économétrie

### 2.2.1 Années 70 et la fin de l'âge d'or de la modélisation macro-économétrique selon la tradition de la Cowles Foundation

- Premier choc pétrolier → Remise en cause des modélisations
- **Sims** → Modèles VAR (*vector auto regressive*)
- Fonctions de réponse impulsionnelle
- Etude de la causalité
- Test de la racine unitaire (Si les séries sont stationnaires ou non)
- Apparition de la modélisation ARCH, instauration d'une économétrie plurielle (modèle logit, probit)
- Introduction des principes Baillisiens

## 3 Le modèle linéaire simple à deux variables et généralisation à k variables

Modèle économétrique : Représentation simplifiée mais la plus exhaustive possible d'une entité économique donnée sous sa forme la plus courante représentée par un système d'équation (souvent linéaire) et relie des types de variables similaires

Variables explicatives → Variables exogènes  
Variables expliquées → Variables endogènes

$Y = f(X_1, X_2, \dots, X_n)$  une seule variable explicative  $Y$  et une expliquée  $X$

$Y = \alpha + \beta X$  ( $\alpha$  et  $\beta$  sont des paramètres inconnus). Cette relation est exacte, or cela est impossible en économie, on doit donc introduire un terme aléatoire (aléa ou erreur). cette variable a pour rôle de synthétiser l'ensemble des influences sur  $Y$  que  $X$  ne peut expliquer.

### 3 Types de modèles :

- Modèle en série chronologique (évolution au cours du temps)

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

- Modèle en coupe instantanée (à un moment donné dans le temps)

$$Y_i = \alpha + \beta X_i + U_i$$

- Modèle de panel

$$Y_{ti} = \alpha + \beta X_{ti} + U_{it}$$

Le fait d'introduire un aléa permet de faire des tests sur  $\alpha$  et  $\beta$ .

Tout modèle non linéaire peut se ramener à un modèle linéaire (transformation par anamorphose).

### Exemples (transformation)

$Y = \alpha e^{\beta X}$	$Y = \alpha X^\beta$	$Y = \frac{\alpha}{X^\beta}$
$\ln Y = \ln \alpha + \beta X$	$\ln Y = \ln \alpha + \beta X \ln$	$\ln Y = \ln \alpha - \beta \ln X$
$Y' = \alpha' + \beta X$	$Y' = \alpha' + \beta X'$	$Y' = \alpha' - \beta X'$

$$Y_t = \frac{Y_0}{1 + \beta X^t} \Leftrightarrow \frac{Y_0}{Y_t} - 1 = \beta X^t$$

$$\ln\left(\frac{Y_0}{Y_t} - 1\right) = \ln(\beta X^t) \Leftrightarrow \ln Y_0 - \ln Y_t = \ln \beta + t \ln X$$

$$Y' = \beta' + t X'$$

### 3.1 Les hypothèses de base du modèle

Il existe deux méthodes qui permettent d'estimer  $\alpha$  et  $\beta$

- Méthode des MCO (moindres carrés ordinaires)
- Méthode du maximum de vraisemblance

#### Hypothèses

- \*  $X$  est une variable contrôlée (indépendante de l'aléa).  $Cov(X, \varepsilon) = 0$
- \*  $\varepsilon \rightarrow$  une hypothèse de normalité.  $E(\varepsilon_t) = 0$   
En moyenne l'ensemble des facteurs non expliqués par la régression (qui se retrouvent dans l'aléa) tendent à se compenser.
- \* Hypothèse d'homosédasticité :  
la variance de  $\varepsilon_t$  est constante quel que soit le sous échantillon prélevé dans l'intervalle  $\{1; n\}$   
 $E(\varepsilon_t^2) = \sigma_t^2$
- \* Hypothèse de non autocorrélation de l'aléa :  
La distribution de  $\varepsilon_t$  qui correspond à  $X_t$  est indépendante de celle de  $\varepsilon_{t'}$  qui correspond à  $X_{t'}$   
 $Cov(\varepsilon_t, \varepsilon_{t'}) = E(\varepsilon_t \varepsilon_{t'}) = 0 \quad (\forall t \neq t')$

### 3.2 Les estimateurs des MCO (moindres carrés ordinaire)

La méthode des MCO consiste à **minimiser la somme des carrés des écarts**. Écarts entre les valeurs observées de la variable  $Y_t$  et la valeur calculée de cette même variable  $\hat{Y}_t$ . Écart mesuré respectivement par les projections parallèlement à l'axe des ordonnées des points sur la droite de régression.

**Programme de Minimisation :**

$$\min \sum_t e_t^2 = \min \sum_t (Y_t - \hat{Y}_t)^2 = \min \sum_t (Y_t - (\hat{\alpha} + \hat{\beta} X_t))^2$$

Conditions de premier ordre :

$$\begin{array}{l} \frac{\partial \varphi}{\partial \hat{\alpha}} = 0 \\ \Leftrightarrow -2 \sum_t (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0 \\ \Leftrightarrow n\bar{Y} - n\hat{\alpha} - \hat{\beta} n\bar{X} = 0 \\ \Leftrightarrow \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \end{array} \quad \left| \begin{array}{l} \frac{\partial \varphi}{\partial \hat{\beta}} = 0 \\ \Leftrightarrow -2 \sum_t (Y_t - \hat{\alpha} - \hat{\beta} X_t) X_t = 0 \\ \Leftrightarrow \sum_t (Y_t - (\bar{Y} - \hat{\beta} \bar{X} - \hat{\beta} X_t)) X_t = 0 \\ \Leftrightarrow \sum_t X_t Y_t - \bar{Y} \sum_t X_t + \hat{\beta} \bar{X} \sum_t X_t - \hat{\beta} \sum_t X_t^2 = 0 \\ \Leftrightarrow \sum_t X_t Y_t - \bar{Y} n\bar{X} + \hat{\beta} \bar{X} n\bar{X} - \hat{\beta} \sum_t X_t^2 = 0 \\ \Leftrightarrow \sum_t X_t Y_t - n\bar{X}\bar{Y} - \hat{\beta} (\sum_t X_t^2 - n\bar{X}^2) = 0 \\ \Leftrightarrow \hat{\beta} = \frac{\frac{1}{n} \sum_t X_t Y_t - \bar{X}\bar{Y}}{\sum_t X_t^2 - n\bar{X}^2} = \frac{nCov(X, Y)}{nV(X)} \end{array} \right.$$

D'où

$$\begin{aligned} \hat{\beta} &= \frac{\frac{1}{n} \sum_t X_t Y_t - \bar{X}\bar{Y}}{\frac{1}{n} \sum_t X_t^2 - \bar{X}^2} = \frac{Cov(X, Y)}{V(X)} \\ \hat{\alpha} &= \bar{Y} - \hat{\beta} \bar{X} \end{aligned}$$

Conditions de second ordre :

$$\begin{aligned} \frac{\partial^2 \varphi}{\partial \hat{\alpha}^2} &= 2 > 0 \\ \frac{\partial^2 \varphi}{\partial \hat{\beta}^2} &= 2 > 0 \end{aligned}$$

La fonction est convexe, on a donc bien un minimum



La droite de régression passe par le point moyen qui est le centre de gravité (couple  $\bar{X}, \bar{Y}$ ).  
Données centrées :

$$\begin{aligned}x_t &= X_t - \bar{X} & \hat{y}_t &= \hat{Y}_t - \bar{Y} \\y_t &= Y_t - \bar{Y} & \hat{x}_t &= \hat{X}_t - \bar{X}\end{aligned}$$

On cherche  $\hat{y}$  :

$$\hat{Y} = \hat{y}_t + \bar{Y} = \hat{\alpha} + \hat{\beta}X_t = \hat{\alpha} + \hat{\beta}(x_t + \bar{X})$$

$$\begin{aligned}\hat{y}_t &= -\bar{Y} + \hat{\alpha} + \hat{\beta}x_t + \hat{\beta}\bar{X} \\&= -\bar{Y} + \hat{\beta}x_t + \hat{\beta}\bar{X} + (\bar{Y} - \hat{\beta}\bar{X}) \\ \hat{y}_t &= \hat{\beta}x_t\end{aligned}$$

## 4 Lois des estimateurs, intervalles de confiance et tests d'hypothèse

Comme les estimateurs sont linéaires de  $Y_t$  et comme  $Y_t$  dépend de  $\varepsilon_t$  et que  $\varepsilon_t$  est aléatoire et qu'il obéit à une loi normale, alors les estimateurs sont donc aléatoires et obéissent à une loi normale.

$$\hat{\alpha} \sim N\left(\alpha, \sigma_\varepsilon \sqrt{\frac{\sum X_t^2}{n \sum x_t^2}}\right) \quad \hat{\beta} \sim N\left(\beta, \sigma_\varepsilon \sqrt{\frac{1}{\sum_{t=1}^n x_t^2}}\right)$$

$$(n-2) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2(n-2)$$

### 4.1 Estimation par intervalle de confiance

#### 4.1.1 Intervalle de confiance de $\beta$

$$\exists \beta_1, \beta_2 / 1 - \alpha = Prob[\beta_1 < \beta < \beta_2]$$

$$\frac{\frac{\hat{\beta} - \beta}{\sigma_\varepsilon / \sqrt{\sum_{t=1}^n x_t^2}}}{\sqrt{(n-2) \cdot \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \cdot \frac{1}{(n-2)}}} \sim T(n-2) \quad \Leftrightarrow \quad \frac{\hat{\beta} - \beta}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2} \sim T(n-2)$$

d'où :

$$\begin{aligned}
1 - \alpha &= Prob[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}] \\
&= Prob\left[t_{\alpha/2} < \frac{\hat{\beta} - \beta}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2} < t_{1-\alpha/2}\right] \\
&= Prob\left[t_{\alpha/2} \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}} < \hat{\beta} - \beta < t_{1-\alpha/2} \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}}\right] \\
&= Prob\left[\hat{\beta} - t_{1-\alpha/2} \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}} < \beta < \hat{\beta} - t_{\alpha/2} \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}}\right] \\
&= Prob\left[\beta \in \left[\hat{\beta} \pm \begin{Bmatrix} t_{\alpha/2} & \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}} \end{Bmatrix}\right]\right]
\end{aligned}$$

#### 4.1.2 Intervalle de confiance de $\hat{\alpha}$

$$\exists \alpha_1, \alpha_2 / 1 - \alpha = Prob[\alpha_1 < \alpha < \alpha_2]$$

$$\frac{\frac{\hat{\alpha} - \alpha}{\sigma_\varepsilon / \sqrt{\frac{\sum x_t^2}{n \sum x_t^2}}}}{\sqrt{\frac{(n-2)}{n-2} \cdot \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}} \sim T(n-2) \quad \Leftrightarrow \quad \frac{\hat{\alpha} - \alpha}{\hat{\sigma}_\varepsilon} \cdot \frac{\sqrt{n \sum x_t^2}}{\sqrt{\sum X_t^2}} \sim T(n-2)$$

d'où :

$$\begin{aligned}
1 - \alpha &= Prob[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}] \\
&= Prob\left[t_{\alpha/2} < \frac{\hat{\alpha} - \alpha}{\hat{\sigma}_\varepsilon} \cdot \frac{\sqrt{n \sum x_t^2}}{\sqrt{\sum X_t^2}} < t_{1-\alpha/2}\right] \\
&= Prob\left[t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} < \hat{\alpha} - \alpha < t_{1-\alpha/2} \cdot \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}}\right] \\
&= Prob\left[\hat{\alpha} - t_{1-\alpha/2} \cdot \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} < \alpha < \hat{\alpha} - t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}}\right] \\
&= Prob\left[\alpha \in \left[\hat{\alpha} \pm \begin{Bmatrix} t_{\alpha/2} & \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} \end{Bmatrix}\right]\right]
\end{aligned}$$

#### 4.1.3 Intervalle de confiance de $\sigma_\varepsilon^2$

$$\exists \sigma_1^2, \sigma_2^2 / 1 - \alpha = Prob[\sigma_1^2 < \sigma_\varepsilon^2 < \sigma_2^2]$$

$$(n-2) \cdot \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2(n-2)$$

d'où

$$\begin{aligned}
1 - \alpha &= Prob \left[ \chi_{\alpha/2}^2 < \chi^2(n-2) < \chi_{1-\alpha/2}^2 \right] \\
&= Prob \left[ \chi_{\alpha/2}^2 < (n-2) \cdot \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} < \chi_{1-\alpha/2}^2 \right] \\
&= Prob \left[ \frac{\chi_{\alpha/2}^2}{(n-2)\hat{\sigma}_\varepsilon^2} < \frac{1}{\sigma_\varepsilon^2} < \frac{\chi_{1-\alpha/2}^2}{(n-2)\hat{\sigma}_\varepsilon^2} \right] \\
&= Prob \left[ \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\chi_{1-\alpha/2}^2} < \sigma_\varepsilon^2 < \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\chi_{\alpha/2}^2} \right]
\end{aligned}$$

## 4.2 Tests d'hypothèse

### 4.2.1 Test de $\beta$

Spécification du test

$$H_0 : \beta = 0 \quad H_1 : \beta \neq 0$$

Statistique de test

$$\frac{\hat{\beta} - \beta}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2} \sim T(n-2)$$

Intervalle d'acceptation

$$\begin{aligned}
1 - \alpha &= Prob \left[ t_{\alpha/2} < T(n-2) < t_{1-\alpha/2} \right] \\
&= Prob \left[ t_{\alpha/2} < \frac{\hat{\beta}}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2} < t_{1-\alpha/2} \right] \text{ si } H_0 \text{ vrai} \\
&= Prob \left[ t_{\alpha/2} \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}} < \hat{\beta} < t_{1-\alpha/2} \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}} \right] \\
&= Prob \left[ \hat{\beta} \in \left[ \pm t_{\alpha/2} \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum x_t^2}} \right] \right] \text{ IA}
\end{aligned}$$

**RDD**

Si  $\hat{\beta} \in \text{IA}$  on accepte  $H_0$  au risque de première espèce  $\alpha$

Si  $\hat{\beta} \notin \text{IA}$  on rejette  $H_0$  au risque de première espèce  $\alpha$  (validité du modèle)

### Autre façon de procéder

$$\begin{aligned}
1 - \alpha &= Prob \left[ t_{\alpha/2} < \frac{\hat{\beta}}{\hat{\sigma}_\varepsilon / \sqrt{\sum x_t^2}} < t_{1-\alpha/2} \right] = Prob \left[ \left| \frac{\hat{\beta}}{\hat{\sigma}_\varepsilon / \sqrt{\sum x_t^2}} \right| < t_{1-\alpha/2} \right] \\
&= Prob \left[ \left| \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \right| < t_{1-\alpha/2} \right] \\
&= Prob \left[ |t_c| < t_{1-\alpha/2} \right]
\end{aligned}$$

Avec  $t_c = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \sim T(n-2)$

### RDD

Si  $|t_c| < t_{1-\alpha/2}$  on accepte  $H_0$  au risque de  $\alpha$

Si  $|t_c| \geq t_{1-\alpha/2}$  on rejette  $H_0$  au risque de  $\alpha$

### 4.2.2 Test de $\alpha$

#### Spécification du test

$$H_0 : \alpha = 0 \quad H_1 : \alpha \neq 0$$

#### Statistique de test

$$\frac{\hat{\alpha} - \alpha}{\hat{\sigma}_\varepsilon} \cdot \frac{\sqrt{n \sum x_t^2}}{\sqrt{\sum X_t^2}} \sim T(n-2)$$

#### Intervalle d'acceptation

$$\begin{aligned}
1 - \alpha &= Prob \left[ t_{\alpha/2} < T(n-2) < t_{1-\alpha/2} \right] \\
&= Prob \left[ t_{\alpha/2} < \frac{\hat{\alpha}}{\hat{\sigma}_\varepsilon} \cdot \frac{\sqrt{n \sum x_t^2}}{\sqrt{\sum X_t^2}} < t_{1-\alpha/2} \right] \text{ si } H_0 \text{ vrai} \\
&= Prob \left[ t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} < \hat{\alpha} < t_{1-\alpha/2} \cdot \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} \right] \\
&= Prob \left[ \hat{\alpha} \in \left[ \pm t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} \right] \right] \text{ IA}
\end{aligned}$$

### RDD

Si  $\hat{\alpha} \in \text{IA}$  on accepte  $H_0$  au risque de première espèce  $\alpha$

Si  $\hat{\alpha} \notin \text{IA}$  on rejette  $H_0$  au risque de première espèce  $\alpha$

### Autre façon de procéder

$$t_c = \frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}} \sim T(n-2)$$

### RDD

Si  $|t_c| < t_{1-\alpha/2}$  on accepte  $H_0$  au risque de  $\alpha$

Si  $|t_c| \geq t_{1-\alpha/2}$  on rejette  $H_0$  au risque de  $\alpha$

### 4.2.3 Test de $\sigma_\varepsilon^2$

Spécification du test

$$H_0 : \sigma_\varepsilon^2 = \sigma_0^2 \quad H_1 : \sigma_\varepsilon^2 \neq \sigma_0^2$$

Statistique de test

$$(n-2) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_0^2} \sim \chi^2(n-2)$$

Intervalle d'acceptation

$$\begin{aligned} 1 - \alpha &= \text{Prob} \left[ \chi_{\alpha/2}^2 < \chi^2(n-2) < \chi_{1-\alpha/2}^2 \right] \\ &= \text{Prob} \left[ \chi_{\alpha/2}^2 < (n-2) \cdot \frac{\hat{\sigma}_\varepsilon^2}{\sigma_0^2} < \chi_{1-\alpha/2}^2 \right] \text{ si } H_0 \text{ vrai} \\ &= \text{Prob} \left[ \chi_{\alpha/2}^2 \cdot \frac{\sigma_0^2}{n-2} < \sigma_\varepsilon^2 < \chi_{1-\alpha/2}^2 \cdot \frac{\sigma_0^2}{n-2} \right] \end{aligned}$$

**RDD**

Si  $\sigma_\varepsilon^2 \in \text{IA}$  on accepte  $H_0$  au risque de première espèce  $\alpha$

Si  $\sigma_\varepsilon^2 \notin \text{IA}$  on rejette  $H_0$  au risque de première espèce  $\alpha$

## 4.3 Étude de la corrélation

Le **coefficient de corrélation** est un coefficient qui mesure le degré de covariation linéaire, c'est à dire la manière dont varient ensemble les variables entre elles.

$$r_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}}$$

### 4.3.1 Propriétés

- $-1 \leq r \leq 1$
- $r$  est sans dimension
- $r$  est symétrique :  $r_{XY} = r_{YX}$
- $r$  n'est pas affecté par un changement de variable  $r_{XY} = r_{xy}$  avec  $r_{xy} = \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}}$
- Relation entre  $\hat{\beta}$  et  $r$  :  $\hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2}$

$$r = \hat{\beta} \cdot \frac{\sqrt{\sum x_t^2}}{\sqrt{\sum y_t^2}} \Leftrightarrow r = \hat{\beta} \frac{\sqrt{\frac{1}{n} \sum (X_t - \bar{X})^2}}{\sqrt{\frac{1}{n} \sum (Y_t - \bar{Y})^2}} \Leftrightarrow r = \hat{\beta} \cdot \frac{s_X}{s_Y}$$

### 4.3.2 Analyse de la Variance

$$e_t = y_t - \hat{y}_t \quad \text{On sait que : } \hat{y}_t = \hat{\beta}x_t$$

$$\sum_{t=1}^n y_t^2 = \sum_{t=1}^n (e_t + \hat{y}_t)^2 = \sum_{t=1}^n \hat{y}_t^2 + \sum_{t=1}^n e_t^2 + 2 \sum_{t=1}^n e_t \hat{y}_t$$

$$\begin{aligned} \sum_{t=1}^n e_t \hat{y}_t &= \sum_{t=1}^n (y_t - \hat{y}_t) \hat{\beta} x_t \\ &= \hat{\beta} \left[ \sum_{t=1}^n (y_t - \hat{y}_t) x_t \right] \\ &= \hat{\beta} \left[ \sum_{t=1}^n (\hat{y}_t - \hat{\beta} x_t) x_t \right] \\ &= \hat{\beta} \left[ \sum_{t=1}^n y_t x_t - \hat{\beta} \sum_{t=1}^n x_t^2 \right] \\ &= \hat{\beta} \underbrace{\left[ \sum_{t=1}^n y_t x_t - \hat{\beta} \sum_{t=1}^n x_t^2 \right]}_{=0} \end{aligned}$$

D'où  $\boxed{\sum_{t=1}^n y_t^2 = \sum_{t=1}^n \hat{y}_t^2 + \sum_{t=1}^n e_t^2}$  EQ de l'analyse de la variance

$$\Leftrightarrow \underbrace{\sum_{t=1}^n (Y_t - \bar{Y})^2}_{\text{Variance totale}} = \underbrace{\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2}_{\text{Variance expliquée}} + \underbrace{\sum_{t=1}^n e_t^2}_{\text{Variance résiduelle}}$$

La fluctuation totale des valeurs de  $Y_t$  autour de la moyenne de l'échantillon peut être décomposée en deux éléments qui sont :

- **La variance expliquée** : Variation des valeurs de  $\hat{Y}$  autour de la moyenne. Somme des carrés expliquée par l'influence linéaire de  $X$
- **La variance résiduelle** : Variation résiduelle des valeurs de  $Y$  autour de la droite des moindres carrés

#### 4.3.3 Calcul du coefficient de détermination $R^2$

$$\sum_{t=1}^n y_t^2 = \sum_{t=1}^n \hat{y}_t^2 + \sum_{t=1}^n e_t^2 \quad \Leftrightarrow \quad \frac{\sum y_t^2}{\sum y_t^2} = \frac{\sum \hat{y}_t^2}{\sum y_t^2} + \frac{\sum e_t^2}{\sum y_t^2} \quad \Leftrightarrow \quad 1 = \frac{\hat{\beta}^2 \sum x_t^2}{\sum y_t^2} + \frac{\sum e_t^2}{\sum y_t^2}$$

Or :  $r = \hat{\beta} \frac{\sqrt{\sum x_t^2}}{\sqrt{\sum y_t^2}}$  D'où :

$$\Leftrightarrow 1 = r^2 + \frac{\sum e_t^2}{\sum y_t^2}$$

$$\Leftrightarrow r^2 = 1 - \frac{\sum e_t^2}{\sum y_t^2}$$

$$\Leftrightarrow r^2 = 1 - \frac{\text{VR}}{\text{VT}} = \frac{\text{VE}}{\text{VT}}$$

Avec :  $0 \leq r^2 \leq 1$

$$r^2 = \hat{\beta}^2 \frac{\sum x_t^2}{\sum y_t^2} \quad \text{Dans le cas d'une regression lineaire : } r_{xy} = \sqrt{r^2}$$

#### 4.3.4 Test du coefficient de corrélation linéaire $r$

Statistique de test :

$$T(n-2) \sim \frac{\hat{\beta} - \beta}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2}$$

Rappel :

$$\hat{\beta} = r \cdot \frac{\sqrt{\sum y_t^2}}{\sqrt{\sum x_t^2}} \quad \text{et} \quad \hat{\sigma}_\varepsilon^2 = \frac{\sum e_t^2}{n-2}$$

D'où :

$$1 = r^2 + \frac{\sum e_t^2}{\sum y_t^2} \quad \Leftrightarrow \quad \sum_{t=1}^n e_t^2 = (1 - r^2) \sum_{t=1}^n y_t^2$$

Donc :

$$\hat{\sigma}_\varepsilon^2 = \frac{(1 - r^2) \sum y_t^2}{n - 2}$$

Posons  $H_0 : \beta = 0$ , sous  $H_0$  :

$$T(n-2) \sim \frac{\hat{\beta}}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2} = r \cdot \frac{\sqrt{\sum y_t^2}}{\sqrt{\sum x_t^2}} \cdot \frac{\sqrt{\sum x_t^2}}{\sqrt{(1-r^2) \sqrt{\sum y_t^2}}} \cdot \sqrt{(n-2)}$$

$$T(n-2) \sim r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$$

Spécification du test

$$H_0 : \begin{matrix} \rho = 0 \\ \hat{\beta} = 0 \end{matrix} \quad H_1 : \rho \neq 0$$

Statistique de test

$$T(n-2) \sim r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$$

Intervalle d'acceptation

$$\begin{aligned} 1 - \alpha &= Prob[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}] \\ &= Prob\left[t_{\alpha/2} < r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}} < t_{1-\alpha/2}\right] \end{aligned}$$

$$= Prob\left[\left|r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}\right| < t_{1-\alpha/2}\right]$$

RDD

Si  $\left|r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}\right| < t_{1-\alpha/2}$  on accepte  $H_0$  au risque de  $\alpha$

Si  $\left|r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}\right| \geq t_{1-\alpha/2}$  on rejette  $H_0$  au risque de  $\alpha$

#### 4.3.5 Tableau de l'analyse de la variance

Origine des variations	$\sum$ des carrés des écarts	DDL	Carrés moyens
VE	$Q_1 = \sum \hat{y}^2 = \hat{\beta}^2 \sum x_t^2$	1	$Q_1/1 = \sum \hat{y}^2$
VR	$Q_2 = \sum e_t^2$	$n-2$	$Q_2/n-2 = \hat{\sigma}_\varepsilon^2$
VT	$Q_3 = \sum y_t^2$	$n-1$	$\times$

Remarque :  $Q_3 \simeq Q_1 + Q_2$

## 5 Test du coefficient de détermination $R^2$

Spécification du test

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

Statistique de test

$$T(n-2) \sim \frac{r}{\sqrt{1-r^2}}$$

$$F_c = \frac{r^2}{1-r^2}(n-2) \sim F(1, n-2)$$

Intervalle d'acceptation

$$1 - \alpha = Prob\left[\frac{r^2}{1-r^2}(n-2) < F_{1-\alpha}(1, n-2)\right]$$

RDD

Si  $F_c < F_{1-\alpha}(1, n-2)$  On accepte  $H_0$  au risque de  $\alpha$

Si  $F_c \geq F_{1-\alpha}(1, n-2)$  On rejette  $H_0$  au risque de  $\alpha$



## 6 Utilisation du modèle de régression en prévision

### 6.1 Intervalle de confiance : de la valeur moyenne de $Y$ connaissant une valeur donnée de $X$

$X_0$  : valeur donnée de  $X$

$$Y_0 = \alpha + \beta X_0 + \varepsilon_0 \quad \hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t \quad \Rightarrow \quad \hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0$$

On appelle valeur moyenne de  $Y$  connaissant une valeur donnée de  $X$  :

$$E(Y_0/X_0) = \alpha + \beta X_0$$

$\hat{Y}_0$  est un estimateur linéaire sans biais de  $\alpha + \beta X_0$  et de  $E(Y_0/X_0)$

$$E(\hat{Y}_0) = \alpha + \beta X_0 \quad E(\hat{Y}_0) = E(\hat{\alpha}) + E(\hat{\beta} X_0) = \alpha + \beta X_0$$

$$\hat{Y}_0 \sim N \left( \underbrace{E(\hat{Y}_0)}_{\alpha + X_0 \beta}, \underbrace{\sqrt{V(\hat{Y}_0)}}_{?} \right)$$

Calcul de  $V(\hat{Y}_0)$

$$\begin{aligned} V(\hat{Y}_0) &= V(\hat{\alpha} + \hat{\beta} X_0) = V(\hat{\alpha}) + V(\hat{\beta} X_0) + 2 \text{Cov}(\hat{\alpha}, \hat{\beta} X_0) \\ &= V(\hat{\alpha}) + X_0^2 V(\hat{\beta}) + 2 \text{Cov}(\hat{\alpha}, \hat{\beta} X_0) \end{aligned}$$

Avec

$$\begin{aligned} V(\hat{\beta}) &= \frac{\sigma_\varepsilon^2}{\sum x_t^2} \\ V(\hat{\alpha}) &= \sigma_\varepsilon^2 \cdot \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} = \sigma_\varepsilon^2 \cdot \left( \frac{\sum x_t^2 + n \bar{X}^2}{n \sum x_t^2} \right) = \sigma_\varepsilon^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum x_t^2} \right) \\ \text{Cov}(\hat{\alpha}, \hat{\beta}) &= -\frac{\bar{X}}{\sum x_t^2} \cdot \sigma_\varepsilon^2 \end{aligned}$$

d'où

$$\begin{aligned} V(\hat{Y}_0) &= \sigma_\varepsilon^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum x_t^2} \right) + X_0^2 \left( \frac{\sigma_\varepsilon^2}{\sum x_t^2} \right) + 2 X_0 \left( -\frac{\bar{X}}{\sum x_t^2} \cdot \sigma_\varepsilon^2 \right) \\ &= \sigma_\varepsilon^2 \left[ \frac{1}{n} + \frac{\bar{X}^2 + X_0^2 - 2 X_0 \bar{X}}{\sum x_t^2} \right] \\ &= \sigma_\varepsilon^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2} \right] \end{aligned}$$

$$\hat{Y}_0 \sim N \left( \alpha + \beta X_0; \sigma_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}} \right)$$

$\sigma_\varepsilon$  Inconnu, d'où studentisation :

$$(n-2) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2(n-2) \quad \hat{Y}_0 \sim N \left( \alpha + \beta X_0; \sigma_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}} \right)$$

$$t_c = \frac{\frac{\hat{Y}_0 - (\alpha + \beta X_0)}{\sigma_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}}}}{\sqrt{\frac{n-2}{n-2} \cdot \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}} \sim T(n-2)$$

$$t_c = \frac{\hat{Y}_0 - (\alpha + \beta X_0)}{\hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}}} \sim T(n-2)$$

Intervalle de confiance :

$$\begin{aligned} 1 - \alpha &= Prob \left[ t_{\alpha/2} \leq T(n-2) \leq t_{1-\alpha/2} \right] \\ &= Prob \left[ t_{\alpha/2} \leq \frac{\hat{Y}_0 - (\alpha + \beta X_0)}{\hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}}} \leq t_{1-\alpha/2} \right] \\ &= Prob \left[ \hat{Y}_0 - t_{1-\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}} \leq \alpha + \beta X_0 \leq \hat{Y}_0 - t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}} \right] \\ &= Prob \left( \begin{pmatrix} \alpha + \beta X_0 \\ E(Y_0/X_0) \\ Y_0 \end{pmatrix} \in \left[ \hat{Y}_0 \pm t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}} \right] \right) \end{aligned}$$

## 6.2 Tests d'hypothèse : comparaison d'une prévision ponctuelle à la droite des moindres carrés

Calcul de la statistique de test :

- Calcul de  $\sigma_{\hat{Y}_0}^2$

$$\begin{aligned} Y_0 &= \alpha + \beta X_0 \\ \hat{Y}_0 &= \hat{\alpha} + \hat{\beta} X_0 \\ Y_0 - \hat{Y}_0 &= \varepsilon_0 - (\hat{\alpha} - \alpha) - (\hat{\beta} - \beta) X_0 = Z_0 \end{aligned}$$

- Détermination de la loi de  $Z_0$

$$\hat{Y}_0 \sim N \Rightarrow Z_0 \sim N(E(Z_0), \sigma_{Z_0})$$

– Espérance de  $Z_0$  :

$$E(Z_0) = E(Y_0 - \hat{Y}_0) = E(\varepsilon) - \underbrace{E(\hat{\alpha} - \alpha)}_{=0} - \underbrace{E[(\hat{\beta} - \beta)X_0]}_{=0} = \underbrace{E(\varepsilon)}_{\text{Hypothèse de normalité}} = 0$$

$\hat{Y}_0$  est un ESB de  $Y_0$

– Variance de  $Z_0$

$$\begin{aligned}
V(Z_0) &= V[(\alpha + \beta X_0) - \hat{Y}_0] \\
&= V(\alpha + \beta X_0) + V(\hat{Y}) - \underbrace{\text{Cov}(\alpha + \beta X_0, \hat{Y}_0)}_{\substack{=0 \\ \text{Indépendance}}} \\
&= V(\varepsilon_0) + V(\hat{Y}_0) \\
&= \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \left( \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2} \right) \\
V(Z_0) &= \sigma_\varepsilon^2 \left( 1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2} \right)
\end{aligned}$$

**Spécification du test**

$$H_0 : \begin{cases} E(Y_0/X_0) \\ \alpha + \beta X_0 \\ Y_0 \end{cases} = \text{Constante} \quad H_1 : E(Y_0/X_0) \neq \text{Constante}$$

**Statistique de test**

$$t_c = \frac{Y_0 - \hat{Y}_0}{\hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}}} \sim T(n-2)$$

**Intervalle d'acceptation**

$$\begin{aligned}
1 - \alpha &= \text{Prob} [t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}] \\
&= \text{Prob} \left[ t_{\alpha/2} < \frac{Y_0 - \hat{Y}_0}{\hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}}} < t_{1-\alpha/2} \right] \\
&= \text{Prob} \left[ \hat{Y}_0 \in \left( \begin{pmatrix} \alpha + \beta X_0 \\ E(Y_0/X_0) \\ Y_0 \end{pmatrix} \pm \begin{pmatrix} t_{\alpha/2} \\ t_{1-\alpha/2} \end{pmatrix} \hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_t^2}} \right) \right]
\end{aligned}$$

**RDD**

Si  $\hat{Y}_0 \in \text{IA}$  on accepte  $H_0$  au risque de  $\alpha$

Si  $\hat{Y}_0 \notin \text{IA}$  on rejette  $H_0$  au risque de  $\alpha$

## 7 Le modèle linéaire simple à plusieurs variables explicatives

### 7.1 Les estimateurs des moindres carrés ordinaires (MCO)

#### 7.1.1 Spécifications matricielles du modèle

#### 7.1.2 Les hypothèses de base du MLGS

#### 7.1.3 Les estimateurs des moindres carrés ordinaires

#### 7.1.4 Les propriétés des estimateurs des moindres carrés ordinaires

### 7.2 Test des estimateurs

### 7.3 L'analyse de la variance et test du coefficient de détermination

### 7.4 Utilisation du modèle en prévision

## Part II

# Modèle linéaire général à $k$ paramètres estimés

Ici nous supposons que les hypothèses de base sur l'aléa ne sont plus vérifiées, à savoir : normalité, auto-corrélation, homoscedasticité, hétéroscedasticité.

## 8 Test de normalité

Il existe deux tests de normalité :

- Test de Skewness (symétrie)
- Kurtosis (aplatissement)

La normalité concerne l'aléa, comme on ne connaît pas l'aléa, on fait les tests sur les résidus.

### 8.1 Test de symétrie normale (Skewness)

**Spécification du test**

$H_0$  : symétrie normale (Skewness)

**Statistique de test**

$\beta_1^{1/2}$  : Coefficient de symétrie de Pearson (Skewness)

$\mu_k$  : Moment centré d'ordre  $k$

$$\beta_1^{1/2} = \frac{\mu_3}{\mu_2^{3/2}} \sim N\left(0, \sqrt{\frac{6}{n}}\right)$$

d'où

$$v_1 = \frac{\beta_1^{1/2} - 0}{\sqrt{\frac{6}{n}}} \sim N(0, 1)$$

**Intervalle d'acceptation**

$$1 - \alpha = Prob \left[ u_{\alpha/2} < \frac{\beta_1^{1/2} - 0}{\sqrt{\frac{6}{n}}} < u_{1-\alpha/2} \right]$$

**RDD**

Si  $\left| \frac{\beta_1^{1/2} - 0}{\sqrt{\frac{6}{n}}} \right| < u_{1-\alpha/2}$  On accepte  $H_0$  au risque de  $\alpha$ . (Il y a symétrie normale)  
Si  $\left| \frac{\beta_1^{1/2} - 0}{\sqrt{\frac{6}{n}}} \right| \geq u_{1-\alpha/2}$  On rejette  $H_0$  au risque de  $\alpha$

## 8.2 Test d'aplatissement normal (Kurtosis)

**Spécification du test**

$H_0$  : Aplatissement normal

**Statistique de test**  $\beta_2$  : Coefficient de Kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \sim N \left( 3, \sqrt{\frac{24}{n}} \right)$$

d'où

$$v_2 = \frac{\beta_2 - 3}{\sqrt{\frac{24}{n}}} \sim N(0, 1)$$

**Intervalle d'acceptation**

$$1 - \alpha = Prob \left[ u_{\alpha/2} < \frac{\beta_2 - 3}{\sqrt{\frac{24}{n}}} < u_{1-\alpha/2} \right]$$

**RDD**

Si  $\left| \frac{\beta_2 - 3}{\sqrt{\frac{24}{n}}} \right| < u_{1-\alpha/2}$  On accepte  $H_0$  au risque de  $\alpha$ . (Il y a aplatissement normale)  
Si  $\left| \frac{\beta_2 - 3}{\sqrt{\frac{24}{n}}} \right| \geq u_{1-\alpha/2}$  On rejette  $H_0$  au risque de  $\alpha$

Si les deux hypothèses sont validées (symétrie et aplatissement normal) alors il y a normalité de l'aléa

## 8.3 Test de Jarque-Bera

**Spécification du test**

$H_0$  : Normalité

### Statistique de test

$$JB = \frac{n}{6}\beta_1 + \frac{n}{24}(\beta_2 - 3)^2 \sim \chi^2(2)$$

### Intervalle d'acceptation

$$1 - \alpha = Prob \left[ \frac{n}{6}\beta_1 + \frac{n}{24}(\beta_2 - 3)^2 < \chi_{1-\alpha}^2(2) \right]$$

### RDD

Si  $JB < \chi_{1-\alpha}^2(2)$  On accepte  $H_0$  au risque de  $\alpha$  (Il y a normalité des résidus)

Si  $JB \geq \chi_{1-\alpha}^2(2)$  On rejette  $H_0$  au risque de  $\alpha$

## 9 Le problème de l'autocorrélation des erreurs

### 9.1 Détection de l'autocorrélation

Comme l'aléa ( $\varepsilon_t$ ) est inconnu, le test se fait sur les résidus  $e_t = Y_t - \hat{Y}_t$ . **Autocorrélation** : On analyse la corrélation à l'intérieur de la distribution des résidus. Il y aura autocorrélation toutes les fois où l'on peut trouver un coefficient de corrélation linéaire significativement différent de 0 entre la chronique des résidus et elle-même décalée d'un ou de plusieurs pas de temps.

Dessin coef autocorrelation\*

On va appeler  $r_k$  le coefficient d'autocorrélation à l'ordre  $k$  (on mesure la corrélation entre la chronique et  $e_t$ )

**Corrélogramme** : représentation graphique de l'ensemble des coefficients de corrélation graph FAC

Il existe 2 types d'autocorrélation.

graph AC

### 9.2 Causes de l'autocorrélation

1. Le modèle ignore une variable exogène.
2. Les variables de départ ont mal été désaisonnalisées (donnée de départ saisonnière)
3. Les variables de départ possédaient des non informations et corrigés par extrapolation linéaire)
4. Les valeurs de départ contiennent des événements exceptionnels mal expliqués par le modèle. Cela peut signifier qu'il y a eu un manque de variable dichotomique ( $= \{0; 1\}$ )
5. Les variables de départ ne sont pas stationnaire, elles peuvent contenir

Graph

### 9.3 Les effets de l'autocorrélation

- Les estimateurs sont sans biais
- Les variances d'échantillon des coefficients de régression ne sont plus minimales (les estimateurs ne sont plus efficaces)

- On ne peut pas utiliser le modèle pour faire la prévision, la méthode des MCO n'est plus la meilleure des méthodes pour estimer le modèle (remise en cause du modèle)

## 9.4 Tests d'autocorrélation

### 9.4.1 Test d'autocorrélation d'ordre 1 (Durbin Watson)

Le test de Durbin Watson permet de mettre en évidence l'autocorrélation à l'ordre 1.

Il y a processus d'AR(1) lorsque :

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

$\eta_t \sim$  Bruit Blanc : Hypothèse de normalité, homosédasticité, non autocorrélation.

**Spécification du test :**

$$H_0 : \rho = 0 \quad H_1 : \rho \neq 0$$

$\rho = 0 \Rightarrow$  Absence d'autocorrélation car :  $\varepsilon_t = 0\varepsilon_{t-1}$

**Statistique de test :**

$$DW = \frac{\sum (e_{t+1} - e_t)^2}{\sum e_t^2}$$

**RDD**

$0; d_1$	$d_1; d_2$	2	$4 - d_2; 4 - d_1$	$4 - d_1; 4$
autocorrélation positive d'ordre 1	Zone de doute ←	Absence d'autocorrélation (H0 vrai)	Zone de doute →	autocorrélation négative d'ordre 1

Remarque :

$$DW \xrightarrow{n \rightarrow \infty} 2(1 - \bar{\rho})$$

Où  $\hat{\rho}$  est un estimateur de  $\rho$  tel que :  $\hat{\rho} = \frac{\sum (e_{t+1} - e_t)^2}{\sum e_t^2}$  avec  $|\rho| < 1$

### 9.4.2 Test d'autocorrélation d'ordre k

Statistique de Ljung-Box

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_k \varepsilon_{t-k} + \eta_t$$

**Spécification du test :**

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k \quad H_1 : \text{au moins un coef} \neq 0$$

**Statistique de test :**

$$Q_{\text{stat}}(K) = n(n+2) \sum_{k=1}^k \frac{r_k^2}{n-k} \sim \chi_{1-\alpha}^2(K)$$

Statistique à l'ordre 3 :

$$\begin{aligned} Q_{\text{stat}}(3) &= n(n+2) \sum_{k=1}^3 \frac{\hat{r}_k^2}{n-k} \\ &= n(n+2) \left( \frac{r_1^2}{n-1} + \frac{r_2^2}{n-2} + \frac{r_3^2}{n-3} \right) \end{aligned}$$

**RDD :** Si  $Q_{\text{stat}} < \chi_{1-\alpha}^2(K)$  On accepte  $H_0$  au risque de  $\alpha$  (absence d'AR à l'ordre k)  
 Si  $Q_{\text{stat}} \geq \chi_{1-\alpha}^2(K)$  On rejette  $H_0$  au risque de  $\alpha$  (AR à l'ordre k)

## 10 Le problème de l'hétéroscédasticité

### 10.1 Définition et conséquences

Il y a hétéroscédasticité lorsque les termes qui se trouvent sur la diagonale de la matrice covariance-variance sont différents entre eux tel que :  $E(\varepsilon_1^2) = \sigma_1^2 \neq E(\varepsilon_2^2) = \sigma_2^2 \neq E(\varepsilon_n^2) = \sigma_n^2$   
 Les conséquences de l'hétéroscédasticité sont les mêmes que celles de l'autocorrélation (l'hétéroscédasticité étant un cas particulier de l'autocorrélation).

### 10.2 Test de Glejser

Le test de Glejser consiste à regresser la valeur absolue des résidus (obtenus lors de la regression) avec chacune des variables explicatives (Il y a donc autant de regressions qu'il y a de variables explicatives).

$$\begin{aligned} |e_t| &= a_0 + a_1 + X_{kt} \\ |e_t| &= a_0 + \frac{a_1}{X_{kt}} \\ |e_t| &= a_0 + a_1 \sqrt{X_{kt}} \\ |e_t| &= a_0 + \frac{a_1}{\sqrt{X_{kt}}} \end{aligned}$$

On teste par la suite la signification de  $a_1$ .

**Spécification du test :**

$$H_0 : a_1 = 0 \quad H_1 : a_1 \neq 0$$

**Statistique de test :**

$$t_c = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} \sim T(n-2)$$

**RDD :** Si  $|t_c| < t_{1-\alpha/2}$  On accepte  $H_0$  au risque de  $\alpha$  (homoscédasticité)  
 Si  $|t_c| \geq t_{1-\alpha/2}$  On rejette  $H_0$  au risque de  $\alpha$  (hétéroscédasticité)



### 10.3 Test de ARCH

Processus autorégressif sur le carré de la variable :

$$\varepsilon_t^2 = \varphi_1 \varepsilon_{t-1}^2 + \varphi_2 \varepsilon_{t-2}^2 + \dots + \varphi_p \varepsilon_{t-p}^2 + \eta_t$$

**Spécification du test**

$$H_0 : \varphi_1 = \varphi_2 = \dots = \varphi_p = 0 \quad H_1 : \text{au moins un coef} \neq 0 \quad (1)$$

**Statistique de test :** Multiplicateur de Lagrange

$$LM = nR^2 \sim \chi_{1-\alpha}^2(p)$$

Où  $p$  est le nombre de retards.

**RDD :**

Si  $nR^2 < \chi_{1-\alpha}^2(p)$  On accepte  $H_0$  au risque de  $\alpha$  (homoscédasticité)

Si  $nR^2 \geq \chi_{1-\alpha}^2(p)$  On rejette  $H_0$  au risque de  $\alpha$  (hétéroscédasticité)

NOTES 08/11/21

### 10.4 Test de Breusch Pagan

Si il y a hétéroscédasticité, il y a une relation entre l'aléa et les variables explicatives.

$$\varepsilon_t = \delta_1 + \delta_2 X_{2t} + \dots + \delta_k X_{kt} + \nu_t$$

**Spécification du test :**

$$H_0 : \delta_1 = \delta_2 = \dots = \delta_k = 0 \quad H_1 : \text{au moins un coef} \neq 0$$

Deux façons de procéder :

**Statistique de test :** Fischer

$$F_c = \frac{R^2}{1-R^2} \cdot \frac{n-K}{K-1} \sim F_{1-\alpha}(K-1, n-K)$$

**RDD :**

Si  $F_c < F_{1-\alpha}(K-1, n-K)$  On accepte  $H_0$

Si  $F_c \geq F_{1-\alpha}(K-1, n-K)$  On rejette  $H_0$

**Statistique de test :** Lagrange

$$LM = nR^2 \sim \chi_{1-\alpha}^2(K-1)$$

**RDD :**

Si  $nR^2 < \chi_{1-\alpha}^2(K-1)$  On accepte  $H_0$

Si  $nR^2 \geq \chi_{1-\alpha}^2(K-1)$  On rejette  $H_0$

Où  $K$  est le nombre de paramètres estimés.

### 10.5 Test de White

Le test de White est fondé sur une relation entre le carré des résidus et une ou plusieurs variables explicatives au carré au sein d'une même régression.

On peut réaliser le test de white sans termes croisés et avec termes croisés. Pour le premier :

$$\varepsilon_t^2 = a_0 + a_1 x_{1t} + b_1 x_{1t}^2 + a_2 x_{2t} + b_2 x_{2t}^2 + \dots + a_k x_{kt} + b_k x_{kt}^2 + \eta_t$$

**Spécification du test :**

$$H_0 : a_1 = b_1 = a_2 = b_2 = \dots = a_k = b_k \quad H_1 : \text{au moins un coef} \neq 0$$

**Statistique de test : Fischer**

$$F_c = \frac{R^2}{1-R^2} \cdot \frac{n-K}{K-1} \sim F_{1-\alpha}(K-1, n-K)$$

**RDD :**

Si  $F_c < F_{1-\alpha}(K-1, n-K)$  On accepte  $H_0$

Si  $F_c \geq F_{1-\alpha}(K-1, n-K)$  On rejette  $H_0$

**Statistique de test : Lagrange**

$$LM = nR^2 \sim \chi^2_{1-\alpha}(K-1)$$

**RDD :**

Si  $nR^2 < \chi^2_{1-\alpha}(K-1)$  On accepte  $H_0$

Si  $nR^2 \geq \chi^2_{1-\alpha}(K-1)$  On rejette  $H_0$

Où  $K$  est le nombre de paramètres estimés.

Avec termes croisés :

$$\varepsilon_t^2 = \lambda_1 + \lambda_2 X_{2t} + \dots + \lambda_k X_{kt} + a_2 X_{2t}^2 + a_k X_{kt}^2 + \mu_t + b_2 X_{2t} X_{3t} + b_3 X_{2t} X_{4t} + \eta_t$$

La manière de procéder est la même que en terme non croisés.

## 10.6 Goldfeld - Quandt

Ce test se construit toutes les fois où l'écart type de l'erreur du modèle s'accroît proportionnellement avec l'une des variables explicative

$$\sqrt{E(\varepsilon_t^2)} = aX_{kt} + \eta_t$$

On ordonne les observations des variables  $Y$  et  $X_{kt}$  en fonction des valeurs croissantes de la variable qui a une relation avec l'écart type de l'alea. On supprime les observations centrales de l'échantillon obtenu et on appelle  $m$  le nombre de ces observations supprimées.

$$n = 30$$

$$m = 8$$

$$n = 60$$

$$m = 16$$

On obtient deux sous échantillons, un qui correspond aux faibles valeurs des variables explicatives, et l'autre qui correspond aux fortes de celle ci, On régresse en appliquant les MCO sur chacun de ces sous échantillons et l'on calcule les SCR de ces deux échantillons  $SCR_1$  et  $SCR_2$ .

**Spécification du test :**

$H_0$  : Homoscédasticité

$H_1$  : Hétéroscédasticité

**Statistique de test :**

$$F_c = \frac{SCR_2}{SCR_1} \sim F_{1-\alpha} \left( \frac{n-m}{2} - k; \frac{n-m}{2} - k \right)$$

**RDD :**

Si  $F_c < F_{1-\alpha}$  On accepte  $H_0$  au risque de  $\alpha$  (homosédasticité)

Si  $F_c \geq F_{1-\alpha}$  On rejette  $H_0$  au risque de  $\alpha$  (heterosedasticité)

## 11 La Multi-colinéarité

### 11.1 Definition

Dans le MLS on a supposé que les variables  $X_k$  étaient linéairement indépendantes, ce que l'on a traduit par le fait que le rang de  $X$  est égal à  $k$ .

On dira qu'il y a colinéarité parfaite entre deux variables explicatives si les deux vecteurs d'observation sont identiques, ce que l'on traduit par le coefficient de corrélation linéaire entre les deux égal à 1.

On dira qu'il y a multicollinéarité parfaite entre plusieurs variables (explicatives) si il existe une combinaison linéaire des observations de plusieurs variables qui donnent le vecteur des observations d'une autre variable. Le coefficient de détermination (multiple)  $R^2$  entre ces variables doit être égal à 1.

Dans le cas de colinéarité parfaite il est nécessaire de modifier le modèle de départ en supprimant la variable qui est à l'origine de la colinéarité.

Dans la pratique la colinéarité parfaite n'existe quasiment pas, par contre il y a une tendance vers la colinéarité, c'est à dire un  $R^2 \rightarrow 1$ .

## 11.2 Les effets de la colinéarité

1. Les estimateurs des MC tendent à être très important en valeur absolue
2. Les variances et les covariances tendent à être importantes

## 11.3 Les tests de colinéarité

### 11.3.1 Test d'Haitvosky

Ce test repose sur l'observation de la matrice des coefficients de corrélation linéaire simple **entre les variables explicatives**. Cette matrice permet de détecter des colinéarités simples mais pas des multi-colinéarités.

$\tilde{R}$  matrice des coefficients de corrélation linéaire simple entre les variables explicatives.

Si un élément de cette matrice tend vers 1, cela veut dire qu'il existe une colinéarité entre les deux variables exogènes utilisées dans le calcul de ce coefficient.:

**Spécification du test :**

$H_0 : |\tilde{R}| = 0$

$H_1 : |\tilde{R}| \neq 0$

**Statistique de test :**

$$H = - \left( n - 1 - \frac{1}{6}(2k + 5) \right) \ln \left( 1 - |\tilde{R}| \right) \sim \chi^2_{1-\alpha} \left( \frac{1}{2}k(k-2) \right)$$

**RDD :**

Si  $H < \chi^2_{1-\alpha} \left( \frac{1}{2}k(k-2) \right)$  On accepte  $H_0$  au risque de  $\alpha$

Si  $H \geq \chi^2_{1-\alpha} \left( \frac{1}{2}k(k-2) \right)$  On rejette  $H_0$  au risque de  $\alpha$

### 11.3.2 VIF

Facteur d'inflation de la variance

$$VIF_i = \frac{1}{1 - R_i^2}$$

$R_i$  : coefficient de détermination multiple entre une variable  $X_i$  et les autres

$$\begin{aligned} \text{VIF}_1 &= \frac{1}{1 - R_1^2} & R_1^2 &\rightarrow X_2 \text{ et } X_3, X_4 \\ \text{VIF}_2 &= \frac{1}{1 - R_2^2} & R_2^2 &\rightarrow X_3 \text{ et } X_2, X_4 \\ \text{VIF}_3 &= \frac{1}{1 - R_3^2} & R_3^2 &\rightarrow X_4 \text{ et } X_2, X_3 \end{aligned}$$

$\text{VIF} \rightarrow +\infty$  Cela veut dire qu'il y a de la colinéarité entre  $X_i$  et les autres variables.

### 11.3.3 Etude des valeurs propres de la matrice des coefficients de corrélation linéaire simple $\tilde{R}$

## Part III

# Bibliographie

- Bourbonnais - Econométrie - Dunod
- Johnston Dinardo - Econométrie
- Greene - Econométrie - Pearson