A decorative graphic on the right side of the page. It features three concentric circles in shades of blue and teal, arranged vertically. The top circle is the largest and darkest blue. The middle circle is smaller and a lighter blue. The bottom circle is the largest and a teal color. Two thin blue lines cross the page diagonally, one from the top-left to the middle-right, and another from the top-right to the bottom-right, framing the circles.

ECONOMETRIE

Partie I

Synthèse

Bokich Sara

Table des matières

CHAPITRE 1: VARIABLE ALÉATOIRE ET ESTIMATEUR	1
A. VARIABLE ALÉATOIRE	1
1. QU'EST CE QU'UNE VARIABLE ALÉATOIRE X ?	1
2. EXEMPLE DE DISTRIBUTION DE PROBABILITÉ: LA SOMME DE DEUX DÉES.....	1
3. ESPÉRANCE D'UNE VARIABLE ALÉATOIRE	2
a) Définition de $E(X)$, l'espérance de X	2
b) Notation alternative de $E(X)$	2
4. ESPÉRANCE D'UNE FONCTION D'UNE VARIABLE ALÉATOIRE.....	3
a) Définition de $Eg(X)$, l'espérance d'une fonction de X	3
5. PROPRIÉTÉ DE L'ESPÉRANCE.....	3
6. VARIANCE DE LA POPULATION DANS LE CAS D'UNE VARIABLE ALÉATOIRE DISCRÈTE.....	4
a) Définition de la variance de la population de X : $EX - \mu^2$	4
7. INDÉPENDANCE DE DEUX VARIABLES ALÉATOIRES	5
8. VARIABLES ALÉATOIRES CONTINUES.....	5
9. COMPOSANTES FIXES ET ALÉATOIRES D'UNE VARIABLE ALÉATOIRE.....	5
B. ESTIMATEURS	6
1. ESTIMATEURS.....	6
a) Définition estimateur et estimation.....	6
b) Les estimateurs sont des variables aléatoires.....	6
2. ESTIMATEUR NON BIAISÉ ET EFFICACITÉ.....	7
a) Estimateur non biaisé X	7
b) Efficacité	7
3. EFFET DE L'AUGMENTATION DE LA TAILLE D'ÉCHANTILLONS SUR LA DISTRIBUTION X	7
a) Convergence.....	7
4. ESTIMATEUR BIAISÉ EN ÉCHANTILLON FINI MAIS CONVERGEANT.....	8
CHAPITRE 2: COVARIANCE, MODÈLE DE RÉGRESSION SIMPLE ET TEST D'HYPOTHÈSE	9
A. COVARIANCE	9
1. COVARIANCE D'ÉCHANTILLON	9
a) Définition de la covariance d'échantillon	9
2. RÈGLES DE LA COVARIANCE	9
3. VARIANCE D'ÉCHANTILLON ET RÈGLES DE LA VARIANCE	10
a) Définition de la variance d'un échantillon	10
b) Propriétés.....	10
4. COEFFICIENT DE CORRÉLATION D'UNE POPULATION	11
a) Coefficient de corrélation d'un échantillon	11
B. MODÈLE DE RÉGRESSION SIMPLE	12
1. MODÈLE DE RÉGRESSION SIMPLE.....	12
a) Critère des Moindres Carrés: (Residual Sum of Squares: RSS).....	13
2. DÉDUIRE LES COEFFICIENTS DE RÉGRESSIONS LINÉAIRES	14
3. INTERPRÉTATION D'UNE ÉQUATION DE RÉGRESSION	18

4.	VALIDITÉ DE L'ESTIMATION	19
a)	Mesures de validité de l'estimation dans les analyses de régression.....	19
b)	Validité de l'estimation.....	20
5.	LES COEFFICIENTS DE RÉGRESSION: DES VARIABLES ALÉATOIRES DISTRIBUÉES COMME UNE NORMALE.	21
6.	PRÉCISION DES COEFFICIENTS DE RÉGRESSION	23
C.	TEST D'HYPOTHÈSE	24
1.	TEST D'HYPOTHÈSE D'UN COEFFICIENT DE RÉGRESSION	24
2.	TEST T D'UNE HYPOTHÈSE PORTANT SUR LE COEFFICIENT DE RÉGRESSION	30
3.	INTERVALLES DE CONFIANCE.....	31
4.	TEST F DE QUALITÉ DE L'ESTIMATION.....	35

CHAPITRE 3: RÉGRESSION MULTIPLE AVEC DEUX VARIABLES EXPLICATIVES – MULTICOLINÉARITÉ – TEST F DE QUALITÉ D'AJUSTEMENT..... 38

A.	RÉGRESSION MULTIPLE AVEC DEUX VARIABLES EXPLICATIVES	38
1.	RÉGRESSION MULTIPLE AVEC DEUX VARIABLES EXPLICATIVES: EXEMPLE	38
2.	METTRE 1 RELATION SOUS FORME DE GRAPHE DANS 1 MODÈLE DE RÉGRESSION MULTIPLE.....	40
B.	MULTICOLINÉARITÉ	41
1.	MULTICOLINÉARITÉ.....	41
a)	Mesures possibles pour réduire la multicollinearité	43
C.	F-TESTS DE QUALITÉ D'AJUSTEMENT	47
1.	QUALITÉ D'AJUSTEMENT DE L'ÉQUATION ENTIÈRE.....	47
2.	TEST DE POUVOIR EXPLICATIF JOINT D'UN GROUPE DE VARIABLES QUAND ELLES SONT AJOUTÉES À UN MODÈLE DE RÉGRESSION	48
3.	F-TEST MARGINAUX ÉQUIVALENTS AU T-TESTS POUR UNE SEULE VARIABLE	49

CHAPITRE 4: NON LINÉARITÉ ET VARIABLES DUMMY (DICHOTOMIQUES)..... 50

A.	LINÉARITÉ ET NON LINÉARITÉ.....	50
1.	DÉFINITION.....	50
2.	ÉLASTICITÉS ET MODÈLES LOGARITHMIQUES DOUBLES.....	52
a)	Modèle Log-Log	52
b)	Modèle Log-Lin (semi-logarithmiques)	53
c)	Modèle Lin-Log (semi-logarithmiques).....	53
d)	Variable Dummy dans les modèles Log-Lin	53
3.	TERME D'ERREUR DANS LES MODÈLES NON LINÉAIRES.....	54
B.	VARIABLES DICHOTOMIQUES (OU DUMMY).....	55
1.	CLASSIFICATION DE VARIABLES BINAIRES AVEC DEUX CATÉGORIES.....	55
2.	CLASSIFICATION DES VARIABLES BINAIRES AVEC PLUS DE DEUX CATÉGORIES	56
3.	TEST DE F	58

CHAPITRE 5: MAUVAISE SPÉCIFICATION DE VARIABLES, VARIABLES PROXY ET TESTER UNE RÉGRESSION LINÉAIRE..... 59

A.	HYPOTHÈSE DE GAUSS –MARKOV (BONNE SPÉCIFICATION)	59
1.	MAUVAISE SPÉCIFICATION DE VARIABLES	59
a)	Mauvaise spécification de variables I: omission d'une variable pertinente	59

b)	Mauvaise spécification de variables II: inclusion d'une variable non pertinente	61
B.	VARIABLES PROXY	62
1.	VARIABLES PROXY	62
C.	TESTER UNE RESTRICTION LINÉAIRE	63

CHAPITRE 6: HETEROSCEDASTICITÉ..... 64

A.	HOMOSCEDASTICITÉ ET HETEROSCEDASTICITÉ	64
1.	HOMOSCEDASTICITÉ	64
2.	HÉTÉROSCÉDASTICITÉ.....	64
3.	COMMENT DÉTECTER L'HÉTÉROSCÉDASTICITÉ ?	65
a)	Test Goldfeld-Quant.....	65
b)	Test de white.....	66
4.	COMMENT RÉSOUDRE L'HÉTÉROSCÉDASTICITÉ.....	66
a)	Matrice sandwich.....	66
b)	$u_i = \sigma_i^2$	67
c)	Variance est proportionnelle à une variable qui n'est pas dans le modèle	67
d)	La variance est proportionnelle à une variable de mon modèle	68
e)	Régression logarithmique	68
5.	RÉSUMÉ DES RÉGRESSIONS AVEC LES 4 SPÉCIFICATIONS ALTERNATIVES DU MODÈLE	69
a)	Quel modèle choisir ?.....	69
6.	PROBLÈME AVEC LE R^2	69

CHAPITRE 7: CONDITIONS DE GAUSS-MARKOV, EXOGÉNÉITÉ, ERREURS DE MESURE, VARIABLES INSTRUMENTALES ET SPÉCIFICATION DU TEST DE DURBIN-WU-HAUSMAN..... 70

A.	CONDITIONS DE GAUSS-MARKOV	70
B.	EXOGÉNÉITÉ	70
1.	RÉGRESSEURS STOCHASTIQUES.....	70
C.	CAS DE FIGURES QUI VIOLENT CETTE HYPOTHÈSE	71
1.	ENDOGENÉITÉ: BIAIS DE VARIABLES OMISES	71
2.	ERREURS DE MESURE.....	71
a)	Erreurs de mesure dans une variable explicative.....	71
b)	Erreur de mesure dans la variable dépendante	72
3.	BIAIS DE SIMULTANÉITÉ	72
D.	VARIABLES INSTRUMENTALES	73
1.	DÉMONSTRATION DES PROPRIÉTÉS DE L'ESTIMATEUR DE VARIABLE INSTRUMENTALE	74
E.	SPÉCIFICATION DU TEST DE DURBIN-WU-HAUSMAN	75

CHAPITRE 8: INTRODUCTION À L'ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE..... 76

A.	MAXIMUM DE VRAISEMBLANCE	76
1.	GRAPHIQUEMENT.....	76
2.	MATHÉMATIQUEMENT.....	77
3.	GÉNÉRALISATION.....	78
4.	ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE QUAND μ ET Σ SONT INCONNUS.....	79
5.	ESTIMATION DES COEFFICIENTS DE RÉGRESSION PAR MAXIMUM DE VRAISEMBLANCE	81

a) Mathématiquement.....	82
--------------------------	----

CHAPITRE 9: MODÈLE DE CHOIX BINAIRE : ANALYSES LINÉAIRE, LOGIT ET PROBIT 84

A. MODÈLE DE CHOIX BINAIRE: MODÈLE DE PROBABILITÉ LINÉAIRE.....	84
1. MODÈLE DE PROBABILITÉ LINÉAIRE	84
a) Quelles sont les limites liées au modèle de probabilité linéaire.....	85
B. ANALYSE LOGIT.....	86
a) Effet marginal.....	88
C. ANALYSE PROBIT.....	92

CHAPITRE 10 : INDÉPENDANCE SÉRIELLE : AUTOCORRELATION, VARIABLE LAGUÉE (SÉRIE CHRONOLOGIQUE), TEST DE FACTEUR COMMUN ET SPÉCIFICATION DE MODÈLE DYNAMIQUE..... 93

A. AUTOCORRELATION	93
a) Dépendance sérielle (autocorrélation) positive	93
b) Dépendance sérielle (autocorrélation) négative.....	93
B. DÉTECTION DE L'AUTOCORRÉLATION	96
1. TEST D'AUTOCORRÉLATION AR(1) DE DURBIN-WATSON.....	96
2. LE TEST DE BREUSH-GODFREY	98
C. ÉLIMINER L'AUTOCORRÉLATION AR(1)	99
D. AUTOCORRÉLATION DANS UN MODÈLE AVEC NE VARIABLE DÉPENDANTE LAGUÉE.....	100
E. TEST DE FACTEUR COMMUN ET SPÉCIFICATION DE MODÈLE DYNAMIQUE	102

Partie 1 : économétrie

Chapitre 1 : Variable aléatoire et estimateur

A. VARIABLE ALÉATOIRE

1. Qu'est ce qu'une variable aléatoire X?

Une variable aléatoire est une variable dont le résultat est déterminé par une expérience.

2. Exemple de distribution de probabilité: la somme de deux dés

La variable aléatoire X, est le résultat de la somme des chiffres apparaissant sur chaque dé. Par exemple: si le dé rouge donne 4 et le vert 6, X est égal à 10. Tous les résultats possibles sont présentés dans le tableau.

Les fréquences de X, f, sont le nombre de fois qu'un résultat concernant X apparait. Par exemple, il y a 4 résultats pour lesquels X est égal à 5.

Enfin on peut en déduire la probabilité d'obtenir chacune des valeurs de X. il y a donc une probabilité de $\frac{1}{36}$ (possibilités = $\frac{1}{6}$ d'obtenir un chiffre du dé rouge * $\frac{1}{6}$ d'obtenir un chiffre du dé vert). De ce fait, pour obtenir la probabilité associée aux différentes valeurs de X, nous faisons:
fréquences
possibilités

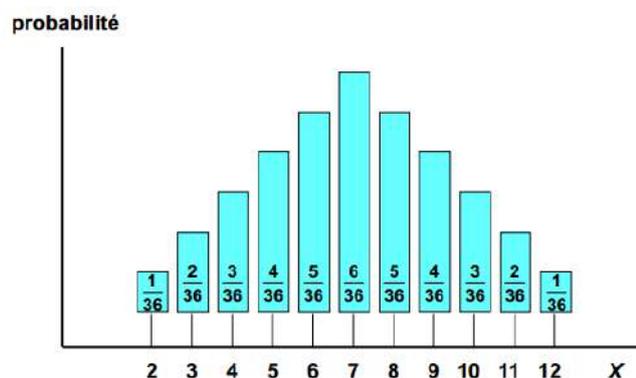
On peut également représenter graphiquement la fonction de probabilité (graph). Dans cet exemple la distribution est symétrique elle est la plus élevée pour X égale à 7.

(Fonctions de Probabilité)

rouge vert	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	P
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

Le 7 apparaît avec la plus grande fréquence.



Partie 1 : économétrie

3. Espérance d'une variable aléatoire

a) Définition de $E(X)$, l'espérance de X

$$E(X) = x_1 p_1 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i$$

Moyenne: $\frac{1}{n} \sum_i x_i$ $\frac{1}{n} \sum_j n_j x_j$ $\sum_j \frac{n_j}{n} x_j$

L'espérance d'une variable aléatoire, aussi appelée la moyenne de sa population, est la moyenne pondérée de ses valeurs possibles.

Dans l'exemple:

x_i	p_i	$x_i p_i$		x_i	p_i	$x_i p_i$
x_1	p_1	$x_1 p_1$		2	1/36	2/36
x_2	p_2	$x_2 p_2$		3	2/36	6/36
x_3	p_3	$x_3 p_3$		4	3/36	12/36
x_4	p_4	$x_4 p_4$		5	4/36	20/36
x_5	p_5	$x_5 p_5$		6	5/36	30/36
x_6	p_6	$x_6 p_6$		7	6/36	42/36
x_7	p_7	$x_7 p_7$		8	5/36	40/36
x_8	p_8	$x_8 p_8$		9	4/36	36/36
x_9	p_9	$x_9 p_9$		10	3/36	30/36
x_{10}	p_{10}	$x_{10} p_{10}$		11	2/36	22/36
x_{11}	p_{11}	$x_{11} p_{11}$		12	1/36	12/36
		$\sum x_i p_i = E(X)$				$252/36 = 7$

L'espérance s'avère être 7. En fait, nous le savions déjà puisque nous avons vu dans la séquence précédente que la distribution était symétrique autour de 7.

b) Notation alternative de $E(X)$

$$E(X) = \mu_x$$

! Lettre latine: on parle d'échantillon - Lettre grecque: on parle de population

Partie 1 : économétrie

4. Espérance d'une fonction d'une variable aléatoire

a) Définition de $E[g(X)]$, l'espérance d'une fonction de X

$$E[g(X)] = g(x_1)p_1 + \dots + g(x_n)p_n = \sum_{i=1}^n g(x_i)p_i$$

Prenons un exemple: $E(X^2) = x_1^2 p_1 + \dots + x_n^2 p_n = \sum_{i=1}^n x_i^2 p_i$

x_i	p_i	$g(x_i)$	$g(x_i) p_i$	x_i	p_i	x_i^2	$x_i^2 p_i$
x_1	p_1	$g(x_1)$	$g(x_1) p_1$	2	1/36	4	0.11
x_2	p_2	$g(x_2)$	$g(x_2) p_2$	3	2/36	9	0.50
x_3	p_3	$g(x_3)$	$g(x_3) p_3$	4	3/36	16	1.33
...	5	4/36	25	2.78
...	6	5/36	36	5.00
...	7	6/36	49	8.17
...	8	5/36	64	8.89
...	9	4/36	81	9.00
...	10	3/36	100	8.83
...	11	2/36	121	6.72
x_n	p_n	$g(x_n)$	$g(x_n) p_n$	12	1/36	144	4.00
			$\sum g(x_i) p_i$				54.83

La somme des valeurs pondérées est l'espérance de la fonction de X.

5. Propriété de l'espérance

Propriétés	Exemples
$E(X + Y) = E(X) + E(Y)$	$E(W + X + Y + Z) = E(W) + E(X) + E(Y) + E(Z)$
$E(bX) = bE(X)$	$E(3X) = 3 E(X)$
$E(b) = b$	$E(3) = 3$
$Y = b_1 + b_2 X$ $E(Y) = E(b_1 + b_2 X)$ $E(Y) = E(b_1) + E(b_2 X)$ $E(Y) = b_1 + b_2 E(X)$	

Partie 1 : économétrie

6. Variance de la population dans le cas d'une variable aléatoire discrète

a) Définition de la variance de la population de X: $E[(X - \mu)^2]$

$$E[(X - \mu)^2] = (x_1 - \mu)^2 p_1 + \dots + (x_n - \mu)^2 p_n = \sum_{i=1}^n (x_i - \mu)^2 p_i$$

La variance (mesure la dispersion): moyenne des écarts au carré par rapport au centre.

Dans l'exemple:

x_i	p_i	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 p_i$
2	1/36	-5	25	0.69
3	2/36	-4	16	0.89
4	3/36	-3	9	0.75
5	4/36	-2	4	0.44
6	5/36	-1	1	0.14
7	6/36	0	0	0.00
8	5/36	1	1	0.14
9	4/36	2	4	0.44
10	3/36	3	9	0.75
11	2/36	4	16	0.89
12	1/36	5	25	0.69
				5.83

La variance de la population de X est habituellement notée σ_x^2

$$E[(X - \mu)^2]$$

$$\sigma_x^2$$

L'écart type de X est la racine carrée de la variance de sa population. Noté σ_x . Il s'agit d'une mesure alternative de dispersion. L'écart type a les mêmes unités que X.

$$\sqrt{E[(X - \mu)^2]}$$

$$\sigma_x$$

Partie 1 : économétrie

7. Indépendance de deux variables aléatoires

Deux variables aléatoires X et Y sont dites indépendantes si et seulement si l'espérance des produits des fonctions est égale au produit des espérances des fonctions.

$$E[f(X)g(Y)] = E[f(X)] E[g(Y)] \text{ pour toutes fonctions } f(X) \text{ et } g(Y)$$

Cas particulier: si X et Y sont indépendants, $E(XY) = E(X)E(Y)$ l'espérance de XY est égale à l'espérance de X multipliée par l'espérance de Y ssi X et Y sont indépendants

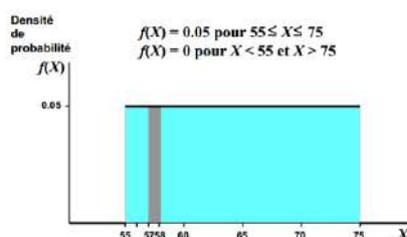
8. Variables aléatoires continues

La plupart des variables aléatoires rencontrées dans les expériences de la vie quotidienne sont continues. Elles peuvent prendre n'importe quelle valeur d'une série infinie de valeurs définies selon **un intervalle** (ou parfois **des intervalles**).

Dans le cas d'une variable aléatoire continue, la probabilité d'être égale à une valeur donnée finie est toujours infinitésimale. Pour cette raison, on peut seulement parler de la probabilité qu'une variable aléatoire continue se trouve entre deux valeurs données.

La représentation graphique est une **aire**. Notons enfin que l'aire continue entre minimum et maximum de l'intervalle est égale à 1 car on est sûr d'être entre les 2 bornes¹.

La probabilité par écart unitaire est appelée la densité de probabilité. Elle est égale à la hauteur du rectangle qui correspond à cet écart unitaire. On nomme l'axe vertical "**densité de probabilité**" plutôt que de le dénommer "hauteur". $f(X)$ est appelé fonction de densité de probabilité et est représenté graphiquement par la **ligne noire épaisse**.



9. Composantes fixes et aléatoires d'une variable aléatoire

Moyenne de la population de X: $E(X) = \mu_x$

Dans l'observation i, la composante aléatoire est donnée par: $u_i = x_i - \mu_x$

Donc, x_i peut être décomposé en composantes fixes et variables: $x_i = \mu_x + u_i$

Notons que l'espérance de u_i est zéro: $E(u_i) = E(x_i - \mu_x) = E(x_i) + E(-\mu_x) = \mu_x - \mu_x = 0$

¹ Voir les exemples dans le syllabus

B. ESTIMATEURS

1. Estimateurs

a) Définition estimateur et estimation

Un estimateur est une formule mathématique.

Une estimation est un nombre obtenu en appliquant cette formule à une série d'échantillons de données.

Exemples: un exemple d'estimateur est la moyenne de l'échantillon, qui est l'estimateur de la moyenne d'une population – Un autre estimateur commun est S^2 , il est utilisé pour estimer la variance de la population, σ_x^2

Caractéristiques d'une population	Estimateur
Moyenne: μ_x	$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$
Variance de la population: σ_x^2	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$

b) Les estimateurs sont des variables aléatoires

L'estimateur de la moyenne changera d'un échantillon à un autre. Nous démontrons cela dans le cas de la moyenne de l'échantillon.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$$

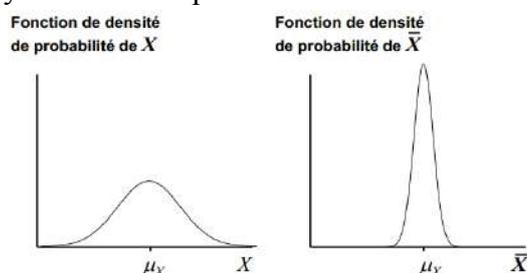
$$x_i = \mu_x + u_i$$

$$\bar{X} = \frac{1}{n} (\mu_x + \dots + \mu_x) + \frac{1}{n} (u_1 + \dots + u_n)$$

$$\bar{X} = \frac{1}{n} (n\mu_x) + \bar{u}$$

$$\bar{X} = \mu_x + \bar{u}$$

Les graphiques ci-dessous comparent les fonctions de densité de probabilité de X et de \bar{X} . La distribution de \bar{X} est plus concentrée. En effet sa composante aléatoire a tendance à être plus petite que celle de X car c'est la moyenne des composantes aléatoires de toutes les observations.



Partie 1 : économétrie

2. Estimateur non biaisé et efficacité

a) Estimateur non biaisé X

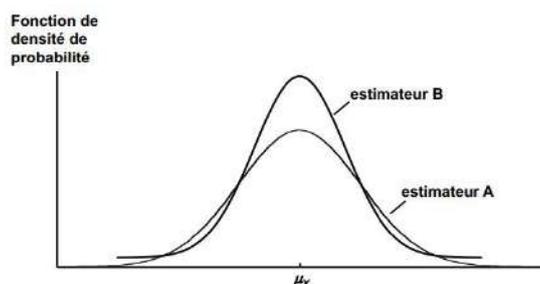
Un **estimateur est dit non biaisé** si l'espérance de mon estimateur est égale à la vraie valeur.

$$\begin{aligned} E(\bar{X}) &= E\left[\frac{1}{n}(x_1 + \dots + x_n)\right] = \frac{1}{n}E(x_1 + \dots + x_n) \\ &= \frac{1}{n}[E(x_1) + \dots + E(x_n)] = \frac{1}{n}n\mu_x = \mu_x \end{aligned}$$

b) Efficacité

Parmi les estimateurs non biaisés, il convient de choisir le plus efficace c'est-à-dire celui qui aura la plus petite variance, car il aura tendance à être le plus précis (il s'écarte moins de la vraie valeur).

Dans le diagramme ci-dessous, A et B sont tous deux des estimateurs non biaisés mais B est supérieur car il est plus efficace (écart-type plus petit)



3. Effet de l'augmentation de la taille d'échantillons sur la distribution \bar{X}

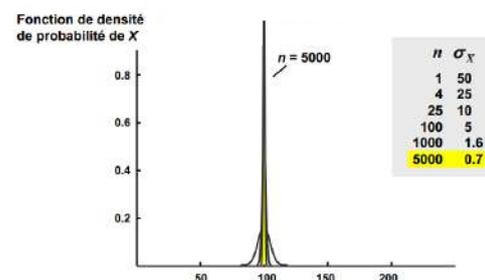
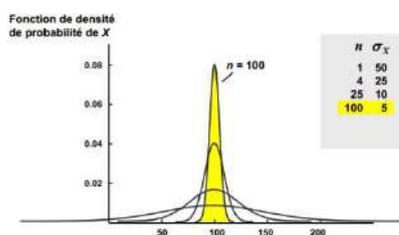
a) Convergence

Regardons à présent quels sont les effets d'une augmentation de la taille de l'échantillon sur la distribution de \bar{X} définissons la notion de convergence d'un estimateur.

Les graphiques ci-dessous montrent que plus l'échantillon est grand, plus l'écart type de la moyenne de l'échantillon est petit. La distribution devient plus concentrée autour de la population. Lorsque nous approchons de la limite, la variance de la distribution tend vers 0. La distribution échoue en un pic à la vraie valeur. **La moyenne de l'échantillon** est de ce fait un **estimateur convergent** de la moyenne de la population.

Convergent: la distribution se serre et le centre se rapproche de la vraie valeur.

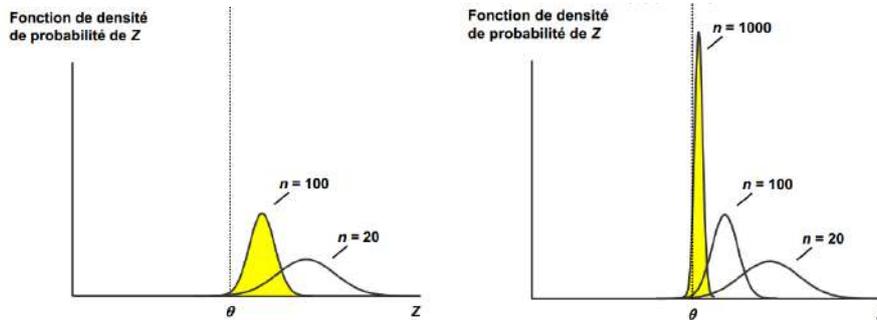
- ⇒ Diminution de la variance
- ⇒ Diminution du biais



Partie 1 : économétrie

4. Estimateur biaisé en échantillon fini mais convergent

Il se peut qu'un estimateur soit convergent mais qu'il soit biaisé en échantillon fini. Pour qu'un estimateur soit convergent, deux phénomènes doivent se produire: **le biais doit diminuer** lorsque n augmente et la **distribution doit échouer en un pic**.



Partie 1 : économétrie

Chapitre 2: Covariance, modèle de régression simple et test d'hypothèse

A. COVARIANCE

1. Covariance d'échantillon

a) Définition de la covariance d'échantillon²

La covariance d'échantillon est la moyenne des produits de leur écart autour de la moyenne de leur échantillon. Elle indique comment deux variables aléatoires varient ensemble mais attention **SEUL son signe est lisible!** La covariance indique si la relation est positive, négative ou nulle.

$$\begin{aligned} \text{Cov}(X,Y) &= \frac{1}{n} [(X_1 - \bar{X})(Y_1 - \bar{Y}) + \dots + (X_n - \bar{X})(Y_n - \bar{Y})] \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \end{aligned}$$

2. Règles de la covariance

Règles	Exemples
Si $Y = V + W$ $\text{Cov}(X,Y) = \text{Cov}(X,V) + \text{Cov}(X,W)$	
Si $Y = bZ$, où b est une constante $\text{Cov}(X,Y) = \text{Cov}(X,bZ) = b\text{Cov}(X,Z)$	$\text{Cov}(X,3Z) = 3 \text{Cov}(X,Z)$
Si $Y = b$, où b est une constante $\text{Cov}(X,Y) = \text{Cov}(X,b) = 0$	$\text{Cov}(X,10) = 0$
$Y = b_1 + b_2Z$ $\text{Cov}(X,Y) = \text{Cov}(X, [b_1 + b_2Z])$ $\text{Cov}(X,Y) = \text{Cov}(X,b_1) + \text{Cov}(X,b_2Z)$ $\text{Cov}(X,Y) = 0 + \text{Cov}(X,b_2Z)$ $\text{Cov}(X,Y) = b_2\text{Cov}(X,Z)$	

² Voir syllabus pour un exemple de calcul de la covariance d'échantillon.

Partie 1 : économétrie

3. Variance d'échantillon et règles de la variance

a) Définition de la variance d'un échantillon

La variance de l'échantillon est donc la moyenne du carré des écarts par rapport à la moyenne. C'est aussi la covariance de X par rapport à X

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})$$

$$\text{Var}(X) = \text{Cov}(X, X)$$

b) Propriétés

Propriétés	Commentaires
Si $Y = V + W$ $\text{Var}(Y) = \text{Var}(V) + \text{Var}(W) + 2\text{Cov}(V, W)$	Sauf si V et W sont indépendants. Dans ce cas, la covariance est nulle.
Si $Y = bZ$, où b est une constante $\text{Var}(Y) = b^2 \text{Var}(Z)$	
Si $Y = b$, où b est une constante $\text{Var}(Y) = 0$	
Si $Y = V + b$, où b est une constante $\text{Var}(Y) = \text{Var}(V)$	

La variance de l'échantillon est un estimateur biaisé de la variance de la population. On peut facilement montrer que cet estimateur est biaisé vers le bas d'un facteur $(n-1)/n$.

$$E[\text{Var}(X)] = \frac{n-1}{n} \sigma_x^2$$

$$s_x^2 = \frac{n}{n-1} \text{Var}(X)$$

$$E[s_x^2] = E\left[\frac{n}{n-1} \text{Var}(X)\right]$$

$$= \frac{n}{n-1} E[\text{Var}(X)]$$

$$= \frac{n}{n-1} \times \frac{n-1}{n} \sigma_x^2 = \sigma_x^2$$

Partie 1 : économétrie

4. Coefficient de corrélation d'une population

On divise par le produit des écarts types pour standardiser les unités c'est-à-dire se débarrasser des unités.

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Les bornes sont **1 et -1**

Lorsqu'il y a **une corrélation parfaite positive** entre les variables.

Lorsqu'il y a **une corrélation parfaite mais négative**.

Le coefficient de corrélation est **nul** s'il n'y a **pas de relation** entre les variables.

a) Coefficient de corrélation d'un échantillon

Pour un échantillon donné, il peut être estimé en remplaçant la covariance de la population ainsi que ses variances par l'estimateur non biaisé.

$$r_{XY} = \frac{\frac{n}{n-1} Cov(X, Y)}{\sqrt{\frac{n}{n-1} Var(X) \frac{n}{n-1} Var(Y)}} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Exemple: Comme le montre l'exemple (voir syllabus pour plus de détail), nous trouvons que le coefficient de corrélation est de 0.55. Il est positif car il y a une corrélation positive entre S et Y.

Obs	S	Y	S - \bar{S}	Y - \bar{Y}	(S - \bar{S})(Y - \bar{Y})	(S - \bar{S}) ²	(Y - \bar{Y}) ²
1	15	17.24	1.75	3.016	5.277	3.063	9.093
2	16	15.00	2.75	0.776	2.133	7.563	0.601
3	8	14.91	-5.25	0.686	-3.599	27.563	0.470
4	6	4.50	-7.25	-9.725	70.503	52.563	94.566
5	15	18.00	1.75	3.776	6.607	3.063	14.254
6	12	6.29	-1.25	-7.935	9.918	1.563	62.956
7	12	19.23	-1.25	5.006	-6.257	1.563	25.055
8	18	18.69	4.75	4.466	21.211	22.563	19.941
9	12	7.21	-1.25	-7.015	8.768	1.563	49.203
10	20	42.06	6.75	27.836	187.890	45.563	187.890
...
...
19	12	7.50	-1.25	-6.725	8.406	1.563	45.219
20	14	8.00	0.75	-6.225	-4.668	0.563	38.744
Total	265	284.49			305.888	217.760	1542.160
Ave	13.25	14.225			15.294	10.888	77.108
					Cov(S, Y)	Var(S)	Var(Y)

Cov(S, Y) = 15.294

Var(S) = 10.888

Var(Y) = 77.108

$$r_{SY} = \frac{Cov(S, Y)}{\sqrt{Var(S)Var(Y)}} = \frac{15.294}{\sqrt{10.888 \times 77.108}} = \frac{15.294}{28.975} = 0.55$$

Notes:

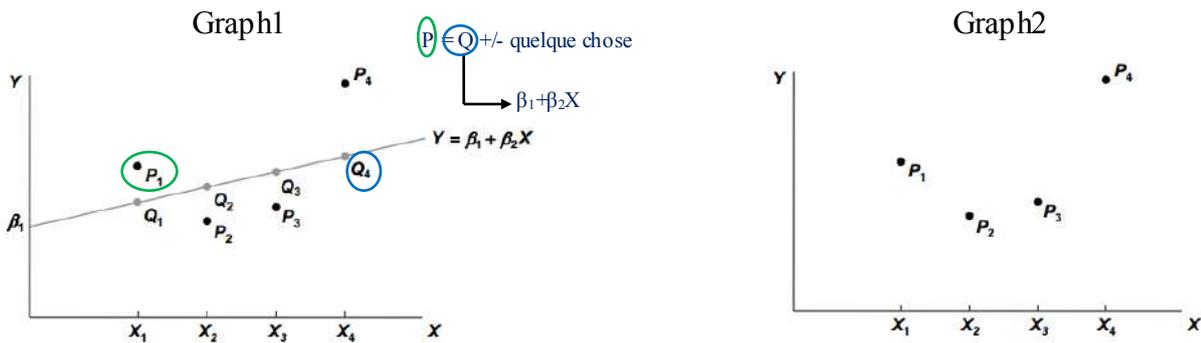
- ⇒ Le signe de la corrélation et la covariance sera toujours le même.
- ⇒ Pour expliquer le coefficient de corrélation il faut l'élever au carré. Ici, $0.55^2 = 30\%$.
Explication: 30 % des variations en termes de salaire sont expliqués en termes d'éducation.

Partie 1 : économétrie

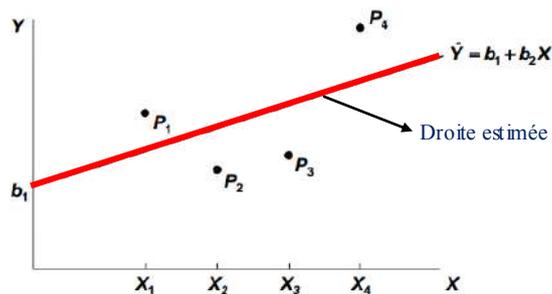
B. MODÈLE DE RÉGRESSION SIMPLE³

1. Modèle de régression simple

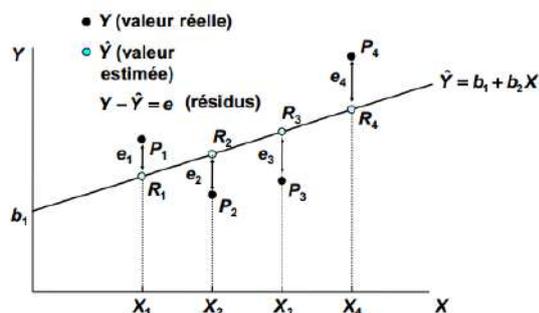
En pratique, la plupart des relations économiques ne sont pas exactes et les valeurs réelles de Y sont différentes de celles qui correspondent à une ligne droite. Pour permettre de telles divergences, nous réécrivons le modèle tel que $Y = \beta_1 + \beta_2 X + u$ où u est le **terme d'erreur** (graph1). Chaque valeur de Y a donc une composante non aléatoire, $\beta_1 + \beta_2 X$, et une composante aléatoire u . En pratique, nous observons uniquement les points P (graph2).



On va donc essayer de chercher une droite qui minimise la somme des distances des points par rapport à la droite. Pour cela, utilisons les points P pour tracer **une droite estimée** $Y = b_1 + b_2 X$ où b_1 et b_2 est une estimation de β_1, β_2 .



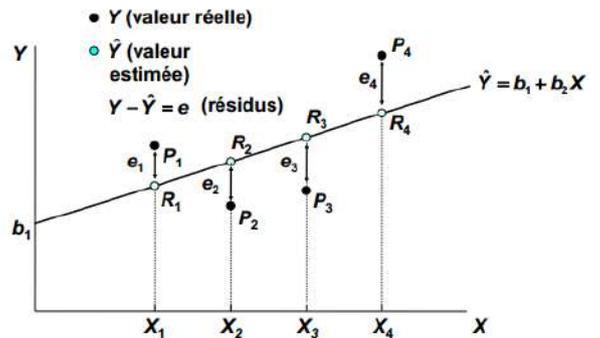
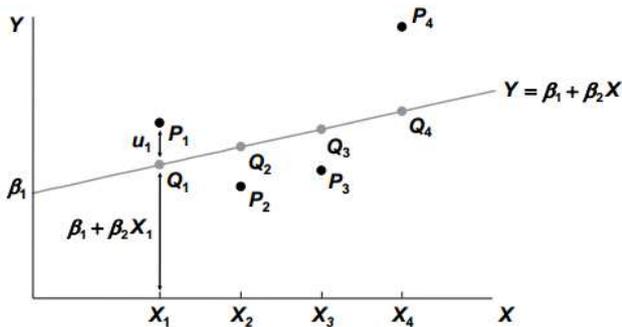
La ligne est appelée "**le modèle estimé**" et les valeurs de Y trouvées sont appelées "**valeurs estimées**". Elles sont données par la hauteur des points R . Les écarts entre les valeurs réelles et les valeurs estimées de Y sont appelés **résidus (e)**.



³ Voir dans le syllabus tout le début sur le modèle de la régression simple. Dans ce point, je passe directement à l'essentiel.

Partie 1 : économétrie

Notons que les valeurs des résidus ne sont pas les mêmes que les valeurs du terme d'erreur. Si l'estimation est bonne ils seront similaires. Mais se sont des concepts différents. En effet, **l'erreur** est la **distance verticale** entre le **point et la vraie droite**. Tandis que **les résidus** est l'**écart** entre les **valeurs réelles et les valeurs estimées** de Y.



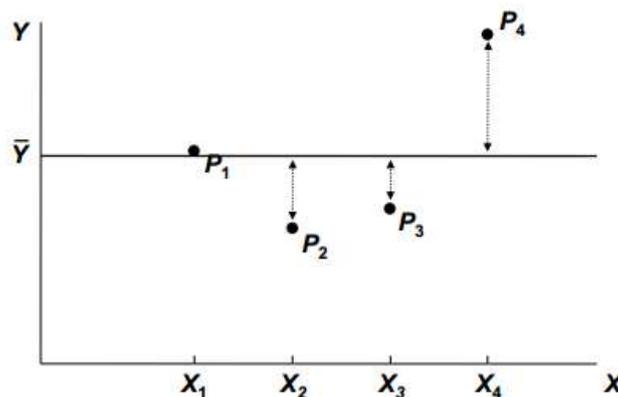
a) Critère des Moindres Carrés: (Residual Sum of Squares: RSS)

La droite estimée est tracée en minimisant la somme des carrés des résidus:

$$RSS = \sum_{i=1}^n e_i^2 = e_1^2 + \dots + e_n^2$$

Pourquoi minimiser la somme des carrées des résidus?

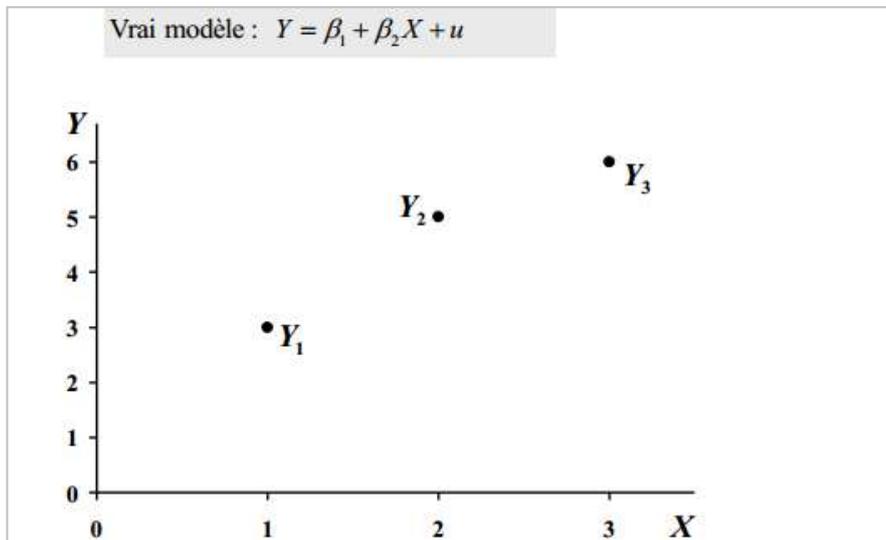
Parce-que l'on obtiendrait une ligne horizontale au niveau de la valeur moyenne de Y. La somme des résidus serait 0. On doit donc empêcher les résidus négatifs d'annuler les positifs et donc utiliser le carré des résidus pour se faire. Représentation graphique:



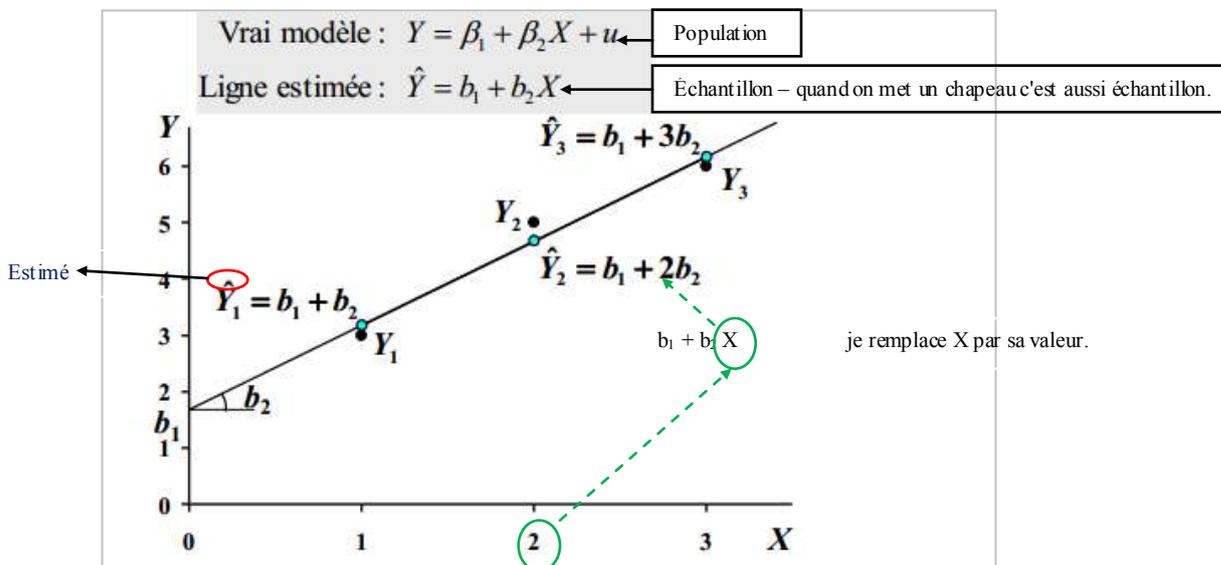
Partie 1 : économétrie

2. Déduire les coefficients de régressions linéaires

Déterminons à présent les valeurs b_1 et b_2 qui minimisent la somme des carrés des résidus via un exemple numérique.

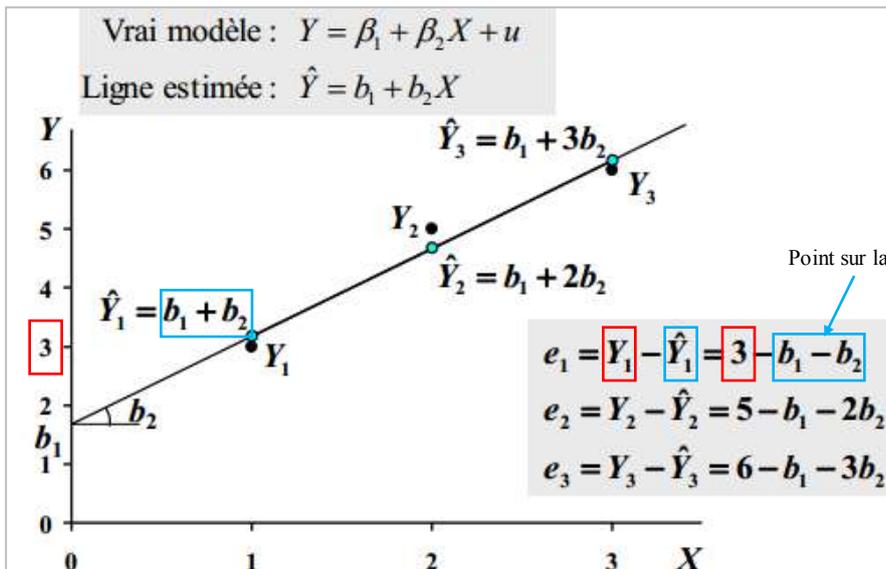


3 observations: (1,3); (2,5) et (3,6)



En écrivant la régression estimée: $\hat{Y} = b_1 + b_2 X$, nous déterminerons les valeurs de b_1 et b_2 qui minimisent RSS, la somme des carrés des résidus.

Partie 1 : économétrie



$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$ → Somme des carrés des résidus

$$= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2$$

$$+ 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2$$

$$+ 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2$$

$$= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

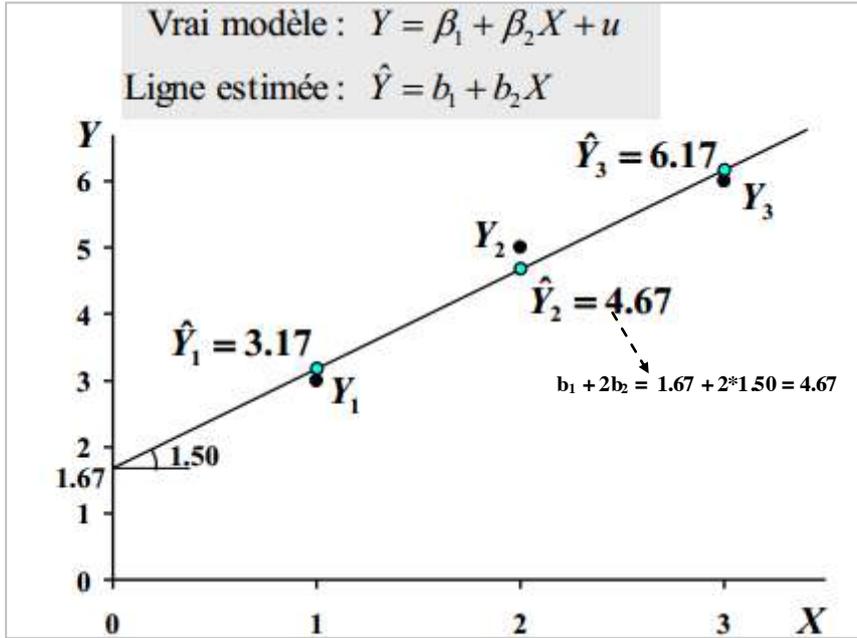
→ Polynôme du second degré

$$\frac{\partial RSS}{\partial b_1} = 0 \Rightarrow 6b_1 + 12b_2 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \Rightarrow 12b_1 + 28b_2 - 62 = 0$$

Constante ← $b_1 = 1.67$, $b_2 = 1.50$ → Pente
 Quand X augmente de 1 unité, Y augmente de 1.50 unité

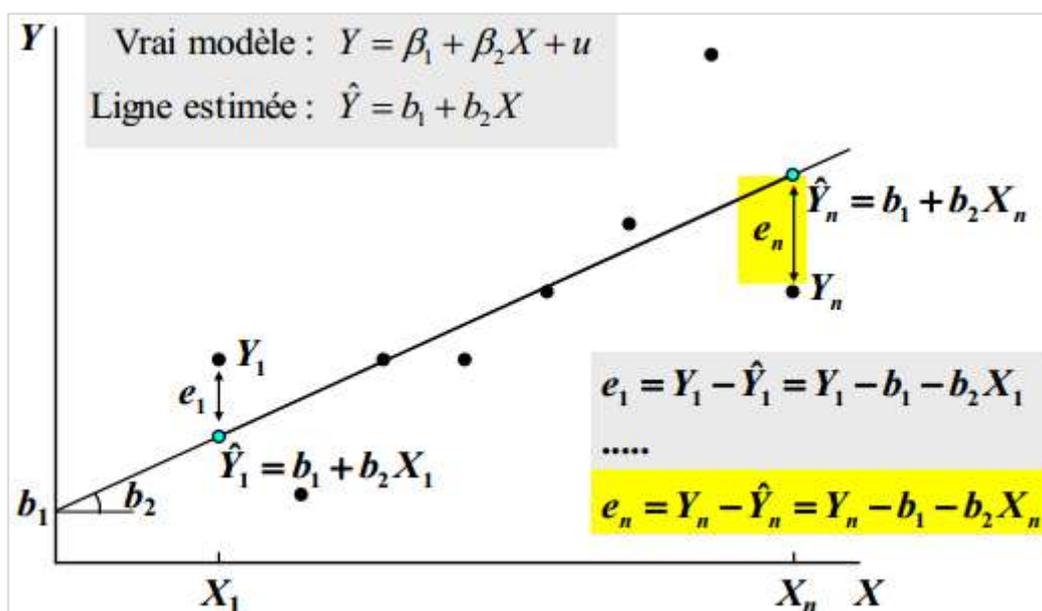
Nous trouvons que RSS est minimisée quand b_1 et b_2 valent 1.67 et 1.50



Partie 1 : économétrie

Nous allons faire la même chose pour le cas général avec n observations.

1) Définir les résidus de toutes les observations



2) Réaliser la somme des carrés des résidus. Les termes équivalents sont additionnés ensemble.

$$\begin{aligned}
 RSS &= e_1^2 + \dots + e_n^2 = (Y_1 - b_1 - b_2 X_1)^2 + \dots + (Y_n - b_1 - b_2 X_n)^2 \\
 &= Y_1^2 + b_1^2 + b_2^2 X_1^2 - 2b_1 Y_1 - 2b_2 X_1 Y_1 + 2b_1 b_2 X_1 \\
 &\quad + \dots \\
 &\quad + Y_n^2 + b_1^2 + b_2^2 X_n^2 - 2b_1 Y_n - 2b_2 X_n Y_n + 2b_1 b_2 X_n \\
 &= \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i
 \end{aligned}$$

3) Dérivée première par rapport à b_1

$$\begin{aligned}
 RSS &= \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i \\
 \frac{\partial RSS}{\partial b_1} &= 0 \Rightarrow 2nb_1 - 2\sum Y_i + 2b_2 \sum X_i = 0 \\
 nb_1 &= \sum Y_i - b_2 \sum X_i \quad b_1 = \bar{Y} - b_2 \bar{X}
 \end{aligned}$$

Partie 1 : économétrie

4) Dérivée première par rapport à b_2 . Substituer b_1 en utilisant l'expression obtenue.

$$\begin{aligned}
 RSS &= \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i \\
 \frac{\partial RSS}{\partial b_2} &= 0 \Rightarrow 2b_2 \sum X_i^2 - 2 \sum \sum X_i Y_i + 2b_1 \sum X_i = 0 \\
 b_2 \sum X_i^2 - \sum X_i Y_i + b_1 \sum X_i &= 0 \\
 b_2 \sum X_i^2 - \sum X_i Y_i + (\bar{Y} - b_2 \bar{X}) \sum X_i &= 0
 \end{aligned}$$

5) Utiliser la définition de la moyenne d'échantillon. Nous obtenus une expression pour b_2 .

$$\begin{aligned}
 \frac{\partial RSS}{\partial b_2} &= 0 \Rightarrow 2b_2 \sum X_i^2 - 2 \sum \sum X_i Y_i + 2b_1 \sum X_i = 0 \\
 b_2 \sum X_i^2 - \sum X_i Y_i + b_1 \sum X_i &= 0 \\
 b_2 \sum X_i^2 - \sum X_i Y_i + (\bar{Y} - b_2 \bar{X}) \sum X_i &= 0 \\
 b_2 \sum X_i^2 - \sum X_i Y_i + (\bar{Y} - b_2 \bar{X}) n \bar{X} &= 0 \\
 b_2 (\sum X_i^2 - n \bar{X}^2) &= \sum X_i Y_i - n \bar{X} \bar{Y} \\
 b_2 \left(\frac{1}{n} \sum X_i^2 - \bar{X}^2 \right) &= \frac{1}{n} \sum X_i Y_i - \bar{X} \bar{Y}
 \end{aligned}$$

$$\begin{aligned}
 \bar{X} &= \frac{\sum X_i}{n} \\
 \sum X_i &= n \bar{X}
 \end{aligned}$$

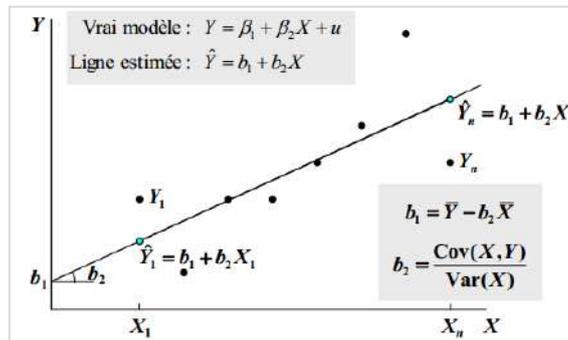
$$b_2 \text{Var}(X) = \text{Cov}(X, Y)$$

$$b_2 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

Partie 1 : économétrie

6) Graphiquement (voir également la matrice sur la feuilles du cours)

Nous avons déterminé les valeurs pour b_1 et b_2 qui minimisaient la somme des carrés des résidus.



3. Interprétation d'une équation de régression

```
. reg EARNINGS S
```

Coefficient de corrélation - 10 % des différences salariales sont expliquées par des différences d'éducation.

Residual	34419.6569	568	60.5979875
Total	38397.0371	569	67.4816117

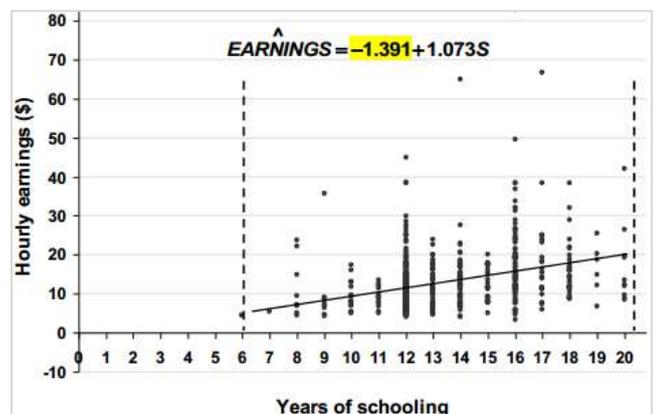
Number of obs =	570
F(1, 568) =	65.64
Prob > F =	0.0000
R-squared =	0.1036
Adj R-squared =	0.1020
Root MSE =	7.7845

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
S	1.073055	.1324501	8.102	0.000	.8129028 1.333206
_cons	-1.391004	1.820305	-0.764	0.445	-4.966354 2.184347

Pour présenter cette section nous nous baserons sur l'exemple montrant la relation entre le plus haut niveau d'étude accompli (S) et le salaire horaire (EARNINGS).

Pour chaque année d'étude supplémentaire, le salaire horaire augmente de 1.073\$. Ce 1.073 est l'estimation de β_2 (coefficient de pente).

Concernant la **constante**, mot à mot elle indique qu'un individu sans année d'étude devrait payer 1.39\$ par heure pour avoir le droit de travailler. Ceci n'est pas plausible. Une solution permettant de solutionner le problème est de limiter l'interprétation à l'intervalle des données de l'échantillon, et de refuser d'extrapoler en partant du principe que nous n'avons pas de preuve à l'extérieur de cet intervalle (graph1).



Avec cette explication, la seule fonction du terme de la constante est de permettre de tracer la ligne de régression à la hauteur correcte sur le diagramme de dispersion. Le terme de la constante n'a pas de sens propre (si le domaine de définition des variables commence à 0 alors OK). En effet, mettre une constante à zéro c'est faux dans cet exemple car on l'oblige à passer par l'origine.

Partie 1 : économétrie

4. Validité de l'estimation

a) Mesures de validité de l'estimation dans les analyses de régression

I. $\bar{e} = 0$ la moyenne des résidus vaut zéro

$$\begin{aligned}
 e_i &= Y_i - \hat{Y}_i = (Y_i - b_1 - b_2 X_i) \\
 \sum e_i &= \sum (Y_i - nb_1 - b_2) \sum X_i \\
 \frac{1}{n} \sum e_i &= \frac{1}{n} \sum Y_i - b_1 - b_2 \frac{1}{n} \sum X_i \quad \text{Divise par n} \\
 \bar{e} &= \bar{Y} - b_1 - b_2 \bar{X} \quad b_1 = \bar{Y} - b_2 \bar{X} \\
 &= \bar{Y} - (\bar{Y} - b_2 \bar{X}) - b_2 \bar{X} = 0
 \end{aligned}$$

II. $\bar{\hat{Y}} = \bar{Y}$ la moyenne des valeurs estimées de Y = la moyenne des valeurs réelles de Y

$$\begin{aligned}
 e_i &= (Y_i - \hat{Y}_i) = Y_i - b_1 - b_2 X_i \\
 \sum e_i &= \sum Y_i - \sum \hat{Y}_i \\
 \frac{1}{n} \sum e_i &= \frac{1}{n} \sum Y_i - \frac{1}{n} \sum \hat{Y}_i \\
 \bar{e} &= \bar{Y} - \bar{\hat{Y}} \quad \bar{\hat{Y}} = \bar{Y}
 \end{aligned}$$

Vaut 0

III. $Cov(\hat{Y}, e) = 0$ la covariance entre les valeurs estimées de Y et les résidus = 0

$$\begin{aligned}
 Cov(\hat{Y}, e) &= Cov([b_1 + b_2 X], e) = Cov(b_1, e) + Cov(b_2 X, e) \\
 &= 0 + b_2 Cov(X, e) = b_2 Cov(X, [Y - b_1 - b_2 X]) \\
 &= b_2 [Cov(X, Y) - Cov(X, b_1) - Cov(X, b_2 X)] \\
 &= b_2 [Cov(X, Y) - b_2 Cov(X, X)] \quad \text{Vaut 0} \\
 &= b_2 \left[Cov(X, Y) - \frac{Cov(X, Y)}{Var(X)} Var(X) \right] = 0
 \end{aligned}$$

Partie 1 : économétrie

b) Validité de l'estimation

Critère du R^2

Le critère principal de validité de l'estimation, décrit comme le coefficient de détermination, noté R^2 est défini comme le ratio de ESS sur TSS, c'est-à-dire la proportion de la variance de Y expliquée par l'équation de régression.

A la fin nous obtiendrons **TSS** ou total sum of squares (somme des carrés totaux). **ESS** (explain sum of squares) ou somme des carrés expliquée. Enfin **RSS** n'est rien d'autre que la somme du carré des résidus.

Le coefficient de détermination R^2 est égal au ratio de ESS sur TSS, c'est-à-dire la proportion de la variance de Y expliquée par l'équation de régression. Borne 0: j'explique tout et borne 1: j'explique rien.

$$e_i = Y_i - \hat{Y}_i \Rightarrow Y_i = \hat{Y}_i + e_i$$

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(\hat{Y} + e) = \text{Var}(\hat{Y}) + \text{Var}(e) + 2\text{Cov}(\hat{Y}, e) \\ &= \text{Var}(\hat{Y}) + \text{Var}(e) \end{aligned}$$

Égal à 0

$$\left(\frac{1}{n} \sum (Y - \bar{Y})^2 \right) = \left(\frac{1}{n} \sum (\hat{Y} - \bar{Y})^2 \right) + \left(\frac{1}{n} \sum (e - \bar{e})^2 \right)$$

Égal à 0

$$\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum e^2$$

Variation totale

$$TSS = ESS + RSS$$

Somme des carrés expliqués (valeur que j'explique -> ce qui est sur la droite).

La moyenne estimée de Y = la moyenne des valeurs réelles.

Somme des carrés des résidus (valeur que je n'explique pas).

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

Un R^2 élevé, un objectif en contradiction avec le critère RSS ?

De manière évidente nous voudrions placer la ligne de régression de manière à avoir une validité de l'estimation la plus élevée possible. Cet objectif n'est pas en contradiction avec le critère de RSS.

Il suffit pour cela d'effectuer la décomposition de la variance pour obtenir une expression alternative de R^2 .

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

Partie 1 : économétrie

Autre critère de validité de l'estimation: corrélation entre valeurs réelles et valeurs estimées de Y

Nous allons démontrer que ceci est maximisé en utilisant le principe des moindres carrés pour déterminer les coefficients de régression.

$$\begin{aligned}
 r_{Y,\hat{Y}} &= \frac{\text{Cov}(Y, \hat{Y})}{\sqrt{\text{Var}(Y) \text{Var}(\hat{Y})}} = \frac{\text{Cov}(\hat{Y} + e, \hat{Y})}{\sqrt{\text{Var}(Y) \text{Var}(\hat{Y})}} \\
 &= \frac{\text{Cov}(\hat{Y}, \hat{Y}) + \text{Cov}(e, \hat{Y})}{\sqrt{\text{Var}(Y) \text{Var}(\hat{Y})}} = \frac{\text{Var}(\hat{Y})}{\sqrt{\text{Var}(Y) \text{Var}(\hat{Y})}} \\
 &= \frac{\sqrt{\text{Var}(\hat{Y}) \text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y) \text{Var}(\hat{Y})}} = \frac{\sqrt{\text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y)}} \quad \text{Égal à 0} \\
 &= \sqrt{R^2}
 \end{aligned}$$

Permet de voir si le modèle est bon ou mauvais

Le coefficient de corrélation est donc la racine carrée du R².

Le R² ajusté

Notons que R² augmente toujours quand le nombre de X augmente. Un R² ajusté peut être utilisé pour prendre en compte la perte de degrés de liberté lié à l'inclusion de variables explicatives dans le modèle estimé.

$$R_a^2 = 1 - \frac{n-1}{n-k} (1 - R^2)$$

5. Les coefficients de régression: des variables aléatoires distribuées comme une normale

Les coefficients de régression sont une sorte spécifique de variables aléatoires: cela peut être montré via la définition de b₂. Démontrons que b₂ dépend à la fois de β₂ (composante commune) et d'une composante aléatoire qui varie selon l'échantillon.

$$Y = \beta_1 + \beta_2 X + u$$

$$\hat{Y} = b_1 + b_2 X$$

Dès que l'on peut développer une variable aléatoire en partie commune et idiosyncratique c'est une variable aléatoire.

$$\begin{aligned}
 b_2 &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\text{Cov}(X, \beta_1 + \beta_2 X + u)}{\text{Var}(X)} \\
 &= \frac{\text{Cov}(X, \beta_1) + \text{Cov}(X, \beta_2 X) + \text{Cov}(X, u)}{\text{Var}(X)} \\
 &= \frac{0 + \beta_2 \text{Cov}(X, X) + \text{Cov}(X, u)}{\text{Var}(X)} \\
 &= \beta_2 + \frac{\text{Cov}(X, u)}{\text{Var}(X)}
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}(X, X) &= \text{Var}(X) \\
 \beta_2 \frac{\text{Var}(X)}{\text{Var}(X)} + \frac{\text{Cov}(X, u)}{\text{Var}(X)}
 \end{aligned}$$

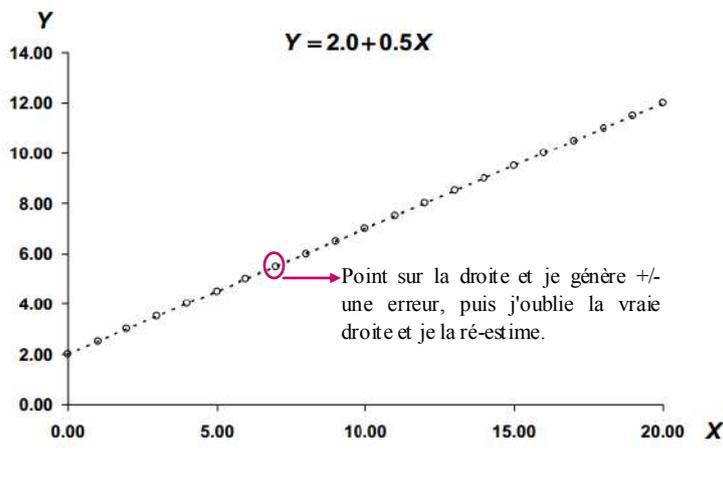
Partie commune

Partie idiosyncratique

Partie 1 : économétrie

Ce terme est également distribué comme une normale. Démontrons-le!

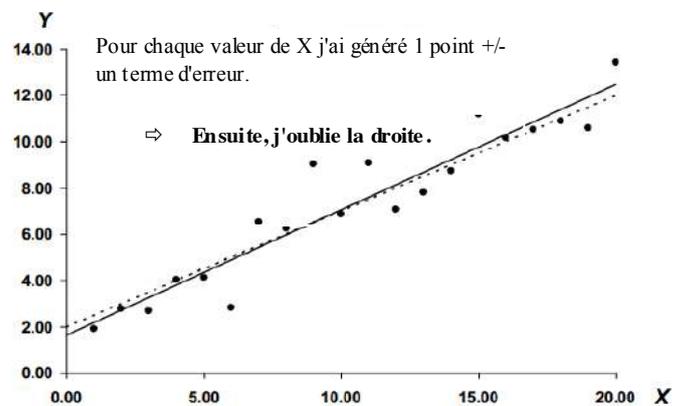
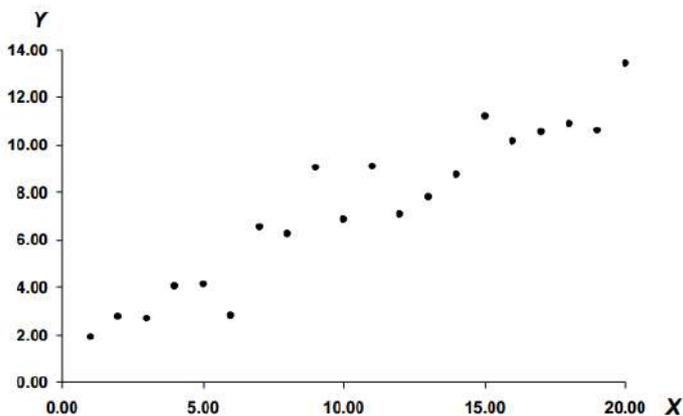
Pour cela nous allons partir de la vraie relation et ensuite générer aléatoirement une valeur du terme résiduel pour chaque observation en utilisant une distribution $N(0,1)$ (distribution normale avec une moyenne égale à 0 et de variance 1). Donc par exemple, la valeur de Y dans la première observation est 1.91 et non 2.5 (tableau 2.14).



$$Y = 2.0 + 0.5X + u$$

X	2.0+0.5X	u	Y	X	2.0+0.5X	u	Y
1	2.5	-0.59	1.91	11	7.5	1.59	9.09
2	3.0	-0.24	2.76	12	8.0	-0.92	7.08
3	3.5	-0.83	2.67	13	8.5	-0.71	7.79
4	4.0	0.03	4.03	14	9.0	-0.25	8.75
5	4.5	-0.38	4.12	15	9.5	1.69	11.19
6	5.0	-2.19	2.81	16	10.0	0.15	10.15
7	5.5	1.03	6.53	17	10.5	0.02	10.52
8	6.0	0.24	6.24	18	11.0	-0.11	10.89
9	6.5	2.53	9.03	19	11.5	-0.91	10.59
10	7.0	-0.13	6.87	20	12.0	1.42	13.42

Mais en réalité nous n'observons que le "nuages" de points. Il faut donc oublier la vraie droite et faire passer une droite estimée à travers ce scatter.



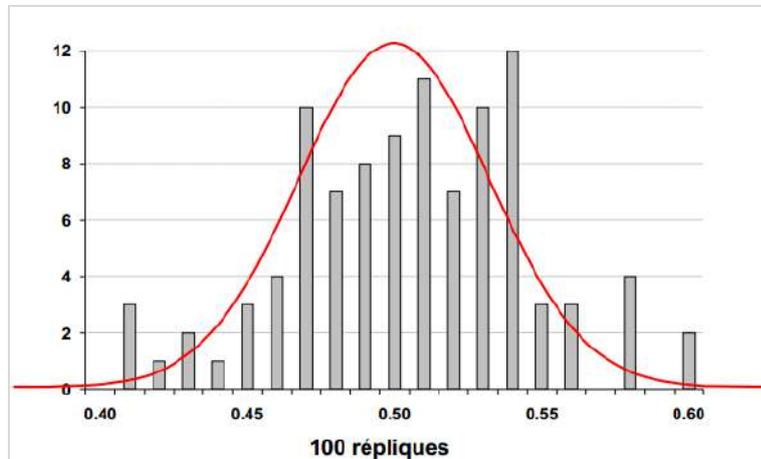
Il faudra ensuite répéter le même processus. La droite estimée sera différente de la vraie droite mais sera toujours proche⁴.

⁴ Voir plus d'exemple dans les slides.

Partie 1 : économétrie

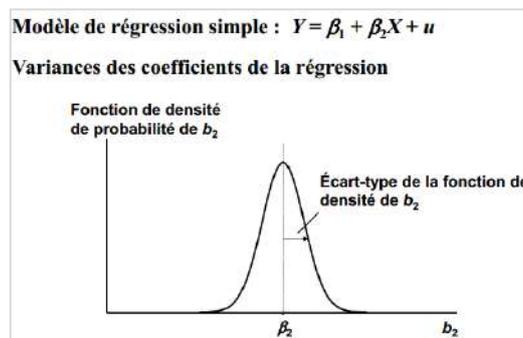
Lorsqu'on regarde l'histogramme (estimations de β_2 après un certains nombres de répliques), on observe que la distribution est normale centrée autour de la vraie valeur. Ce qui implique que **l'estimateur soit non biaisé**.

En conclusion les coefficients de régression sont des **variables aléatoires** distribués comme une **Normale**.



6. Précision des coefficients de régression

Nous allons nous intéresser aux estimations des écart-types des distributions. Ce qui nous donnera une idée quant à leur fiabilité et nous donnera une base pour tester les hypothèses.



Modèle de régression simple : $Y = \beta_1 + \beta_2 X + u$

Variances des coefficients de la régression

$$\sigma_{b_1}^2 = \frac{\sigma_u^2}{n} \left\{ 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right\} \quad \sigma_{b_2}^2 = \frac{\sigma_u^2}{n \text{Var}(X)}$$

Variance résiduelle

$$s_u^2 = \frac{n}{n-k} \text{Var}(e) = \frac{n}{n-2} \text{Var}(e)$$

$$\text{s.e.}(b_1) = \sqrt{\frac{s_u^2}{n} \left\{ 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right\}} \quad \text{s.e.}(b_2) = \sqrt{\frac{s_u^2}{n \text{Var}(X)}}$$

Dans une analyse de régression simple, $k = 2$

Partie 1 : économétrie

Modèle de régression simple : $Y = \beta_1 + \beta_2 X + u$

Variances des coefficients de la régression

Statistique qui test si le coefficient est égal à 0.

$$\frac{\beta_2 - \text{valeur du test}}{\text{écart-type}} = \frac{1.0 - 0}{0.13} = 8.102$$

 On rejette car > 2 (1.96)

MS	
3977.38016	
60.5979875	
67.4816117	

Donne les valeurs à partir desquelles je ne rejette pas l'H₀

Number of obs =	570
F(1, 568) =	65.64
Prob > F =	0.0000
R-squared =	0.1036
Adj R-squared =	0.1020
Root MSE =	7.7845

Exemple: on teste l'hypothèse que la vraie valeur = 1,5. On rejette car elle n'est pas dans l'intervalle de confiance.

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
s	1.073055	.1324501	8.102	0.000	.8129028 1.333206
_cons	-1.391004	1.820305	-0.764	0.445	-4.966354 2.184347

C. TEST D'HYPOTHÈSE

1. Test d'hypothèse d'un coefficient de régression

TESTS BILATERAUX

Supposons que nous voulons tester l'hypothèse H₀ que le coefficient est égal à une certaine valeur. Nous la testerons contre l'hypothèse alternative H₁, qui est simplement que β₂ n'est pas égal à β₂⁰.

Modèle: $Y = \beta_1 + \beta_2 X + u$

Hypothèse nulle: $H_0 : \beta_2 = \beta_2^0$ ← On travail sur l'H₀ car l'H₁ on ne la connaît pas.

Hypothèse alternative: $H_1 : \beta_2 \neq \beta_2^0$

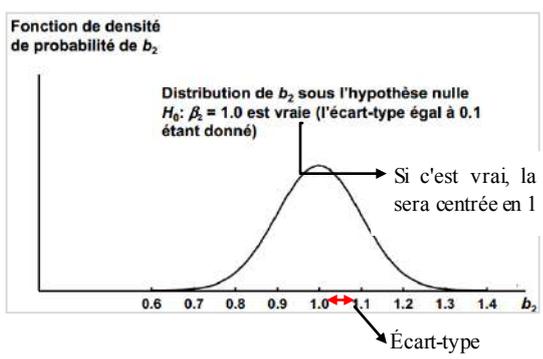
Inférer: conclure des choses à partir de l'échantillon au niveau de la population car ce qui m'intéresse c'est la population.
 ⇒ Tests d'hypothèses

Comme illustration, nous considérons un modèle reliant l'inflation des prix à l'inflation des salaires. Nous testerons l'hypothèse que le taux d'inflation des prix est égal au taux d'inflation des salaires. L'hypothèse nulle est donc H₀: β₂ = 1.0

Modèle d'exemple: $p = \beta_1 + \beta_2 w + u$
Hypothèse nulle: $H_0 : \beta_2 = 1.0$
Hypothèse alternative: $H_1 : \beta_2 \neq 1.0$

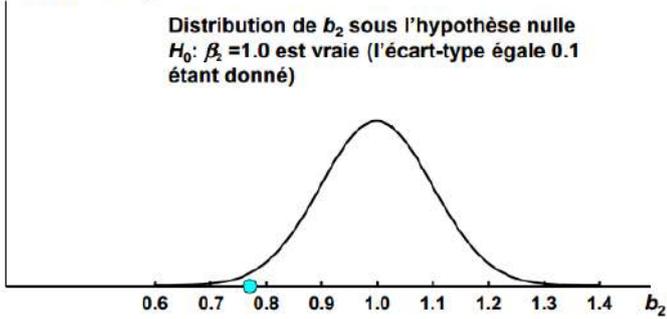
Si H₀ est vraie, b₂ aura une distribution avec une moyenne de 1.0. Nous faisons l'hypothèse que nous connaissons l'écart-type (sd) de cette distribution (en pratique nous devons l'estimer). La deuxième figure montre le cas général.

Écart-type

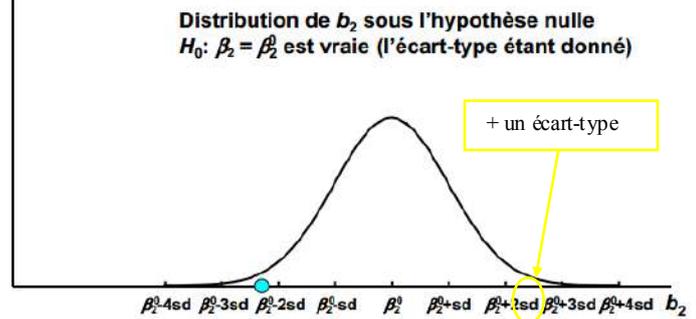


Partie 1 : économétrie

Fonction de densité de probabilité de b_2

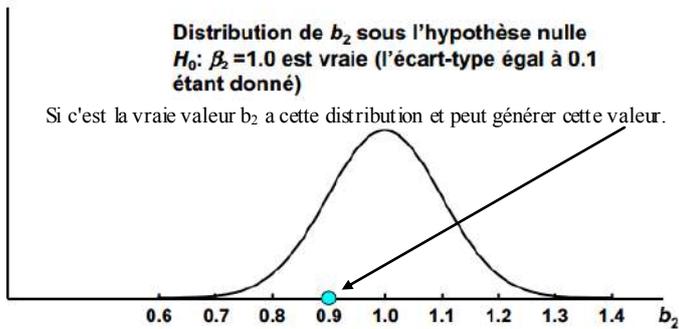


Fonction de densité de probabilité de b_2

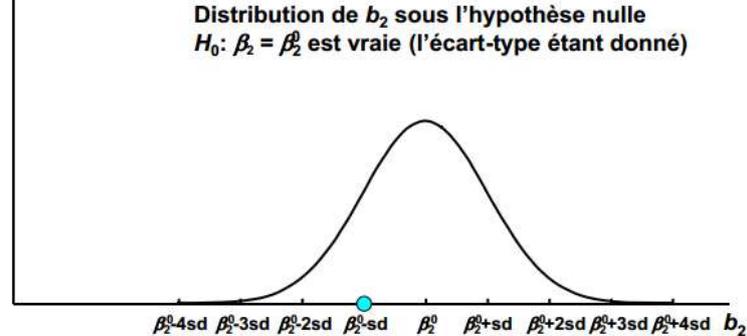


Supposons maintenant que nous avons un échantillon de données pour le modèle inflation des prix/salaire et que l'estimation de b_2 est 0.9. Ceci n'est pas une preuve contre H_0 . Nous ne **rejetons** pas H_0 . En terme général on dira que l'estimation est à **un écart-type en dessous** de la valeur d'hypothèse.

Fonction de densité de probabilité de b_2

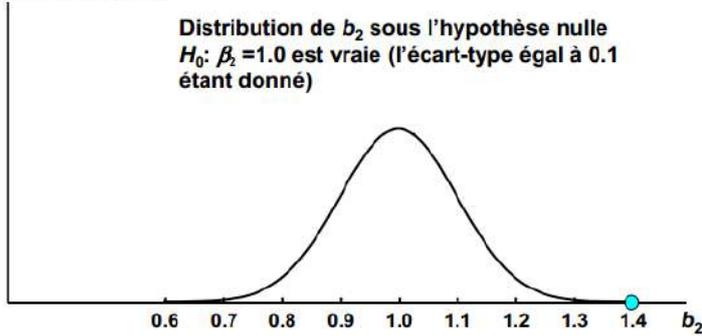


Fonction de densité de probabilité de b_2

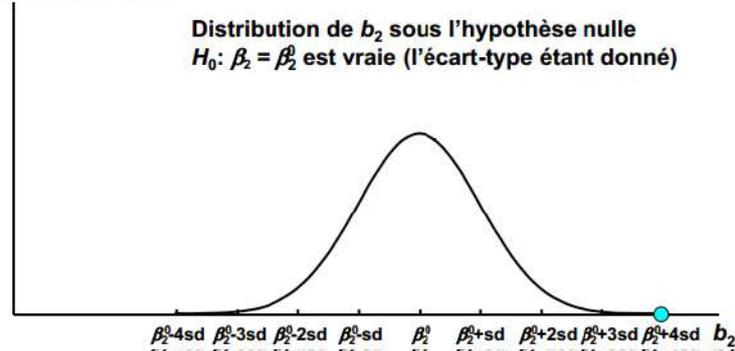


En revanche si nous obtenions une estimation de 1.4. Ceci va clairement à l'encontre de l'hypothèse nulle. Donc, **nous rejetterons** H_0 . En terme général on dira que l'estimation est un écart-type **au-dessus** de la valeur d'hypothèse.

Fonction de densité de probabilité de b_2



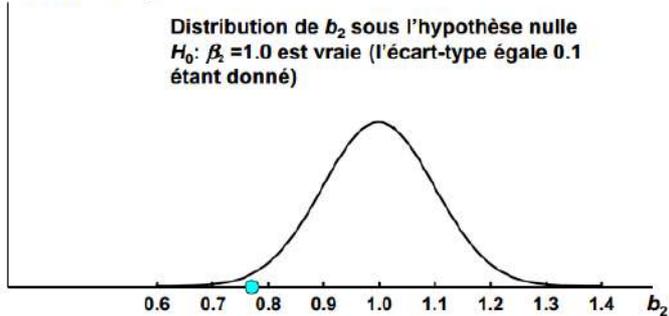
Fonction de densité de probabilité de b_2



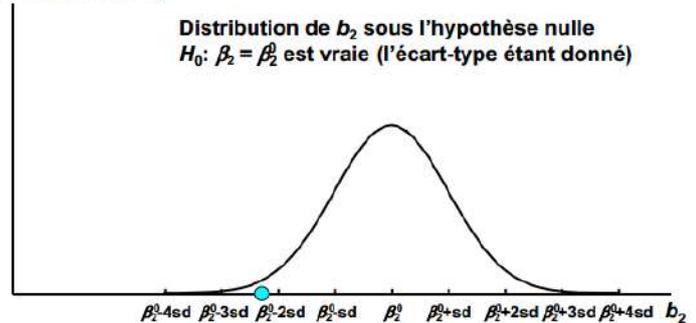
Partie 1 : économétrie

Enfin supposons que l'estimation nous donne 0.77. Ce résultat est ambigu et il y a deux possibilités. D'une part H_0 est vrai et nous avons une estimation étrange. D'autre part, cette dernière est simplement fausse.

Fonction de densité de probabilité de b_2

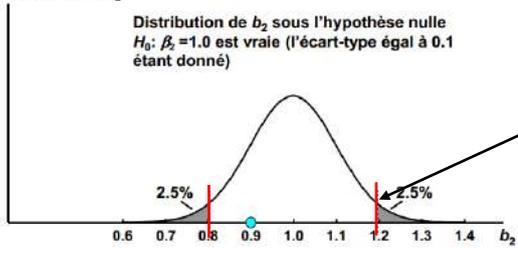


Fonction de densité de probabilité de b_2



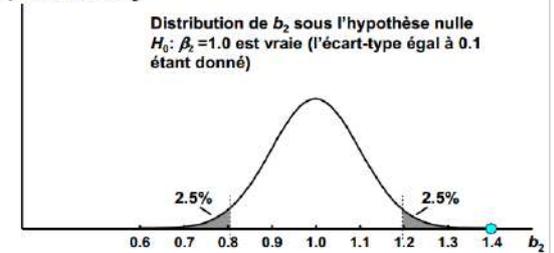
La procédure d'usage pour prendre une décision concernant H_0 est de fixer une borne à partir du moment où l'on rentre dans la zone de rejet. Par exemple 5%. D'après cette règle, nous rejeterions H_0 si l'estimation tombe respectivement dans le flanc des 2,5% les plus bas ou les plus hauts. Si nous appliquons cette règle:

Fonction de densité de probabilité de b_2

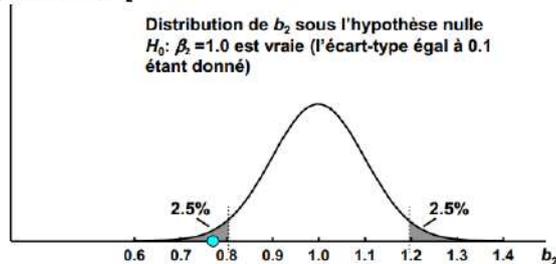


On fixe une borne à partir du moment où l'on rentre dans la zone de rejet.

Fonction de densité de probabilité de b_2

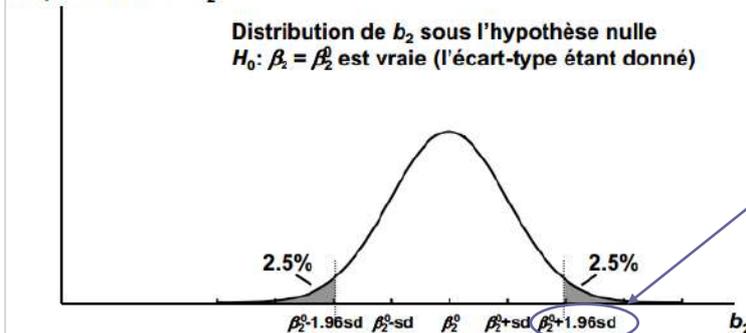


Fonction de densité de probabilité de b_2



En terme général la queue des 2.5% les plus éloignés du centre d'une distribution normale commence toujours à **1.96 écart types de sa moyenne**.

Fonction de densité de probabilité de b_2



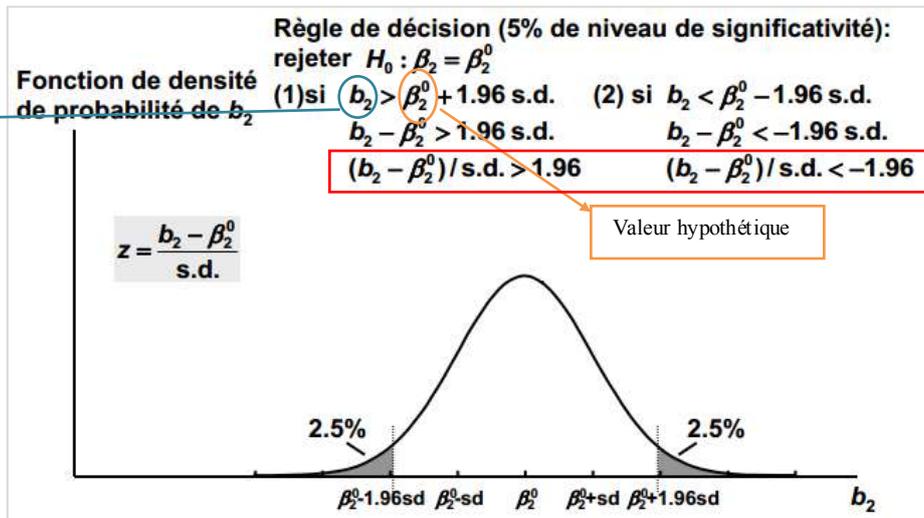
Toujours la valeur centrale + 1.96 écart-type

Partie 1 : économétrie

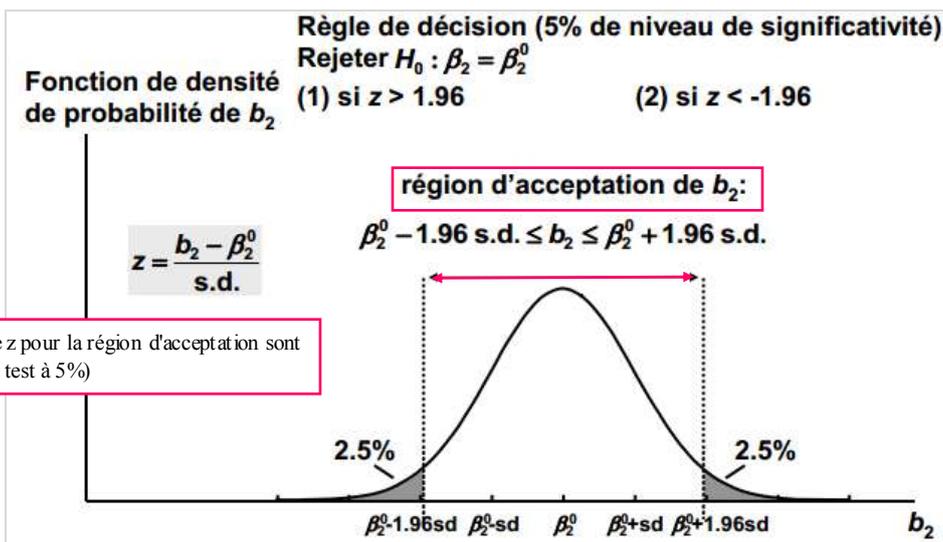
Nous rejetterions H_0 si la différence exprimée en termes d'écart-types, était de plus de 1.96 en termes absolus (positif ou négatif). Nous exprimons cette différence en termes d'écart types, z.

Valeur estimée

La règle est donc de rejeter H_0 si z est plus grand que 1.96 en termes absolus.



L'intervalle de valeurs de b_2 qui ne mène pas au rejet d' H_0 est appelée région d'acceptation.



NB: attention on n' "accepte" jamais une hypothèse nulle, on parle toujours en termes de REJET !!!

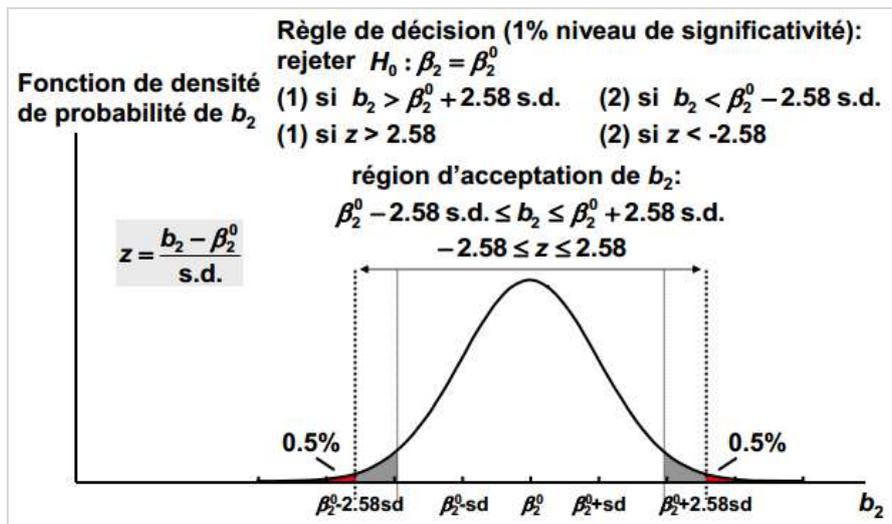
Partie 1 : économétrie

Erreur de Type I: rejet H_0 quand elle est vraie

Avec le test actuel, si l'hypothèse nulle est vraie, une erreur de Type I aura lieu 5 % du temps car 5 % du temps, nous obtiendrons des estimations dans les 2.5 % supérieurs ou inférieurs.

Nous pouvons réduire le risque de faire une erreur de Type I en réduisant l'étendue de la région de rejet. Par exemple nous pourrions fixer le rejet à la probabilité d'obtenir une estimation d'échantillon inférieur à 1 %.

La région de rejet devient le flan supérieur et inférieur à 0.5 %. Le flan des 0.5 % d'une distribution normale commence à 2.58 écart type de la moyenne, donc nous rejeterons désormais H_0 si z est plus grand que 2.58, en termes absolus.

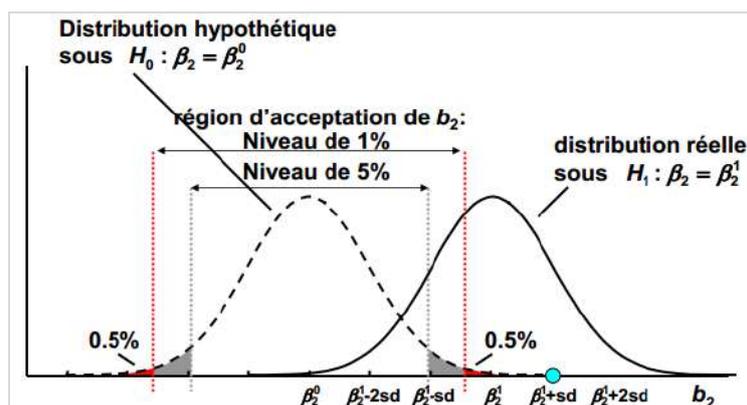


Puisque que la région d'acceptation est plus évasée que celui du test à 5 % le risque de faire une erreur de Type I si l' H_0 est vraie, est plus faible.

Erreur de Type II: on ne rejette pas l' H_0 alors qu'elle est fausse

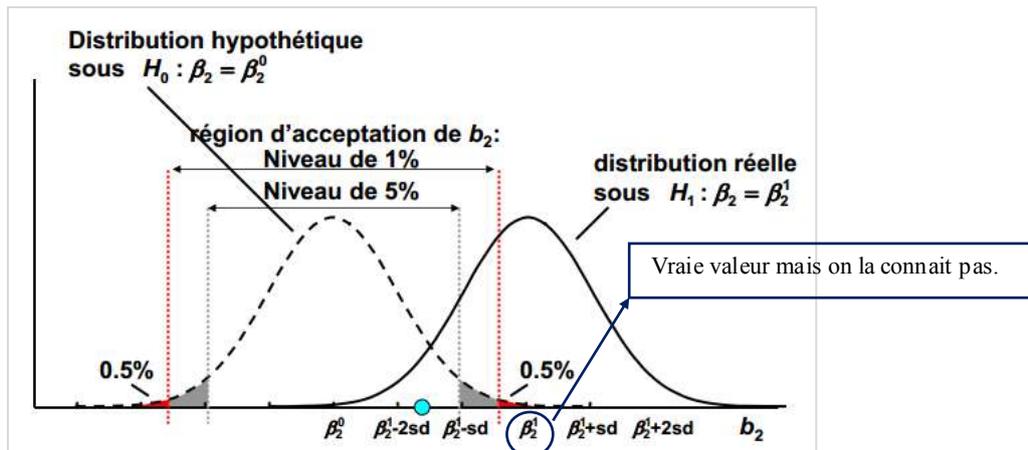
Cette séquence montre qu'il y a un trade-off entre le risque de faire une erreur de Type I et celui de faire une erreur de Type II. Nous allons analyser plusieurs cas de figure afin de mettre en lumière le trade off.

Dans ce cas, nous rejetons l'hypothèse nulle et dans ce cas nous prenons toujours la bonne décision peu importe quel test est utilisé (5 % ou 1 %).

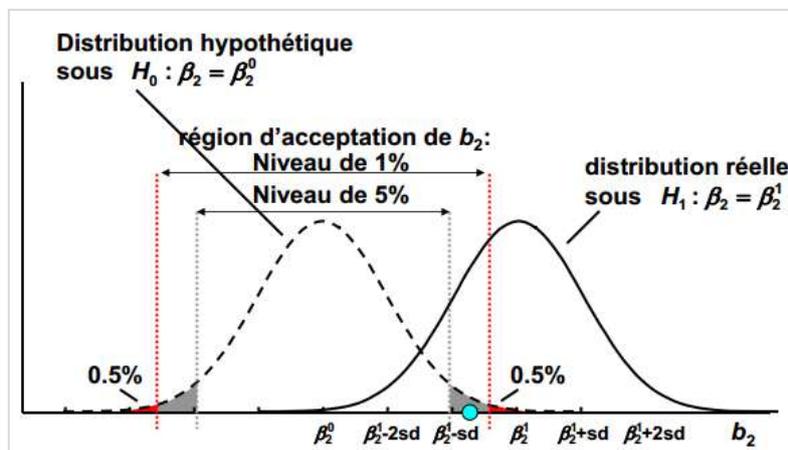


Partie 1 : économétrie

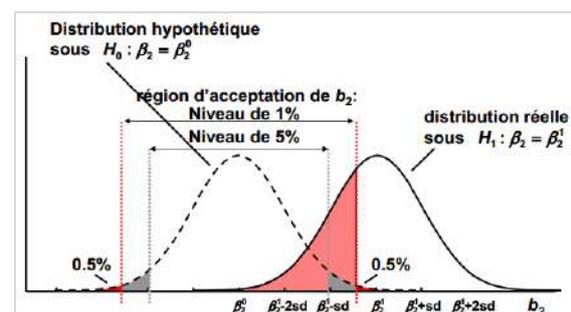
Dans ce cas, nous ferions une erreur de type II et manquerons de rejeter l'hypothèse nulle, peu importe quel test est utilisé.



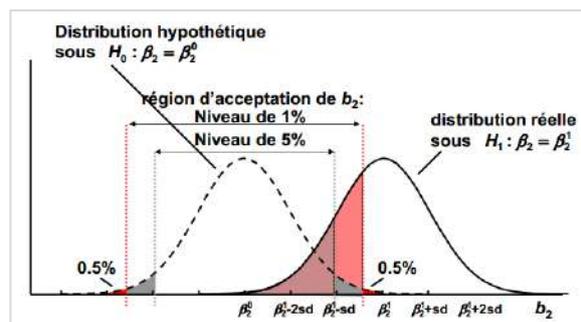
Enfin, dans ce cas, nous prendrions la bonne décision si nous avons utilisé un test à 5 % mais nous aurions commis une erreur de type II si nous avons utilisé un test à 1 %.



La probabilité de faire une erreur de Type II si nous utilisons un test à 1 % est donnée par la probabilité que b_2 se trouve dans la région d'acceptation à 1 % (aire en rose)⁵.



La probabilité de commettre une erreur de Type II (si nous utilisons un test à 5 %) est donnée par l'aire sous la distribution de H_1 dans la région d'acceptation pour le test à 5 % (aire en grise).



⁵ Voir slides pour plus d'exemples.

Partie 1 : économétrie

2. Test t d'une hypothèse portant sur le coefficient de régression

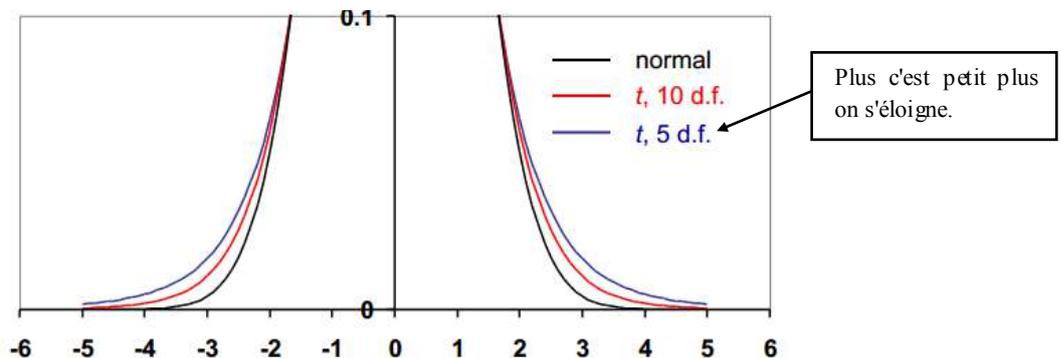
Jusqu'ici nous avons fait l'hypothèse que nous connaissions l'écart type (sd). Ceci est très irréaliste, nous devons l'estimer. Étant donné que nous avons remplacé s.d par s.e, le test statistique aura une distribution dite de t au lieu d'une distribution normale. La procédure est similaire que vu supra : si t est plus grande que sa valeur critique en valeur absolue alors on rejette H_0 si ce n'est pas le cas, nous ne la rejetons pas.

s.d. de b_2 connu	s.d. de b_2 non connu
<p>Divergence entre la valeur hypothétique et l'estimation d'échantillon, en termes de s.d.:</p> $z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$	<p>Divergence entre la valeur hypothétique et l'estimation d'échantillon, en termes de s.d.:</p> $t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$
<p>Test de significativité à 5%: Rejet de $H_0: \beta_2 = \beta_2^0$ si $z > 1.96$ ou $z < -1.96$</p>	<p>Test de significativité à 5%: Rejet de $H_0: \beta_2 = \beta_2^0$ si $t > t_{\text{crit}}$ ou $t < -t_{\text{crit}}$</p>

Écart-type estimé donc on aura non plus une distribution normale mais estimée. La différence est la valeur critique: elle est de 1.96 pour la normal mais pur l'estimée il faut regarder dans les tables.

Au plus le nombre de degré de liberté est élevé, au plus la distribution de t tendra vers la normale.

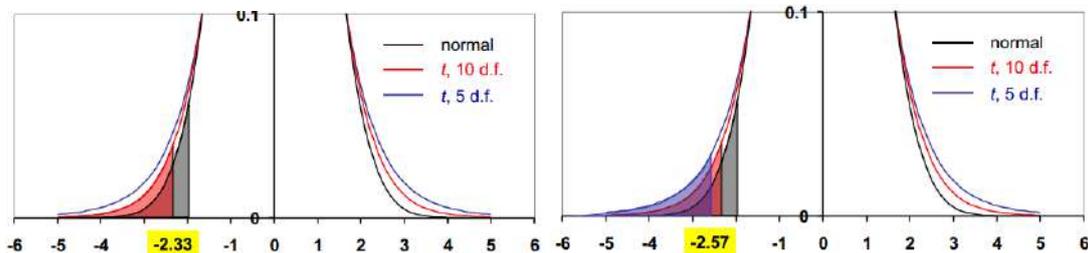
Notons que plus le nombre de degré de liberté devient grand, plus la valeur critique converge à 1.96^6 pour un niveau de significativité de 5 %. Cela est cohérent avec ce qui a été dit supra. *Dans ce cas, pourquoi s'acharner à prendre t si elle ressemble à une normale?* Tout simplement parce-que les queues de distribution de t sont plus longues que la normale, la différence étant plus grande plus le nombre de degré de liberté est faible.



⁶ C'est parce que la distribution t converge vers la distribution normale.

Partie 1 : économétrie

La queue de distribution à 2.5 % d'une distribution t avec 10 degré de liberté commence à 2.33 écart-types à partir de sa moyenne. Et commence à 2.57 écart-types pour une distribution avec 5 degré de liberté. C'est pour cette raison qu'il est nécessaire de se référer à une table de valeurs critiques de t quand on effectue un test de significativité.



Distribution t : valeurs critiques de t

Degrés de liberté	test bilatéral test unilatéral	10%	5%	2%	1%	0.2%	0.1%
		5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
...	
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
...	
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

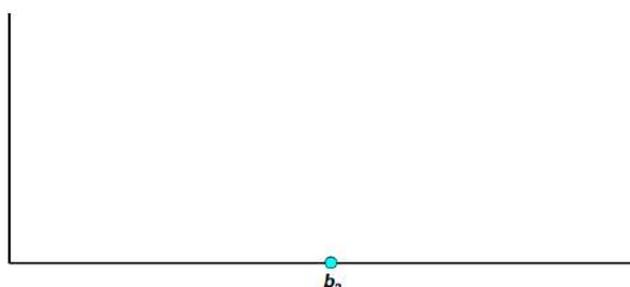
Exemple: 9 observations et 4 valeurs estimées (9 - 4 = 5 degré de libertés).

Le nombre de degré de liberté dans une régression = nombre d'observations - nombre de paramètres estimés. Dans une régression simple, nous estimons simplement deux paramètres (la constante et la pente), donc le nombre de degré de liberté est $n - 2$. (S'il y a n observations).

3. Intervalles de confiance

Dans la séquence sur les tests d'hypothèse, nous commençons avec une hypothèse donnée, par exemple: $H_0: \beta_2 = \beta_2^0$ et nous considérons si une estimation b_2 dérivée d'un échantillon mènerait/mènerait pas à son rejet.

Maintenant, nous faisons exactement l'opposé. Une estimation d'échantillon étant donnée, nous pouvons trouver l'ensemble des hypothèses qui ne seraient pas contredites par celle-ci.

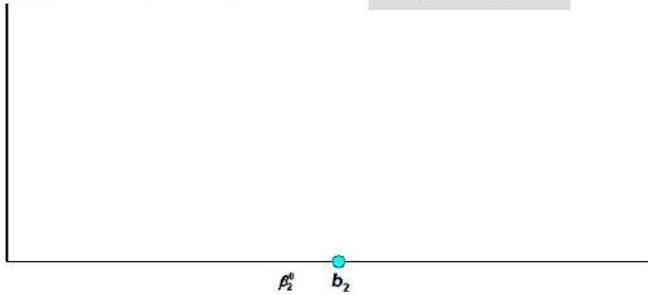


Partie 1 : économétrie

Supposons que quelqu'un vienne avec l'hypothèse $H_0: b_2 = b_2^0$. Pour voir si c'est compatible avec notre estimation d'échantillon b_2 , nous devons tracer la distribution de b_2 conditionnellement à ce que $H_0: \beta_2 = \beta_2^0$ soit vraie. Nous faisons l'hypothèse que nous connaissons l'écart type de la distribution. Après avoir dessiné la distribution, il est clair que H_0 n'est pas contredite par notre estimation d'échantillon.

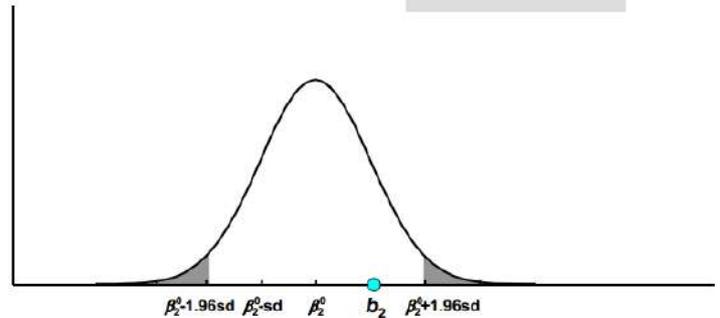
Fonction de densité de probabilité de b_2 conditionnelle à ce que $\beta_2 = \beta_2^0$ soit vrai

hypothèse nulle
 $H_0: \beta_2 = \beta_2^0$



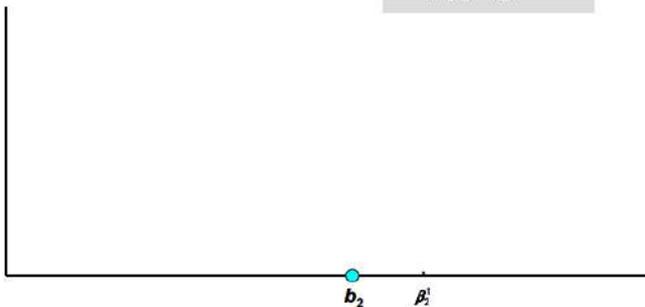
Fonction de densité de probabilité de b_2 conditionnelle à ce que $\beta_2 = \beta_2^0$ soit vrai

hypothèse nulle
 $H_0: \beta_2 = \beta_2^0$



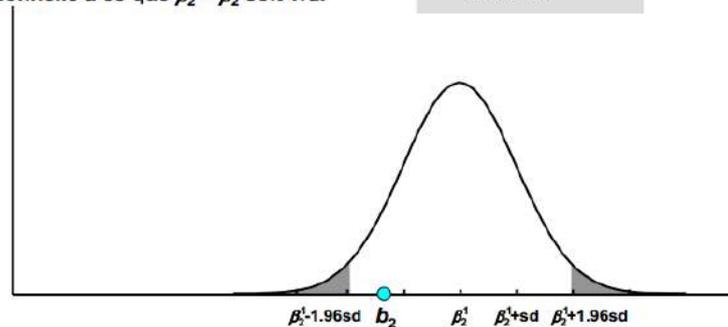
Prenons une autre hypothèse nulle: $H_0: \beta_2 = \beta_2^1$. Nous pouvons voir que cette hypothèse est compatible avec notre estimation.

hypothèse nulle
 $H_0: \beta_2 = \beta_2^1$



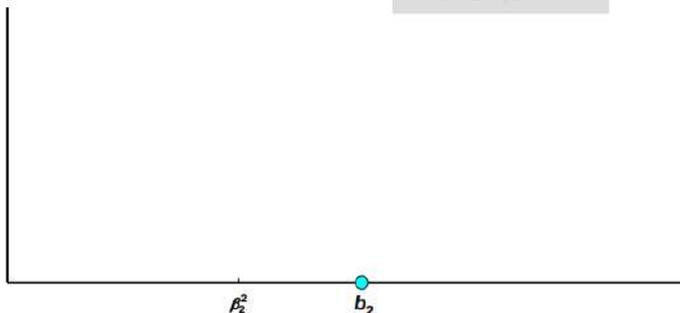
Fonction de densité de probabilité de b_2 conditionnelle à ce que $\beta_2 = \beta_2^1$ soit vrai

hypothèse nulle
 $H_0: \beta_2 = \beta_2^1$



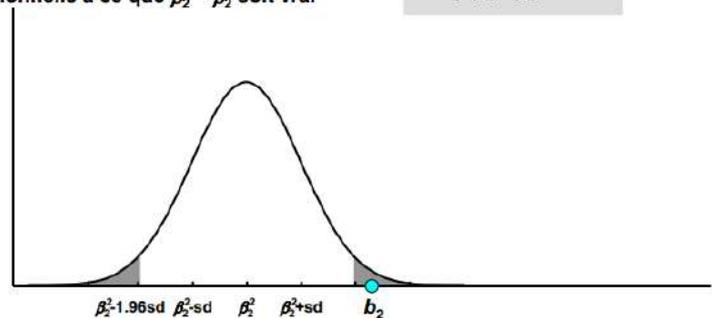
Prenons une autre hypothèse nulle: $H_0: \beta_2 = \beta_2^2$. Elle est incompatible avec notre estimation car l'estimation mènerait à son rejet.

hypothèse nulle
 $H_0: \beta_2 = \beta_2^2$



Fonction de densité de probabilité de b_2 conditionnelle à ce que $\beta_2 = \beta_2^2$ soit vrai

hypothèse nulle
 $H_0: \beta_2 = \beta_2^2$

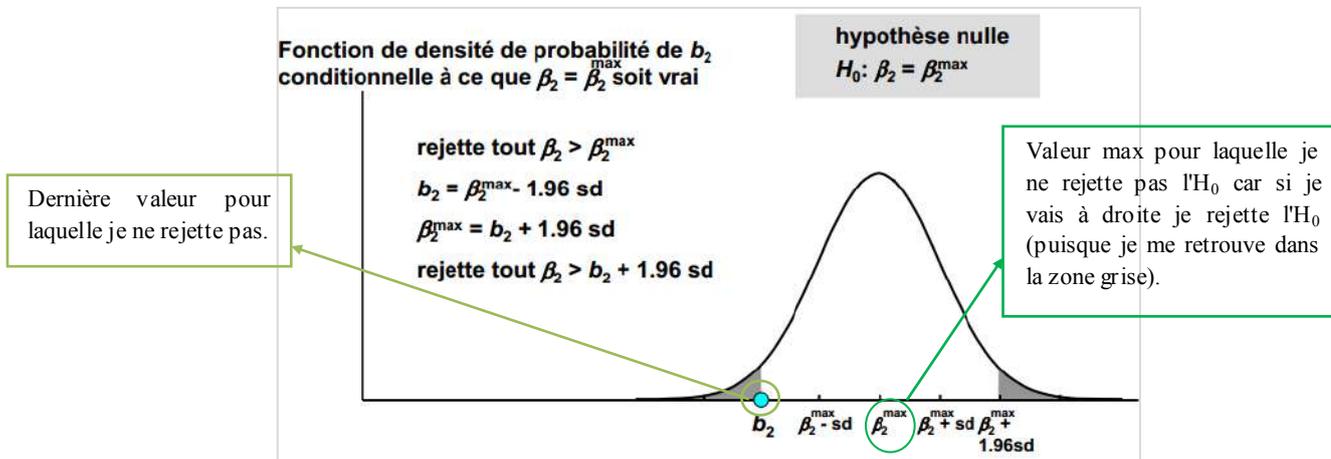


Nb: afin de voir si une variable est statistiquement significative, il suffira d'aller regarder si elle se trouve dans l'intervalle de confiance du tableau de régression calculé par STATA.

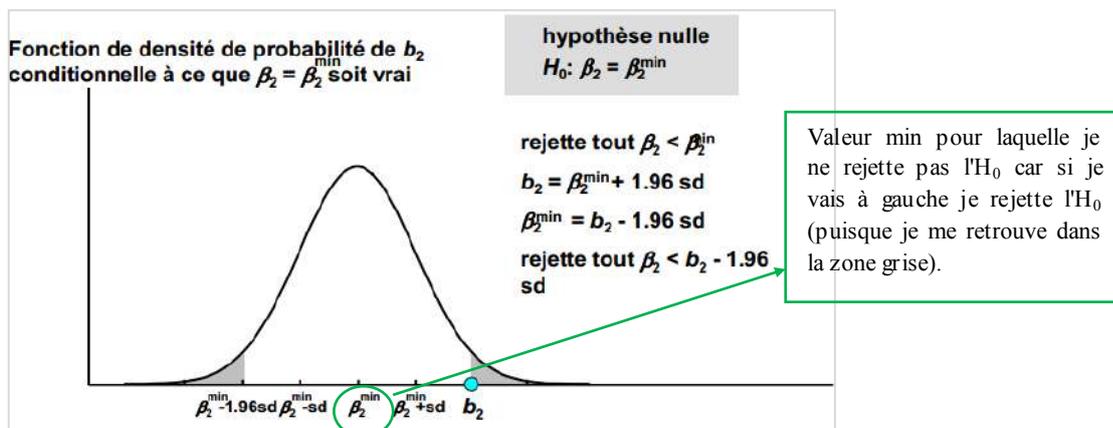
Partie 1 : économétrie

Quand la distribution de probabilité de distribution de b_2 , conditionnelle à l'hypothèse, est dessinée, elle implique que b_2 se trouve sur la limite gauche de la queue de distribution à 2.5 %. Puisque b_2 se trouve à la limite de la queue gauche à 2.5 %, il doit être à 1.96 écart-type de moins que β_2^{max} .

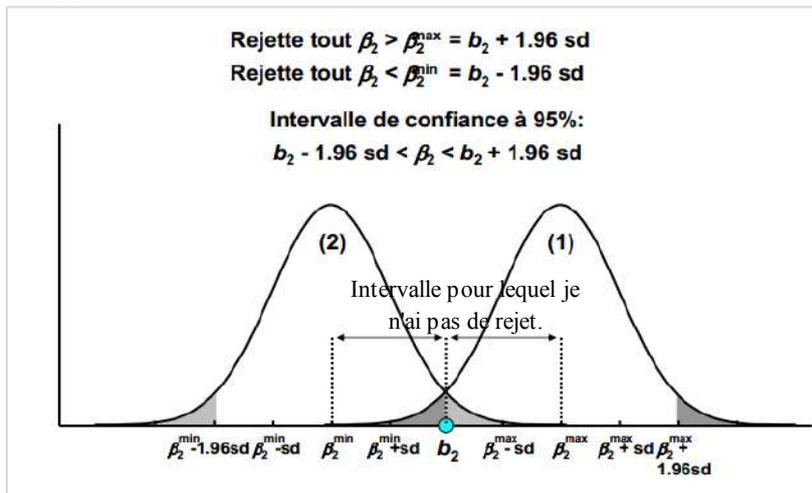
Comme nous connaissons b_2 et l'écart-type, nous pouvons calculer β_2^{max} .



De même, nous pouvons identifier la plus faible valeur hypothétique de β_2 . Ceci implique que b_2 se trouve sur la limite de la queue droite à 2.5 %. Nous l'appelons β_2^{min} . Donc β_2^{min} est égal à b_2 moins 1.96 écart-types.



Toute hypothèse se trouvant dans l'intervalle de β_2^{min} à β_2^{max} serait compatible avec l'estimation d'échantillon. Nous appelons cet intervalle, l'intervalle de confiance à 95 % (test à 5 %).



Partie 1 : économétrie

De la même manière, en utilisant un test à 1 % nous pouvons construire un intervalle de confiance à 99 %. β_2^{min} et β_2^{max} seront respectivement à 2.58 écart type à la gauche et à la droite de b_2 . Jusqu'à présent nous avons fait l'hypothèse que nous connaissons l'écart type de la distribution mais en pratique, nous devons l'estimer. Par conséquent, la distribution t doit être utilisée à la place de la distribution normale pour la localisation de β_2^{min} et β_2^{max} .

Ecart-type connu

intervalle de confiance à 95%

$$b_2 - 1.96 \text{ sd} \leq \beta_2 \leq b_2 + 1.96 \text{ sd}$$

intervalle de confiance à 99%

$$b_2 - 2.58 \text{ sd} \leq \beta_2 \leq b_2 + 2.58 \text{ sd}$$

Ecart-type estimé par l'écart-type de l'estimateur

intervalle de confiance à 95%

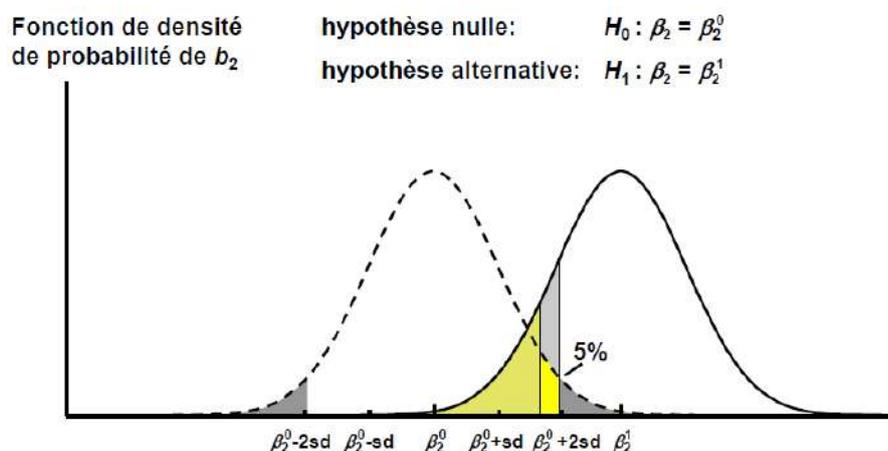
$$b_2 - t_{\text{crit}}(5\%) \text{ se} \leq \beta_2 \leq b_2 + t_{\text{crit}}(5\%) \text{ se}$$

intervalle de confiance à 99%

$$b_2 - t_{\text{crit}}(1\%) \text{ se} \leq \beta_2 \leq b_2 + t_{\text{crit}}(1\%) \text{ se}$$

TESTS UNILATERAUX

L'idée est la suivante: en utilisant un test unilatéral (nous mettons toutes la masse d'un côté), nous réduisons la probabilité de commettre une erreur de Type II⁷.



⁷ Voir slides pour tous les exemples.

Partie 1 : économétrie

4. Test F de qualité de l'estimation

Reprenons la définition du R^2 vu supra. L' H_0 que nous testons est que le modèle ne possède aucun pouvoir explicatif.

$$\text{Var}(Y) = \text{Var}(\hat{Y}) + \text{Var}(e)$$

$$\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum e^2$$

$$TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$Y = \beta_1 + \beta_2 X + u$$

$$H_0 : \beta_2 = 0, \quad H_1 : \beta_2 \neq 0$$

Les hypothèses concernant la qualité de l'estimation sont testées via la statistique de F. K est le nombre de paramètres et n le nombre d'observations. **La statistique F de Fisher** n'est rien d'autre qu'un ratio de deux χ^2 divisé par leur degré de liberté. Une χ^2 étant une somme de normales centrées réduites au carré. La F stat permet de faire un test sur plusieurs variables.

La statistique F peut être écrite alternativement en termes de R^2 .

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

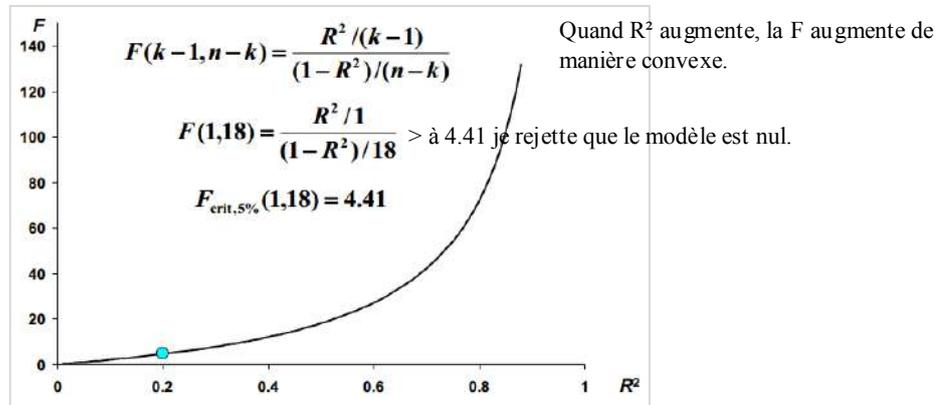
$$F(k-1, n-k) = \frac{\frac{ESS}{(k-1)}}{\frac{RSS}{(n-k)}} = \frac{\frac{ESS}{TSS}}{\frac{RSS}{TSS}} = \frac{R^2}{\frac{(1-R^2)}{(n-k)}}$$

$$\text{Nb: } \frac{RSS}{TSS} = \frac{TSS - ESS}{TSS} = 1 - \frac{ESS}{TSS} = 1 - R^2$$

F est une fonction croissante de R^2 . Si R^2 augmente, le numérateur augmente et le dénominateur diminue, donc F augmente.

Partie 1 : économétrie

Voici une F en fonction du R^2 pour le cas où il y a une variable explicative et 20 observations. Si nous faisons un test de significativité à 5 %, nous **rejetterions l' H_0 si la F stat est plus grande que la valeur critique**.



Notons que pour des analyses de régressions simples, l'hypothèse nulle et l'alternative sont mathématiquement exactement les mêmes que celles d'un test t bilatéral. Nous allons démontrer que pour l'analyse de régressions simples, **la F stat est le carré de la t statistique**.

Rappelons que $t = \frac{b_2 - \beta_2^0}{s.e.(b_2)} = \frac{b_2}{\sqrt{\frac{s_u^2}{n \text{Var}(X)}}}$

$$R^2 = \frac{\text{Var}(\hat{Y})}{\text{Var}(Y)} = \frac{\text{Var}(b_1 + b_2 X)}{\text{Var}(Y)} = \frac{\text{Var}(b_2 X)}{\text{Var}(Y)} = \frac{b_2^2 \text{Var}(X)}{\text{Var}(Y)}$$

$$R^2 = 1 - \frac{\text{Var}(e)}{\text{Var}(Y)} \Rightarrow 1 - R^2 = \frac{\text{Var}(e)}{\text{Var}(Y)} \quad s_u^2 = \frac{n}{n-2} \text{Var}(e)$$

$$F(k-1, n-k) = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)} = \frac{R^2}{(1-R^2) / (n-2)}$$

$$= \frac{\frac{b_2^2 \text{Var}(X)}{\text{Var}(Y)}}{\frac{\text{Var}(e)}{\text{Var}(Y)} / (n-2)} = \frac{b_2^2 \text{Var}(X)}{\frac{1}{n} \frac{n}{n-2} \text{Var}(e)} = \frac{b_2^2 \text{Var}(X)}{\frac{s_u^2}{n}} = \frac{b_2^2}{\frac{s_u^2}{n \text{Var}(X)}} = t^2$$

$$t = \left(\frac{b_2}{\text{s.e.}(b_2)} \right)^2 \quad \sigma_{b_2}^2 = \frac{\sigma_u^2}{n \text{Var}(X)}$$

La F stat est le carré de la t-stat mais cela n'est valable que pour $k=2$ (aussi non la démo n'est plus valable). De plus, la valeur critique sera le carré de l'autre.

Partie 1 : économétrie

L'exemple nous permet de vérifier que la valeur critique de F avec 18 degrés de liberté est égale au carré de la valeur critique de t, tel que la règle l'avait annoncé.

$$F_{\text{crit},5\%}(1,18) = 4.41$$

$$t_{\text{crit},5\%}(18) = 2.10$$

$$4.41 = 2.10^2$$

TEST F DE QUALITE DE L'ESTIMATION

. reg EARNINGS S					
Source	SS	df	MS		
Model	3977.38016	1	3977.38016		
Residual	34419.6569	568	60.5979875		
Total	38397.0371	569	67.4816117		

				Number of obs =	570
				F(1, 568) =	65.64
				Prob > F =	0.0000
				R-squared =	0.1036
				Adj R-squared =	0.1020
				Root MSE =	7.7845

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	1.073055	.1324501	8.102	0.000	.8129028	1.333206
_cons	-1.391004	1.820305	-0.764	0.445	-4.966354	2.184347

% des variations de Y expliqué par des variations de X.

t-stat est > à la valeur critique on rejette l'H₀.

P-valeur (l'air droite du point estimé) est > à 0.05 on ne rejette pas l'H₀.

$$F(1, n-2) = \frac{R^2}{(1-R^2)/(n-2)} = \frac{0.1036}{(1-0.1036)/(570-2)} = 65.65$$

Partie 1 : économétrie

Chapitre 3 : régression multiple avec deux variables explicatives - multicolinéarité - test F de qualité d'ajustement

A. RÉGRESSION MULTIPLE AVEC DEUX VARIABLES EXPLICATIVES

1. Régression multiple avec deux variables explicatives: exemple

La régression linéaire est appelée multiple lorsque **le modèle est composé d'au moins deux variables**. Comme il est excessivement rare, voire impossible, de prédire un phénomène à l'aide d'une seule variable, cette section porte sur la régression linéaire multiple.

Avec la régression multiple nous séparons les variables et de ce fait les effets. Les coefficients de régression sont calculés en utilisant le principe des RSS au même titre que dans les analyses de régression simple.

On passe à des plans: il y a plusieurs variables explicatives.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

On cherche la valeur de b_1 , b_2 et b_3 qui minimise la somme des carrés des résidus.

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i}$$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i}$$

*Démonstration*⁸

$$\begin{aligned} RSS &= \sum e_i^2 = \sum (Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i})^2 \\ &= \sum (Y_i^2 + b_1^2 + b_2^2 X_{2i}^2 + b_3^2 X_{3i}^2 - 2b_1 Y_i - 2b_2 X_{2i} Y_i \\ &\quad - 2b_3 X_{3i} Y_i + 2b_1 b_2 X_{2i} + 2b_1 b_3 X_{3i} + 2b_2 b_3 X_{2i} X_{3i}) \\ &= \sum Y_i^2 + n b_1^2 + b_2^2 \sum X_{2i}^2 + b_3^2 \sum X_{3i}^2 - 2b_1 \sum Y_i \\ &\quad - 2b_2 \sum X_{2i} Y_i - 2b_3 \sum X_{3i} Y_i + 2b_1 b_2 \sum X_{2i} \\ &\quad + 2b_1 b_3 \sum X_{3i} + 2b_2 b_3 \sum X_{2i} X_{3i} \end{aligned}$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \frac{\partial RSS}{\partial b_2} = 0 \quad \frac{\partial RSS}{\partial b_3} = 0$$

$$b_1 = \bar{Y} - b_2 \bar{X}_2 - b_3 \bar{X}_3$$

$$b_2 = \frac{\text{Cov}(X_2, Y) \text{Var}(X_3) - \text{Cov}(X_3, Y) \text{Cov}(X_2, X_3)}{\text{Var}(X_2) \text{Var}(X_3) - [\text{Cov}(X_2, X_3)]^2}$$

$$b_3 = \frac{\text{Cov}(X_3, Y) \text{Var}(X_2) - \text{Cov}(X_2, Y) \text{Cov}(X_2, X_3)}{\text{Var}(X_2) \text{Var}(X_3) - [\text{Cov}(X_2, X_3)]^2}$$

⁸ NB: lorsque le nombre de variables explicatives augment, il est plus aisé de passer au calcul matriciel.

Partie 1 : économétrie

STATA nous donne un aperçu d'une régression multiple mêlant salaire, éducation et intelligence.

Lorsqu'on lit une régression multiple il faut bien faire attention de lire bien séparément les coefficients des variables. Ainsi on dira que lorsque l'éducation augmente d'une année, le salaire augmente de 0.73\$ indépendamment de l'intelligence de l'individu.

```
. reg EARNINGS S ASVABC
```

Source	SS	df	MS
Model	4745.74965	2	2372.87483
Residual	33651.2874	567	59.3497133
Total	38397.0371	569	67.4816117

1 unité d'étude en plus rapporte 0.73 de plus quelque soit l'intelligence

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
S	.7390366	.1606216	4.601	0.000	.4235506 1.054523
ASVABC	.1545341	.0429486	3.598	0.000	.0701764 .2388918
_cons	-4.624749	2.0132	-2.297	0.022	-8.578989 -.6705095

1 unité d'intelligence en plus rapporte 0.15 quelque soit le niveau d'étude.

$$\hat{EARNINGS} = -4.62 + 0.74S + 0.15ASVABC$$

12 % des variations de salaire sont expliquées par des variations d'éducation et d'intelligence.

```
Number of obs = 570
F( 2, 567) = 39.98
Prob > F = 0.0000
R-squared = 0.1236
Adj R-squared = 0.1205
Root MSE = 7.7039
```

Utiliser le R ajusté avec le modèle de régression multiple c'est mieux.

Partie 1 : économétrie

2. Mettre 1 relation sous forme de graphe dans 1 modèle de régression multiple

Supposons que nous sommes particulièrement intéressés par le lien entre les salaires et l'éducation. Un scatter entre les deux variables donnerait une vue disproportionnée de la relation parce qu'en l'occurrence dans notre exemple l'intelligence est fortement positivement corrélée à l'éducation (logique). Ainsi lorsque S augmente, EARNINGS augmentera mais cette augmentation sera accentuée par une augmentation connexe de ASVABC.

Le but va être de purger EARNINGS et S de leurs composantes provenant de ASVABC. Afin de le faire nous allons prédire les résidus qui sont:

- La part de l'éducation qui n'est pas expliquée par l'intelligence. (ES)
- La part du revenu qui n'est pas expliquée par l'intelligence. (EEARN)

```
. reg S ASVABC
```

Si je ne mets pas une variable, elle va dans le terme d'erreur.

Source	SS	df	MS	Number of obs	=	570
Model	1153.80864	1	1153.80864	F(1, 568)	=	284.89
Residual	2300.43873	568	4.05006818	Prob > F	=	0.0000
Total	3454.24737	569	6.07073351	R-squared	=	0.3340
				Adj R-squared	=	0.3329
				Root MSE	=	2.0125

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ASVABC	.1545378	.0091559	16.879	0.000	.1365543 .1725213
_cons	5.770845	.4668473	12.361	0.000	4.853888 6.687803

```
. predict ES, resid
```

Si je régresse l'éducation sur l'intelligence, la partie de l'erreur est: la part d'éducation expliquée par tout sauf par l'intelligence.

Partie de l'éducation non expliquée par l'intelligence.

```
. reg EARNINGS ASVABC
```

Source	SS	df	MS	Number of obs	=	570
Model	3489.30726	1	3489.30726	F(1, 568)	=	56.78
Residual	34907.7298	568	61.4572708	Prob > F	=	0.0000
Total	38397.0371	569	67.4816117	R-squared	=	0.0909
				Adj R-squared	=	0.0893
				Root MSE	=	7.8395

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ASVABC	.2687432	.035666	7.535	0.000	.1986898 .3387966
_cons	-.359883	1.818571	-0.198	0.843	-3.931829 3.212063

```
. predict EEARN, resid
```

Partie du revenu non expliquée par l'intelligence.

Partie 1 : économétrie

Ainsi nous obtiendrons la part respectivement des salaires et de l'éducation qui n'est pas expliquée par l'intelligence.

Il nous suffira enfin de régresser EARN sur ES pour avoir la relation. On peut voir l'estimation du coefficient de pente, son écart-type ainsi que sa t-statistique sont les mêmes qu'en régression multiple.

. reg EARN ES					
Source	SS	df	MS		
Model	1256.44239	1	1256.44239	Number of obs =	570
Residual	33651.2873	568	59.2452241	F(1, 568) =	21.21
Total	34907.7297	569	61.3492613	Prob > F =	0.0000
				R-squared =	0.0360
				Adj R-squared =	0.0343
				Root MSE =	7.6971

EARN	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ES	.7390366	.1604802	4.605	0.000	.4238296 1.054244
_cons	-5.99e-09	.3223957	0.000	1.000	-.6332333 .6332333

On retrouve le même résultat. Il est plus pratique d'ajouter des paramètres.

De la régression multiple:

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
s	.7390366	.1606216	4.601	0.000	.4235506 1.054523
ASVABC	.1545341	.0429486	3.598	0.000	.0701764 .2388918
_cons	-4.624749	2.0132	-2.297	0.022	-8.578989 -.6705095

B. MULTICOLLINÉARITÉ

1. Multicollinéarité

Que se passerait-il si nous essayons de faire une régression lorsqu'une variable aléatoire est une combinaison linéaire parfaite de une ou plusieurs variables.

Démonstration

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \quad X_3 = \lambda + \mu X_2$$

Variable indépendante (celle que je veux expliquée).

$$b_2 = \frac{\text{Cov}(X_2, Y) \text{Var}(\mu X_2) - \text{Cov}(\mu X_2, Y) \text{Cov}(X_2, \mu X_2)}{\text{Var}(X_2) \text{Var}(\mu X_2) - [\text{Cov}(X_2, \mu X_2)]^2}$$

$$= \frac{\text{Cov}(X_2, Y) \mu^2 \text{Var}(X_2) - \mu \text{Cov}(X_2, Y) \mu \text{Cov}(X_2, X_2)}{\text{Var}(X_2) \mu^2 \text{Var}(X_2) - [\mu \text{Cov}(X_2, X_2)]^2}$$

$$= \frac{\mu^2 \text{Cov}(X_2, Y) \text{Var}(X_2) - \mu^2 \text{Cov}(X_2, Y) \text{Var}(X_2)}{\mu^2 \text{Var}(X_2) \text{Var}(X_2) - [\mu \text{Var}(X_2)]^2} = \frac{0}{0}$$

Si une variable aléatoire est une combinaison linéaire de 1 ou plusieurs, il n'y a pas de solution. On ne SAIT PAS ESTIMER LE MODELE.

Partie 1 : économétrie

On ne sait pas estimer le modèle mais cela est très rare dans la vraie vie. Il arrive souvent qu'il y ait une relation approximative entre des variables.

```
. reg EARNINGS S ASVABC ASVAB5
```

Source	SS	df	MS			
Model	4909.11468	3	1636.37156	Number of obs =	570	
Residual	33487.9224	566	59.1659406	F(3, 566) =	27.66	
Total	38397.0371	569	67.4816117	Prob > F =	0.0000	
				R-squared =	0.1279	
				Adj R-squared =	0.1232	
				Root MSE =	7.6919	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	.7115506	.1612235	4.413	0.000	.3948811	1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219	.2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319	.1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034	-1.699616

Le principal problème de la multicolinéarité est qu'elle fausse notre inférence en **gonflant** de manière artificielle les **écarts types** des coefficients, en **élargissant les intervalles de confiance** et en **diminuant la t stat** ce qui affaiblit la robustesse des résultats. Une manière de la détecter est d'effectuer la corrélation entre 2 variables "suspectes" (ici ASVABC et ASVAB5: 0.64).

```
. reg EARNINGS S ASVABC ASVAB5
```

Source	SS	df	MS			
Model	4909.11468	3	1636.37156	Number of obs =	570	
Residual	33487.9224	566	59.1659406	F(3, 566) =	27.66	
Total	38397.0371	569	67.4816117	Prob > F =	0.0000	
				R-squared =	0.1279	
				Adj R-squared =	0.1232	
				Root MSE =	7.6919	

EARNINGS	Coef.	Std. Err.	t	P>	. cor ASVABC ASVAB5 (obs=570)	
S	.7115506	.1612235	4.413	0.		
ASVABC	.1104595	.0504223	2.191	0.	ASVABC	ASVAB5
ASVAB5	.0770794	.0463868	1.662	0.	ASVABC	1.0000
_cons	-5.944977	2.161409	-2.751	0.	ASVAB5	0.6371 1.0000

De plus cette dernière peut être détectée **via un R² élevé** mais avec **peu de variables explicatives** et une **t qui diminue**.

Précision, la multicollinéarité ne biaise pas les coefficients de régression, elles restent centrées autour de la vraie valeur mais elles ont des écart-types (des variances) trop larges. Et nous préférons un écart-type plus faible car on est plus précis (on se rapproche de la vraie valeur).

Partie 1 : économétrie

a) Mesures possibles pour réduire la multicollinearité

Afin d'analyser ces méthodes, il est nécessaire de reprendre la formule de la variance de population de b_2 (r étant le coefficient de corrélation):

$$\text{Variance de population de } b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\text{Var}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(1) Réduire la variance du terme d'erreur (σ_u^2) en incluant plus de variables (pertinentes) dans le modèle

```
. reg EARNINGS S ASVABC ASVAB5
```

Source	SS	df	MS	
Model	4909.11468	3	1636.37156	Number of obs = 570
Residual	33487.9224	566	59.1659406	F(3, 566) = 27.66
Total	38397.0371	569	67.4816117	Prob > F = 0.0000
				R-squared = 0.1279
				Adj R-squared = 0.1232
				Root MSE = 7.6919

```
. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN
```

Source	SS	df	MS	
Model	7715.87322	6	1285.97887	Number of obs = 570
Residual	30681.1638	563	54.4958505	F(6, 563) = 23.60
Total	38397.0371	569	67.4816117	Prob > F = 0.0000
				R-squared = 0.2009
				Adj R-squared = 0.1924
				Root MSE = 7.3821

```
. reg EARNINGS S ASVABC ASVAB5
```

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
S	.7115506	.1612235	4.413	0.000	.3948811 1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219 .2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319 .1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034 -1.699616

```
. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN
```

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
S	.8137184	.1563975	5.203	0.000	.5065245 1.120912
ASVABC	.0442801	.049716	0.891	0.373	-.0533714 .1419317
ASVAB5	.1113769	.0458757	2.428	0.016	.0212685 .2014853
TENURE	.287038	.0676471	4.243	0.000	.1541665 .4199095
MALE	3.123929	.64685	4.829	0.000	1.853395 4.394463
URBAN	2.061867	.7274286	2.834	0.005	.6330618 3.490672
_cons	-10.60023	2.195757	-4.828	0.000	-14.91311 -6.287358

Partie 1 : économétrie

(2) Augmenter le nombre d'observations

En augmentant le nombre d'observations on réduit l'écart type et on augmente la t stat. Cependant, dans le cas de ASVAB5, la t-statistique est plus grande malgré le fait que le coefficient est plus petit.

```
. reg EARNINGS S ASVABC ASVAB5
```

Number of obs = 570

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	.7115506	.1612235	4.413	0.000	.3948811	1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219	.2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319	.1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034	-1.699616


```
. reg EARNINGS S ASVABC ASVAB5
```

Number of obs = 2868

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	1.002693	.0787447	12.733	0.000	.8482905	1.157095
ASVABC	.1448345	.0241135	6.006	0.000	.097553	.1921161
ASVAB5	.0483846	.0218352	2.216	0.027	.0055703	.091199
_cons	-9.654593	1.033311	-9.343	0.000	-11.6807	-7.628485

(3) Augmenter Var(X_2)

Augmenter la Var(X_2) en prenant des individus extrêmes (riches, pauvres, etc.). De ce fait, la variance sera plus grande et l'écart-type sera plus faible.

(4) Diminuer r_{X_2, X_3}^2

Une autre possibilité est de réduire la corrélation entre les variables explicatives. C'est loin d'être facile.

(5) Combiner les variables corrélées (collinéaires)

Combiner les variables en créant un indice composite.

C'est ce qui a été fait avec les 3 variables. ASVABC a été calculé comme une moyenne pondérée de ASVAB2, ASVAB3 et ASVAB4. L'écart type d'estimateur de ASVABC est plus petit que ceux de ses composantes dans la seconde régression.

```
. reg EARNINGS S ASVABC
```

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	.7390366	.1606216	4.601	0.000	.4235506	1.054523
ASVABC	.1545341	.0429486	3.598	0.000	.0701764	.2388918
_cons	-4.624749	2.0132	-2.297	0.022	-8.578989	-.6705095

Créer un indice composite des 3 formes d'intelligence en regroupant les variables colinéaires. L'écart-types diminue.

```
. reg EARNINGS S ASVAB2 ASVAB3 ASVAB4
```

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	.7362439	.1586812	4.640	0.000	.4245668	1.047921
ASVAB2	.2472668	.0472249	5.236	0.000	.154509	.3400246
ASVAB3	.0137422	.058716	0.234	0.815	-.1015861	.1290705
ASVAB4	-.1051868	.0544682	-1.931	0.054	-.2121716	.001798
_cons	-4.734303	2.06706	-2.290	0.022	-8.794363	-.6742428

Partie 1 : économétrie

(6) Enlever certaines variables corrélées

Nous pouvons remarquer que l'écart-type de l'estimateur est plus petit que dans la régression qui inclut ASVAB3 et ASVAB4. Cependant, cette approche est dangereuse car certaines variables enlevées pourraient réellement appartenir au modèle et leur omission pourrait causer un biais de variables omises.

```
. reg EARNINGS S ASVAB2
```

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	.6449415	.1519755	4.244	0.000	.3464378	.9434452
ASVAB2	.2019724	.0376567	5.364	0.000	.1280086	.2759361
_cons	-5.796398	1.957987	-2.960	0.003	-9.642191	-1.950605


```
. reg EARNINGS S ASVAB2 ASVAB3 ASVAB4
```

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	.7362439	.1586812	4.640	0.000	.4245668	1.047921
ASVAB2	.2472668	.0472249	5.236	0.000	.154509	.3400246
ASVAB3	.0137422	.058716	0.234	0.815	-.1015861	.1290705
ASVAB4	-.1051868	.0544682	-1.931	0.054	-.2121716	.001798
_cons	-4.734303	2.06706	-2.290	0.022	-8.794363	-.6742428

(7) Restrictions empiriques

Utiliser de l'information étrangère, si disponible, concernant le coefficient d'une des variables.

$$Y = \beta_1 + \beta_2 X + \beta_3 P + u \quad Y' = \beta'_1 + \beta'_2 X' + u$$

$$\hat{Y}' = b'_1 + b'_2 X'$$

$$Y = \beta_1 + b'_2 X + \beta_3 P + u$$

$$Z = Y - b'_2 X = \beta_1 + \beta_3 P + u$$

(8) Restriction théorique

La restriction théorique est définie comme une relation hypothétique parmi les paramètres d'un modèle de régression.

Illustrons ce point par un modèle d'accès aux études. Partons de l'hypothèse que le plus haut grade accompli, S, dépend de ASVABC, et des grades accomplis par la mère et le père, SM et SF, respectivement.

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + u$$

Partie 1 : économétrie

Cependant, les couples se forment souvent entre personnes au niveau d'éducation similaire, la corrélation entre SM et SF est élevée. La régression souffre de multicollinéarité.

```
. reg S ASVABC SM SF
```

Source	SS	df	MS	Number of obs = 570		
Model	1278.24153	3	426.080508	F(3, 566)	=	110.83
Residual	2176.00584	566	3.84453329	Prob > F	=	0.0000
Total	3454.24737	569	6.07073351	R-squared	=	0.3700
				Adj R-squared	=	0.3667
				Root MSE	=	1.9607

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	.1295006	.0099544	13.009	0.000	.1099486	.1490527
SM	.069403	.0422974	1.641	0.101	-.013676	.152482
SF	.1102684	.0311948	3.535	0.000	.0489967	.1715401
_cons	4.914654	.5063527	9.706	0.000	3.920094	5.909214


```
. cor SM SF
(obs=570)
```

	SM	SF
SM	1.0000	
SF	0.6391	1.0000

Dans ce cas, nous posons l'hypothèse que l'éducation de chacun des parents a la même importance pour celle de leur(s) enfant(s). Nous pouvons imposer la restriction telle que $\beta_3 = \beta_4$

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + u$$

$$\beta_3 = \beta_4 \quad \text{On regroupe les deux.}$$

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 (SM + SF) + u$$

$$= \beta_1 + \beta_2 ASVABC + \beta_3 SP + u$$

L'écart type de l'estimateur de SP est plus petit que celui de SM et SF. L'utilisation de la restriction a mené à un gain d'efficacité et le problème de multicollinéarité a été éliminé. De plus la t est très élevée.

```
. g SP=SM+SF
```

```
. reg S ASVABC SP
```

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	.1295653	.0099485	13.024	0.000	.1100249	.1491057
SP	.093741	.0165688	5.658	0.000	.0611973	.1262847
_cons	4.823123	.4844829	9.955	0.000	3.871523	5.774724

```
. reg S ASVABC SM SF
```

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	.1295006	.0099544	13.009	0.000	.1099486	.1490527
SM	.069403	.0422974	1.641	0.101	-.013676	.152482
SF	.1102684	.0311948	3.535	0.000	.0489967	.1715401
_cons	4.914654	.5063527	9.706	0.000	3.920094	5.909214

Écart-types diminuent.

La restriction théorique n'est pas idiote car l'intervalle de confiance se chevauche.

Partie 1 : économétrie

C. F-TESTS DE QUALITÉ D'AJUSTEMENT

1. Qualité d'ajustement de l'équation entière

Considérons le cas général où il y a $k-1$ variables explicatives. Pour le test F de qualité d'ajustement de l'équation entière, l'hypothèse nulle, en un mot, est que le modèle n'a pas de pouvoir explicatif du tout (aucunes variables n'expliquent Y). L'hypothèse nulle est donc que tous les coefficients de β sont nuls. L'alternative possible est qu'au moins un de ces coefficients β est différent de 0.

Ainsi les tests F testent le pouvoir explicatif joint des variables, tandis que les tests t testent leur pouvoir explicatif individuellement. Illustration du test pour une régression multiple.

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$$

$$H_0 : \beta_2 = \dots = \beta_k = 0 \quad \text{Aucune variable n'explique Y.}$$

$$\text{Au moins une variable explique quelque chose.} \quad H_1 : \text{au moins un } \beta \neq 0$$

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)}$$

$$= \frac{ESS}{TSS} \bigg/ \frac{(k-1)}{(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

Reprenons notre exemple de réussite scolaire. Nous voyons que la F dépasse largement la valeur critique donc nous pouvons rejeter l'hypothèse nulle et affirmer que notre modèle n'est pas nul. Attention si on a une bonne stat F mais toutes des t très basse cela veut probablement dire qu'il y a de la multicolinéarité dans le modèle.

```
. reg S ASVABC SM SF
```

Source	SS	df	MS	Number of obs = 570		
Model	1278.24153	3	426.080508	F(3, 566)	=	110.83
Residual	2176.00584	566	3.84453329	Prob > F	=	0.0000
Total	3454.24737	569	6.07073351	R-squared	=	0.3700
				Adj R-squared	=	0.3667
				Root MSE	=	1.9607

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	.1295006	.0099544	13.009	0.000	.1099486	.1490527
SM	.069403	.0422974	1.641	0.101	-.013676	.152482
SF	.1102684	.0311948	3.535	0.000	.0489967	.1715401
_cons	4.914654	.5063527	9.706	0.000	3.920094	5.909214

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} \quad F(3,566) = \frac{1278/3}{2176/566} = 110.8$$

$$F_{\text{crit},0.1\%}(3,120) = 5.78 \quad 110.8 > 5.78 \text{ je rejette l}'H_0.$$

Partie 1 : économétrie

2. Test de pouvoir explicatif joint d'un groupe de variables quand elles sont ajoutées à un modèle de régression

L'hypothèse nulle du test de F est que ni X_3 ni X_4 n'appartiennent au modèle. L'hypothèse alternative est qu'au moins une des variables y appartient, peut-être les 2.

Par exemple, dans la spécification originale, Y pourrait être écrit comme une fonction simple de X_2 . Dans la seconde, nous ajoutons X_3 et X_4 comme variables explicatives.

$$\begin{array}{ll} Y = \beta_1 + \beta_2 X_2 + u & RSS_1 \\ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u & RSS_2 \end{array}$$

L'un est un modèle restreint de l'autre.

$$\begin{array}{l} H_0 : \beta_3 = \beta_4 = 0 \\ H_1 : \beta_3 \neq 0 \text{ ou } \beta_4 \neq 0 \end{array}$$

Reconsidérons la structure de F pour un tel test et parlons en termes de coût/amélioration.

$$F(\text{coût, d.l. dispo.}) = \frac{\text{amélioration} / \text{coût}}{\text{Partie résiduelle non expliquée} / \text{Degrés de liberté disponibles}}$$

Nombre de paramètre estimé en plus.

L'amélioration est la réduction de la somme des carrés résiduelle une fois le changement fait, dans ce cas, lorsque le groupe de nouvelles variables est ajouté.

Le coût est la réduction du nombre de degré de liberté restant après avoir fait le changement. Dans ce cas-ci, la perte est égale au nombre de nouvelles variables ajoutées.

Nombre de degré de liberté est le nombre d'observations moins le nombre de paramètres estimés.

La partie résiduelle non expliquée est la somme des carrés résiduels après avoir fait le changement.

Partie 1 : économétrie

Dans notre exemple, après avoir rajouté deux variables nous nous apercevons que le test de F nous dit que nous pouvons rejeter l'hypothèse nulle.

. reg S ASVABC				Ce que je n'explique pas dans le modèle	
Source	SS	df	MS		
Model	1153.80864	1	1153.80864	Number of obs =	570
Residual	2300.43873	568	4.05006818	F(1, 568) =	284.89
Total	3454.24737	569	6.07073351	Prob > F =	0.0000
				R-squared =	0.3340
				Adj R-squared =	0.3329
				Root MSE =	2.0125

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ASVABC	.1545378	.0091559	16.879	0.000	.1365543 .1725213
_cons	5.770845	.4668473	12.361	0.000	4.853888 6.687803

. reg S ASVABC SM SF				Ce que je n'explique pas dans le modèle	
Source	SS	df	MS		
Model	1278.24153	3	426.080508	Number of obs =	570
Residual	2176.00584	566	3.84453329	F(3, 566) =	110.83
Total	3454.24737	569	6.07073351	Prob > F =	0.0000
				R-squared =	0.3700
				Adj R-squared =	0.3667
				Root MSE =	1.9607

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ASVABC	.1295006	.0099544	13.009	0.000	.1099486 .1490527
SM	.069403	.0422974	1.641	0.101	-.013676 .152482
SF	.1102684	.0311948	3.535	0.000	.0489967 .1715401
_cons	4.914654	.5063527	9.706	0.000	3.920094 5.909214

Je réduits la partie inconnue.

$$F(2, 570 - 4) = \frac{(RSS_1 - RSS_2) / 2}{RSS_2 / (570 - 4)} = \frac{(2300.4 - 2176.0) / 2}{2176.0 / 566} = 16.18$$

Je rejette l'H0 (SM=SF) car $16.18 > 7.32$ $F_{crit, 0.1\%}(2, 120) = 7.32$

3. F-test marginaux équivalents au t-tests pour une seule variable

Nous avons montré par exemple⁹ que la F-statistiques d'une seule variable est égal au carré de la t-statistique de cette variable. Il peut également être montré que la valeur critique de la F doit être égale au carré de la valeur critique de la t.

⁹ Voir exemple dans les slides.

Partie 1 : économétrie

Chapitre 4: non linéarité et variables dummy (dichotomiques)

A. LINÉARITÉ ET NON LINÉARITÉ

1. Définition

Bonne spécification: il faut que le modèle que l'on estime soit le bon modèle. On supposait que le modèle était linéaire. Le premier modèle est linéaire dans les 2 sens : dans les variables et paramètres. Le second, lui, est non linéaire dans les variables. Enfin concernant le dernier, il est non linéaire dans les paramètres, le coefficient X_4 étant le produit des coefficients de X_2 et X_3 .

Un paramètre différent apparaît dans chaque terme.

Linéaire dans les variables et les paramètres:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

Les variables ne sont pas sous la forme de fonction.

Linéaire dans les paramètres, non linéaire dans les variables:

$$Y = \beta_1 + \beta_2 X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log X_4 + u$$

On transforme les variables en créant une nouvelle variable et on obtient un modèle linéaire (dans les variables et les paramètres).

$$Z_2 = X_2^2, \quad Z_3 = \sqrt{X_3}, \quad Z_4 = \log X_4$$

$$Y = \beta_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + u$$

Nonlinear in parameters:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_2 \beta_3 X_4 + u$$

Non linéaire dans les paramètres. Ce type de modèle peut être linéarisé par des transformations appropriées.

Commençons par un exemple : données sur la consommation annuelle de bananes et le revenu annuel pour un échantillon de 10 ménages. C'est un modèle qui est non linéaire dans les variables mais qui peut être linéarisé par transformation.

household	bananas (lbs) Y	income (\$10,000) X
1	1.71	1
2	6.88	2
3	8.25	3
4	9.52	4
5	9.81	5
6	11.43	6
7	11.09	7
8	10.87	8
9	12.15	9
10	10.94	10

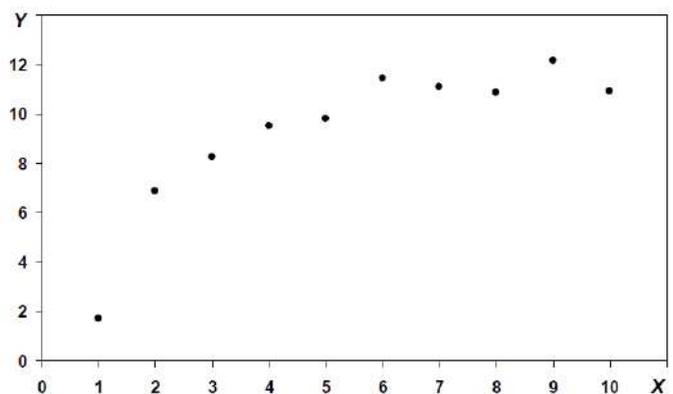
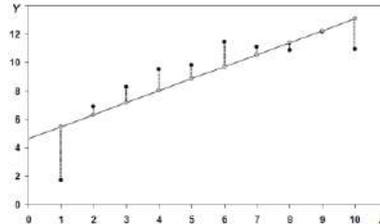


Diagramme de dispersion

Partie 1 : économétrie

Lorsque l'on régresse l'un sur l'autre on remarque que la relation entre X et Y n'est pas linéaire (problème de linéarité). En effet si vous est milliardaire quand votre revenu est élevé vous n'allez pas acheter un milliard de bananes! (Bill Gates). Il y a donc une relation concave (quand le revenu augmente, la consommation de banane diminue).

De plus les résidus indiquent que le **modèle est mal spécifié**. Les résidus ne sont pas distribués aléatoirement en-dessus et au dessous de la ligne de régression (un résidu négatif suivi de 6 positifs....).



Nous créons donc une nouvelle variable Z pour linéariser le modèle (on transforme X pour que la relation entre X et Y soit linéaire).

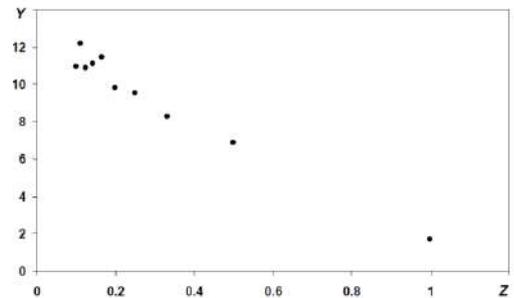
$$\text{Modèle révisé: } Y = \beta_1 + \frac{\beta_2}{X} + u$$

$$z = \frac{1}{X}$$

$$Y = \beta_1 + \beta_2 Z + u$$

Le tableau ci-dessous nous donne les données de Z calculées à partir des données de X.

household	bananas (lbs) Y	Income (\$10,000) X	Z
1	1.71	1	1.00
2	6.88	2	0.50
3	8.25	3	0.33
4	9.52	4	0.25
5	9.81	5	0.20
6	11.43	6	0.17
7	11.09	7	0.14
8	10.87	8	0.13
9	12.15	9	0.11
10	10.94	10	0.10



En générant une nouvelle variable ($Z = 1/X$), en régressant Y sur Z et en remplaçant Z en terme de X, l'ajustement est bien meilleur (relation concave).

```

. g z=1/x
. reg Y z

```

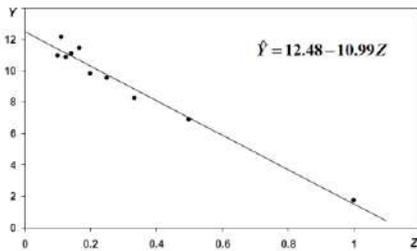
Source	SS	df	MS
Model	83.5451508	1	83.5451508
Residual	2.33609666	8	.292012083
Total	85.8812475	9	9.54236083

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
z	-10.98865	.6496573	-16.915	0.000	-12.48677 -9.490543
_cons	12.48354	48.811	0.000	0.000	11.89378 13.07331

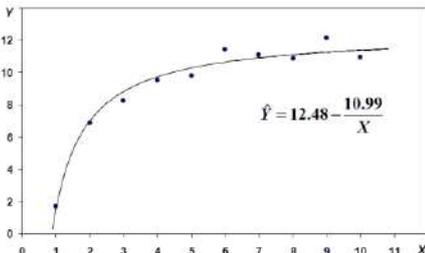
On explique beaucoup mieux le modèle. 97 % des variations de Y sont expliquées par les variations de Z.

$\hat{Y} = 12.48 - 10.99Z$ Quand Z augmente de 1 unité, Y diminue de 10,98.

Partie 1 : économétrie



Estimation de la droite liant Y à Z



On remarque que lorsque X augmente, Y augmente peu. En remplaçant Z en termes de X j'obtiens bien la relation voulue (l'ajustement est bien meilleur).

Il faut donc trouver la bonne forme fonctionnelle aussi non on a un estimateur biaisé.

$$\hat{Y} = 12.48 - 10.99 \circledast Z \rightarrow z = \frac{1}{X}$$

2. Élasticités et modèles logarithmiques doubles

Cette séquence montre comment il est possible d'ajuster des modèles non linaires avec des élasticités constantes.

Commençons par définir **l'élasticité** : c'est le changement proportionnel dans Y par changement proportionnel dans X. On parle en termes de changement relatif (**on ne travaille plus en changement unitaire**). Par exemple, l'élasticité permet de voir comment **Y varie en pourcentage** suite à une augmentation de 1 % de X (exemple: si X augmente de 3 %, Y augmente de 3 %).

$$\text{Elasticité} = \frac{\frac{dY}{Y}}{\frac{dX}{X}} \rightarrow \frac{dY}{dX} \frac{X}{Y}$$

a) Modèle Log-Log¹⁰

Pour passer à l'élasticité, il faut que les variables soient positives.

$$\ln Y = \beta_1 + \beta_2 \ln X + \varepsilon$$

Log ou Ln c'est la même chose.

$$\frac{\partial \ln Y}{\partial X} = \beta_2 \frac{\partial \ln X}{\partial X}$$

$$\frac{\partial \ln Y}{\partial Y} \frac{\partial Y}{\partial X} = \beta_2 \frac{1}{X}$$

$$\frac{X}{1} \frac{\partial \ln Y}{\partial Y} \frac{\partial Y}{\partial X} = \beta_2$$

$$\frac{X}{1} \frac{1}{Y} \frac{\partial Y}{\partial X} = \beta_2 \rightarrow \frac{X}{Y} \frac{\partial Y}{\partial X} = \beta_2 \rightarrow \frac{\partial Y/Y}{\partial X/X} = \beta_2$$

Interprétation du modèle: on dira que lorsque X augmente de 100 % (ou double), Y augmente de β_2 % .

Un changement relatif suite à un changement relatif. Quand je prends ma variable en Log (des 2 côtés), le β_2 est bien une élasticité.

Attention: modèle linéaire (changement unitaire) quand on prend les Log (variations en % -> élasticités).

¹⁰ Voir tableau de régression STATA pour les interprétations des coefficients.

Partie 1 : économétrie

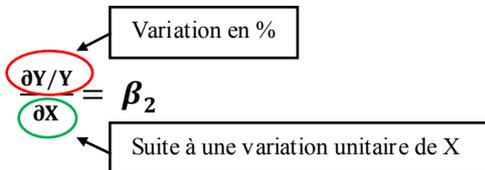
b) Modèle Log-Lin (semi-logarithmiques)¹¹

$$\ln Y = \beta_1 + \beta_2 X + \varepsilon$$

$$\frac{\partial \ln Y}{\partial X} = \beta_2$$

$$\frac{\partial \ln Y}{\partial Y} \frac{\partial Y}{\partial X} = \beta_2$$

$$\frac{1}{Y} \frac{\partial Y}{\partial X} = \beta_2$$



Interprétation du modèle: on dira que lors que X augmente de 1 unité, Y augmente de β_2 % .

Rappel! Dérivé: comment une variable change suite à une variation d'une autre.

c) Modèle Lin-Log (semi-logarithmiques)

$$Y = \beta_1 + \beta_2 \ln X + \varepsilon$$

$$\frac{\partial Y}{\partial X} = \beta_2 \frac{\partial \ln X}{\partial X} = \beta_2 \left(\frac{1}{X} \right)$$

$$\frac{\partial Y}{\partial X} \frac{X}{1} = \beta_2$$

$$\frac{\partial Y}{(\partial X/X)} = \beta_2$$

Changement relatif

Interprétation du modèle: on dira que lors que X augmente de 100 % (ou double), Y augmente de β_2 unité.

d) Variable Dummy dans les modèles Log-Lin

Pour pouvoir lire correctement le coefficient, il faut passer par quelques transformations.

$$\ln Y = b_1 + b_2 D + \varepsilon$$

Variable dichotomique explicative qui peut prendre 2 valeurs (respectivement 0 ou 1)

$$\ln Y_0 = b_1 + \varepsilon \quad \text{Quand } D=0$$

$$\ln Y_1 = b_1 + b_2 + \varepsilon \quad \text{Quand } D=1$$

$$Y_0 = e^{b_1 + \varepsilon}$$

$$Y_1 = e^{b_1 + b_2 + \varepsilon}$$

On prend le modèle sans Log. Quand on a un modèle semi-log et une variable qui peut prendre 2 valeurs, on doit prendre l'exponentielle. Car pour dériver il faut que se soit continue.

Comment Y augmente en % quand D passe de 0 à 1

$$g = \frac{Y_1 - Y_0}{Y_0} = \frac{e^{b_1 + b_2 + \varepsilon} - e^{b_1 + \varepsilon}}{e^{b_1 + \varepsilon}} = e^{b_2} - 1$$

Tx de croissance de Y qd D passe de 0 à 1.

¹¹ Voir tableau de régression STATA pour les interprétations des coefficients concernant chacun des modèles.

Partie 1 : économétrie

3. Terme d'erreur dans les modèles non linéaires

Pour que les résultats de régression en modèle linéarisé aient les propriétés désirées, **le terme d'erreur** dans le **modèle transformé** doit être **additif** et doit satisfaire les conditions de Gauss-Markov. Et il doit être normalement distribué dans le modèle transformé.

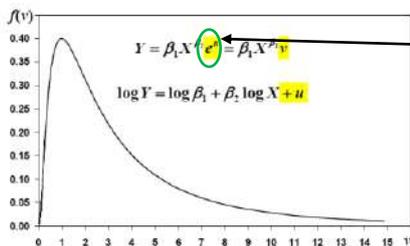
Pour qu'il puisse avoir un terme d'erreur additif dans le modèle transformé, il faut que le terme d'erreur dans le modèle original entre en exponentielle et soit un terme multiplicatif (e^u).

$$Y = \beta_1 X^{\beta_2} e^u \quad \text{Modèle original}$$

Rappel! La somme des Log c'est la somme des produits.

$$\log Y = \log \beta_1 + \beta_2 \log X + u \quad \text{Modèle transformé}$$

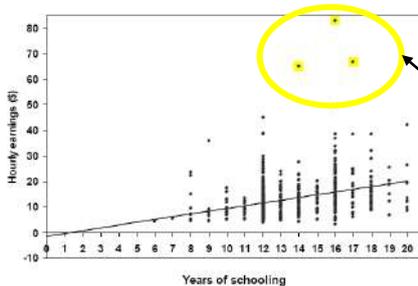
De plus, u doit être normalement distribué pour pouvoir faire des tests t et F



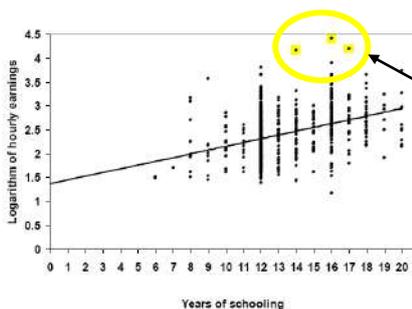
Entre en multiplicatif et de forme normale de sorte que lorsque l'on applique le Log on a une forme additive et normale.

Je veux une forme qui après avoir appliqué le Log devient normal. On prend le Log de manière à ce que tout se rapproche (pour avoir une cloche).

Pour avoir une distribution normale on prend le Log.



Effet levier -> ils attirent la droite car on essaye de minimiser la somme des résidus (et en plus on élève au carré) -> ils ont tendance à biaiser l'estimation.



Quand on prend le Log, les individus sont moins loin, ils attirent la droite vers eux. On a des observations moins extrêmes.

Que se passerait-il si les termes d'erreurs dans le modèle double-logarithmique ou semi-logarithmique étaient additifs au lieu d'être multiplicatifs?

On ne serait pas linéariser le modèle en prenant les Log (on ne sait pas s'en sortir car le Log d'une somme = Log d'une somme)

$$Y = \beta_1 X^{\beta_2} + u$$

$$\log Y = \log(\beta_1 X^{\beta_2} + u)$$

Comment à partir de $\widehat{\log Y}$ je peux estimer \hat{Y}

$\Rightarrow \widehat{\log Y}$ Je ne prédis pas cela $\rightarrow \widehat{\log \hat{Y}}$

Partie 1 : économétrie

B. VARIABLES DICHOTOMIQUES (OU DUMMY)

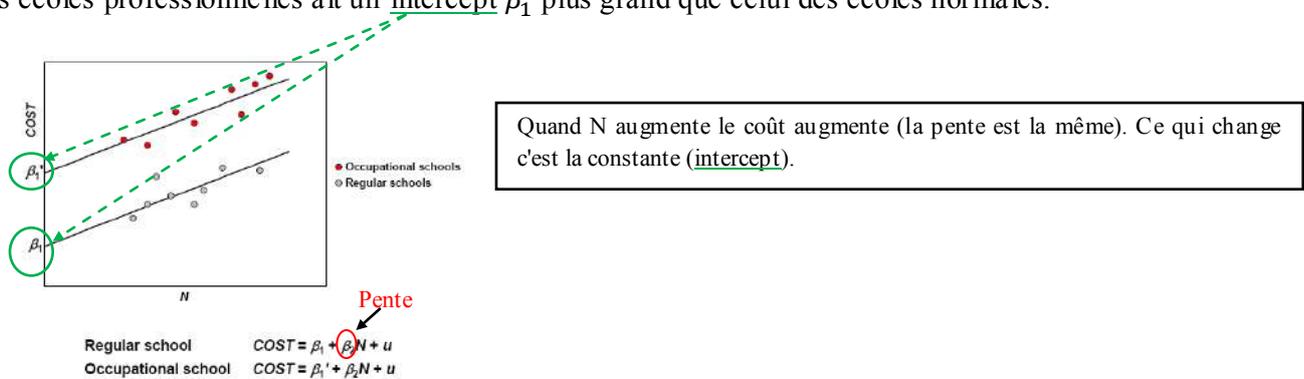
Cette séquence explique comment inclure des variables explicatives qualitatives dans un modèle de régression.

1. Classification de variables binaires avec deux catégories

Rappel! Une variable dichotomique peut prendre deux valeurs (0 ou 1).

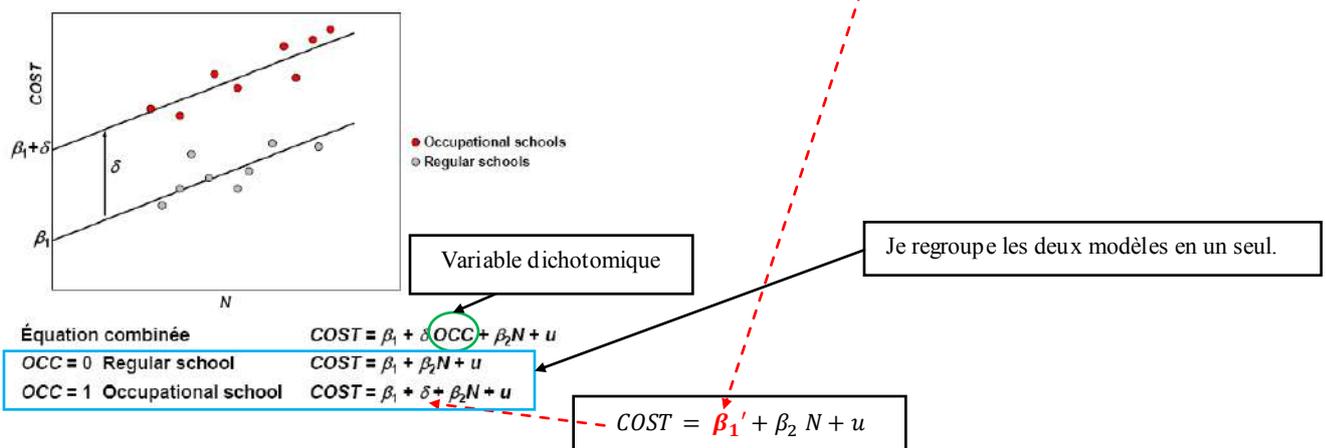
Pour expliquer cette partie, nous avons pris l'exemple des dépenses annuelles (COST), le nombre d'étudiant (N), pour un échantillon de 2 types d'écoles secondaires (normale et professionnelle)¹².

Une façon¹³ de traiter la différence dans le coût serait de poser l'hypothèse que la fonction de coût pour les écoles professionnelles ait un intercept β'_1 plus grand que celui des écoles normales.



Nous poserons $\delta =$ la différence dans le intercepts: $\delta = \beta'_1 - \beta_1$ donc $\beta'_1 = \beta_1 + \delta$

Nous pouvons maintenant combiner les deux fonctions de coût en définissant une variable binaire OCC qui prend la valeur 0 (fonction coûts écoles normales) et 1 (fonction coûts écoles professionnelles). Nous obtenons donc ceci:



¹² Voir l'entièreté de l'exemple dans le syllabus. Je ne détaille que l'essentiel.

¹³ Une autre façon serait de faire des régressions séparées pour les deux types d'écoles. Mais cela signifie que l'on doit estimer les droites de régression sur deux petits échantillons (au lieu d'un seul grand), ce qui détériore la précision des estimations des coefficients.

Partie 1 : économétrie

Nous faisons désormais la régression de COST sur N et OCC, en traitant OCC juste comme n'importe quelle autre variable explicative. Voici le tableau fournit par STATA.

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

Attention: l'école régulière est la référence c'est-à-dire que l'on regarde chaque type (coût) par rapport à cette référence.

< 0,005 on rejette qu'elles ont le même coût.

Chaque élève en plus coûte 331,4 indépendamment de l'école.

En moyenne les écoles professionnelles coûtent 133259,1 indépendamment du nombre d'élèves. Écoles professionnelles coûtent plus cher (car signe +, si - elles sont moins cher).

Nous pouvons réécrire les résultats de la régression sous forme d'une équation. A partir de celle-ci nous pouvons dériver les fonctions de coûts pour les deux types d'école en posant OCC égale à 0 ou 1.

$\hat{COST} = -34,000 + 133,000OCC + 331N$	
Regular School (OCC = 0)	$\hat{COST} = -34,000 + 331N$
Occupational School (OCC = 1)	$\hat{COST} = -34,000 + 133,000 + 331N$ $= 99,000 + 331N$

Le coût marginal par étudiant/année dans ces écoles est de 331 yuans et le coût fixe estimé pour une école normale sans élèves est de -34,000

Le coût moyen d'une école professionnelle est de 99,000 yuans. Le coût marginal est le même que celui des écoles normales.

Référence: je ne mets pas la référence et tout est calculé par rapport à elle.

2. Classification des variables binaires avec plus de deux catégories

Cette séquence explique comment étendre la technique des variables binaires pour traiter une variable explicative qualitative qui a plus de deux catégories.

Reprenons l'exemple avec maintenant 4 catégories.¹⁴

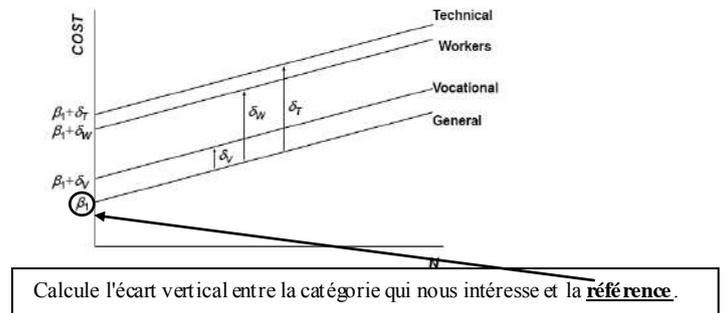
La procédure standard est de choisir une **catégorie de référence** et de définir des variables dichotomiques pour chacune des autres. En général la catégorie de référence est la plus "normale" ou "basique". Donc, dans l'exemple, nous choisissons l'école générale car elle est plus nombreuse et les autres écoles en sont des variations.

¹⁴ Voir l'entièreté de l'exemple dans le syllabus. Je ne détaille que l'essentiel.

Partie 1 : économétrie

Le tableau nous donne un résumé tandis que la figure le représente graphiquement. Les coefficients δ étant les coûts supplémentaires pour chacun des 3 types d'écoles par rapport aux écoles générales¹⁵.

$COST = \beta_1 + \delta_T TECH + \delta_W WORKER + \delta_V VOC + \beta_2 N + u$	
General School (TECH = WORKER = VOC = 0)	$COST = \beta_1 + \beta_2 N + u$
Technical School (TECH = 1; WORKER = VOC = 0)	$COST = (\beta_1 + \delta_T) + \beta_2 N + u$
Skilled Workers' School (WORKER = 1; TECH = VOC = 0)	$COST = (\beta_1 + \delta_W) + \beta_2 N + u$
Vocational School (VOC = 1; TECH = WORKER = 0)	$COST = (\beta_1 + \delta_V) + \beta_2 N + u$



Intéressons nous maintenant aux données :

Tableau de données

Quand j'ai 4 catégories, j'ai 3 dichotomiques car je garde la référence.

School	Type	COST	N	TECH	WORKER	VOC	G	Σ
1	Technical	345,000	623	1	0	0	0	1
2	Technical	537,000	653	1	0	0	0	1
3	General	170,000	400	0	0	0	1	1
4	Workers	526,000	663	0	1	0	0	1
5	General	100,000	563	0	0	0	1	1
6	Vocational	28,000	236	0	0	1	0	1
7	Vocational	160,000	307	0	0	1	0	1
8	Technical	45,000	173	1	0	0	0	1
9	Technical	120,000	146	1	0	0	0	1
10	Workers	61,000	99	0	1	0	0	1

NOTONS que nous devons **garder la référence** car si nous mettons la référence et que nous sommes, nous obtenons une colonne avec des 1 (il y a donc une colinéarité parfaite avec la constante).

Tableau de la régression (STATA)

```
. reg COST N TECH WORKER VOC
```

Source	SS	df	MS		
Model	9.2996e+11	4	2.3249e+11	Number of obs =	74
Residual	5.4138e+11	69	7.8461e+09	F(4, 69) =	29.63
Total	1.4713e+12	73	2.0155e+10	Prob > F	= 0.0000
				R-squared	= 0.6320
				Adj R-squared	= 0.6107
				Root MSE	= 88578

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
N	342.6335	40.2195	8.519	0.000	262.3978 422.8692
TECH	154110.9	26760.41	5.759	0.000	100725.3 207496.4
WORKER	143362.4	27852.8	5.147	0.000	87797.57 198927.2
VOC	53228.64	31061.65	1.714	0.091	-8737.646 115194.9
_cons	-54893.09	26673.08	-2.058	0.043	-108104.4 -1681.748

Chaque élève en plus coûte 342,63 yuans indépendamment de l'école. (Vaut pour type d'école).

On ne peut pas dire que le coût de la Tech et Worker sont différents car l'intervalle de confiance se chevauche (il y a une série de valeur pour lesquelles elles ne sont pas différentes).

Le coût de l'école Voc n'est pas significativement plus grand que celui de l'école générale (si on regarde la différence avec la constante (école générale)).

École Worker coûte 143362,4 de plus que l'école générale

École Tech coûte 154110,9 de plus que l'école générale

¹⁵ Nous n'incluons bien évidemment pas de variable dichotomique pour la catégorie de référence. Si une observation fait référence à une école générale, les 3 variables dichotomiques sont toutes égales à 0 et le modèle de régression est réduit à ses composantes basiques.

Partie 1 : économétrie

Résultat de la régression sous forme équationnelle

$\hat{COST} = -55,000 + 154,000TECH + 143,000WORKER + 53,000VOC + 343N$	
General School (TECH = WORKER = VOC = 0)	$\hat{COST} = -55,000 + 343N$
Technical School (TECH = 1; WORKER = VOC = 0)	$\hat{COST} = -55,000 + 154,000 + 343N$ $= 99,000 + 343N$
Skilled Workers' School (WORKER = 1; TECH = VOC = 0)	$\hat{COST} = -55,000 + 143,000 + 343N$ $= 88,000 + 343N$
Vocational School (VOC = 1; TECH = WORKER = 0)	$\hat{COST} = -55,000 + 53,000 + 343N$ $= -2,000 + 343N$

Chaque élève en plus coûte 343 yuans indépendamment de l'école. (Vaut pour type d'école).

3. Test de F

On peut effectuer un test de F sur le pouvoir explicatif des variables dichotomiques considérées.

$H_0: \delta_T = \delta_W = \delta_V = 0$ Le coût d'un type d'école n'est pas différent de celui de l'école générale (référence).

Hypothèse alternative: au moins un δ est différents de 0

Rappel de la formule de $F(\text{coût}, d.l. \text{dispo}) = \frac{\text{amélioration/coût}}{\text{partie résiduelle non expliquée} / d.l. \text{disponible}}$

. reg COST N							
Source	SS	df	MS	Number of obs			
Model	5.7974e+11	1	5.7974e+11	74	F(1, 72)	= 46.82	
Residual	8.9160e+11	72	1.2383e+10		Prob > F	= 0.0000	
Total	1.4713e+12	73	2.0155e+10		R-squared	= 0.3940	
					Adj R-squared	= 0.3856	
					Root MSE	= 1.1e+05	

. reg COST N TECH WORKER VOC							
Source	SS	df	MS	Number of obs			
Model	9.2996e+11	4	2.3249e+11	74	F(4, 69)	= 29.63	
Residual	5.4138e+11	69	7.8461e+09		Prob > F	= 0.0000	
Total	1.4713e+12	73	2.0155e+10		R-squared	= 0.6320	
					Adj R-squared	= 0.6107	
					Root MSE	= 88576	

$$F(3,69) = \frac{(8.92 \times 10^{11} - 5.41 \times 10^{11}) / 3}{5.41 \times 10^{11} / 69} = 14.92$$

$$F(3,60)_{\text{crit}, 0.1\%} = 6.17$$

On rejette l' H_0 : on rejette que toutes les écoles ont le mêmes coût.

Partie 1 : économétrie

Chapitre 5: mauvaise spécification de variables, variables proxy et tester une régression linéaire

A. HYPOTHÈSE DE GAUSS-MARKOV (BONNE SPÉCIFICATION)

1. Mauvaise spécification de variables

Pour garder l'analyse simple, nous ferons l'hypothèse que Y dépend soit de X₂ ou il dépend à la fois de X₂ et X₃.

Dans cette séquence, nous allons voir les conséquences d'une mauvaise spécification d'un modèle de régression en termes de variables explicatives. Pour suivre l'analyse durant ce chapitre nous nous baserons sur le tableau (ci-dessous) qui résume l'ensemble des cas possibles.

Conséquences de la mauvaise spécification			
		True Model	
		$Y = \beta_1 + \beta_2 X_2 + u$	$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$
Fitted Model	$\hat{Y} = b_1 + b_2 X_2$	Spécification correcte, Pas de problème	Coefficients biaisés (en général). Ecart-types invalides.
	$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$	Coefficients non biaisés (en général), mais inefficaces. Ecart-types valides (en général)	Spécification correcte, Pas de problème

L'omission d'une variable pertinente à pour causes:

- Coefficients sont biaisés
- Invalidité des écarts-types des estimateurs.
- ➔ Plus grave car on oublie une variable importante => estimation biaisée.

L'inclusion d'une variable non pertinente à pour causes:

- Coefficients non biaisés
- Validité des écarts-types
- ➔ Moins grave car j'ajoute une variable qui n'a pas d'importance => écarts-types gonflent et on perd en efficacité.

a) Mauvaise spécification de variables I: omission d'une variable pertinente

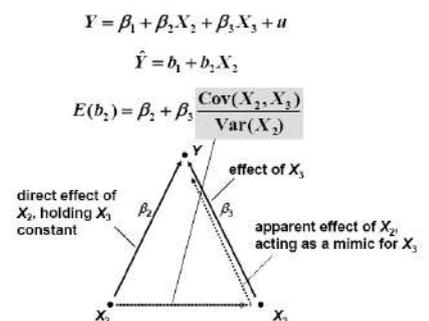
Nous allons démontrer que l'omission d'une variable explicative pertinente a pour cause **un biais dans les coefficients**. En conséquence, les **écarts-types** des estimateurs, les **tests t** et le **test F** sont **non valides**.

Rappel! un **estimateur est dit non biaisé** si l'espérance de mon estimateur est égale à la vraie valeur.

Démonstration intuitive

Si X₃ est omis, X₂ aura un double effet. Il aura un effet direct (β_2) et aussi un effet indirect apparent car il agit comme une proxy de X₃ manquante.

La force de l'effet de proxy dépend de deux facteurs : la force de l'effet de X₃ sur Y qui est donné par β_3 et la capacité de X₂ à imiter X₃. La capacité de X₂ à imiter X₃ est déterminée par le coefficient de pente obtenu lorsque X₃ est régressé sur X₂ qui est $\frac{Cov(X_2, X_3)}{Var(X_2)}$.



⇒ Si j'oublie une variable pertinente et qu'elle est corrélée avec une variable d'un modèle, on a un biais.

Partie 1 : économétrie

Démonstration mathématique

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$\begin{aligned} b_2 &= \frac{\text{Cov}(X_2, Y)}{\text{Var}(X_2)} = \frac{\text{Cov}(X_2, [\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u])}{\text{Var}(X_2)} \\ &= \frac{\text{Cov}(X_2, \beta_1) + \text{Cov}(X_2, \beta_2 X_2) + \text{Cov}(X_2, \beta_3 X_3) + \text{Cov}(X_2, u)}{\text{Var}(X_2)} \\ &= \frac{0 + \beta_2 \text{Cov}(X_2, X_2) + \beta_3 \text{Cov}(X_2, X_3) + \text{Cov}(X_2, u)}{\text{Var}(X_2)} \\ &= \beta_2 + \beta_3 \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)} + \frac{\text{Cov}(X_2, u)}{\text{Var}(X_2)} \end{aligned}$$

$$E(b_2) = E\left\{ \beta_2 + \beta_3 \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)} + \frac{\text{Cov}(X_2, u)}{\text{Var}(X_2)} \right\} \leftarrow \text{Pour voir s'il y a bien un biais, nous prenons l'espérance de } b_2$$

$$= E(\beta_2) + E\left\{ \beta_3 \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)} \right\} + E\left\{ \frac{\text{Cov}(X_2, u)}{\text{Var}(X_2)} \right\} \leftarrow \text{Égale à 0 voir démonstration ci-après}$$

$$= \beta_2 + \beta_3 \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)} \leftarrow b_2 \text{ est biaisé d'un montant } \beta_3 \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)}$$

$$E\left\{ \frac{\text{Cov}(X_2, u)}{\text{Var}(X_2)} \right\} = E\left(\frac{\sum (X_{2i} - \bar{X}_2)(u_i - \bar{u})}{\sum (X_{2i} - \bar{X}_2)^2} \right)$$

$$\begin{aligned} E\left(\frac{\sum (X_{2i} - \bar{X}_2)(u_i - \bar{u})}{\sum (X_{2i} - \bar{X}_2)^2} \right) &= \frac{1}{\sum (X_{2i} - \bar{X}_2)^2} E(\sum (X_{2i} - \bar{X}_2)(u_i - \bar{u})) \\ &= \frac{1}{\sum (X_{2i} - \bar{X}_2)^2} \sum E\{(X_{2i} - \bar{X}_2)(u_i - \bar{u})\} \\ &= \frac{1}{\sum (X_{2i} - \bar{X}_2)^2} \sum (X_{2i} - \bar{X}_2) E(u_i - \bar{u}) \\ &= 0 \end{aligned}$$

NB: avec la formule, nous pouvons également prédire le sens du biais du coefficient (vers le haut ou vers le bas)¹⁶.

¹⁶ Voir le syllabus pour divers exemples concernant le sens du biais.

Partie 1 : économétrie

Comment le R^2 se comporte quand une variable est omise?

En regardant le tableau, on pourrait dire que ASVABC explique 33 % de la variance de S et SM 13 %. **Mais le R^2 n'est pas la somme R^2 individuelle.** La régression multiple révèle que le pouvoir explicatif joint est de 0,36 % et non 0,46 %. Le R^2 nous donne le % des variations de Y par rapport à X, **il ne sert pas à choisir un modèle.**

. reg S ASVABC SM					
Source	SS	df	MS		
Model	1230.2039	2	615.101949	Number of obs =	570
Residual	2224.04347	567	3.92247526	F(2, 567) =	156.81
Total	3454.24737	569	6.07073351	Prob > F =	0.0000
				R-squared =	0.3561
				Adj R-squared =	0.3539
				Root MSE =	1.9805

. reg S ASVABC					
Source	SS	df	MS		
Model	1153.80864	1	1153.80864	Number of obs =	570
Residual	2300.43873	568	4.05006818	F(1, 568) =	284.89
Total	3454.24737	569	6.07073351	Prob > F =	0.0000
				R-squared =	0.3340
				Adj R-squared =	0.3329
				Root MSE =	2.0125

. reg S SM					
Source	SS	df	MS		
Model	443.110436	1	443.110436	Number of obs =	570
Residual	3011.13693	568	5.30129742	F(1, 568) =	83.59
Total	3454.24737	569	6.07073351	Prob > F =	0.0000
				R-squared =	0.1283
				Adj R-squared =	0.1267
				Root MSE =	2.3025

b) Mauvaise spécification de variables II: inclusion d'une variable non pertinente

Dans ce cas, les écarts-types des estimateurs restent valides mais sont inutilement grands.

Lorsque l'on rajoute une variable non pertinente cela revient à:

$$Y = \beta_1 + \beta_2 X_2 + u$$

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

Variable de trop alors qu'elle ne sert à rien.

$$Y = \beta_1 + \beta_2 X_2 + 0 X_3 + u$$

Lorsque l'on rajoute X_3 au modèle avec un coefficient de 0, **le vrai modèle et le modèle ajusté coïncident.** Donc b_2 sera un estimateur non biaisé de β_2 et b_3 sera un estimateur non biaisé de 0.

Cependant, la variance de la population de b_2 sera plus grande

$$\text{Variance de population de } b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\text{Var}(X_2)} \times \frac{1}{1-r_{X_2 X_3}^2} \quad (\text{régression multiple})$$

Sert à rien mais X_3 est corrélé à X_2 -> écarts-types gonflent
Si X_2 et X_3 sont non corrélés, il n'y aura pas de perte d'efficacité.

$$b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\text{Var}(X_2)} \quad (\text{régression simple})$$

L'estimateur de b_2 sera moins efficace dans le modèle de régression multiple que dans le modèle de régression simple. Car les écarts-types auront tendance à être plus grand que ceux obtenus en régression simple (ce qui reflète la perte d'efficacité)¹⁷.

Si la variable omise est corrélée avec une variable dans le modèle cela devient un problème car elle va dans le terme d'erreur. Mais si on oublie de mettre des variables dans le modèle et qu'elles sont indépendantes c'est moins grave.

¹⁷ Voir exemples dans le syllabus.

Partie 1 : économétrie

B. VARIABLES PROXY

1. Variables proxy

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + u$$

Supposons que nous n'avons pas de données sur X_2 . Une régression sans X_2 mènerait à des estimations des **coefficients biaisés** et des **écarts-types invalides**. Ces problèmes peut-être réduits ou éliminés par l'utilisation d'une proxy à la place de X_2 .

Une **variable proxy** est une variable qui est liée à la variable manquante.

Si une variable proxy appropriée a été trouvée, le modèle peut-être réécrit de cette manière:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + u$$

Manque une variable X_2 (or elle est indispensable) mais on sait qu'elle est liée à Z que l'on a.

$$X_2 = \gamma + \mu Z$$

$$Y = \beta_1 + \beta_2(\gamma + \mu Z) + \beta_3 X_3 + \dots + \beta_k X_k + u$$

$$Y = (\beta_1 + \beta_2 \gamma) + \beta_2 \mu Z + \beta_3 X_3 + \dots + \beta_k X_k + u$$

On obtient donc un modèle avec toutes les variables observables. Si la relation proxy est parfaite et que nous ajustons cette relation, la plupart des résultats de régression seront sauvés.

Qu'est-ce qui change dans le modèle après intégration de la proxy ?

Les estimations des coefficients X_3, \dots, X_k seront les mêmes que ceux qui auraient été obtenus si il avait été possible de régresser Y sur X_2, \dots, X_k .

Les écart type des estimateurs et les statistiques t des coefficients de X_3, \dots, X_k **seront les mêmes** que ceux que nous aurions obtenus s'il avait été possible de régresser Y sur X_2, \dots, X_k .

R^2 sera le même que ce qu'il aurait été s'il avait été possible de régresser Y sur X_2, \dots, X_k .

Le coefficient de Z sera une estimation de $\beta_2 \mu$, et donc il ne sera **pas possible d'obtenir un estimateur de β_2** .

$$Y = (\beta_1 + \beta_2 \gamma) + \beta_2 \mu Z + \beta_3 X_3 + \dots + \beta_k X_k + u$$

Je suis incapable d'identifier le β_1 mais bien le bloc

Cependant, **la statistique t de Z sera la même** que celle qui aurait été obtenue pour X_2 si il avait été possible de régresser Y sur X_2, \dots, X_k et donc nous sommes **capable d'affirmer la significativité de X_2** , même si nous ne sommes pas capable d'estimer son coefficient.

Partie 1 : économétrie

On cherche des variables qui ressemblent le plus possible à la variable d'origine de sorte à diminuer le biais. Cependant il est rare de rencontrer des proxys parfaits d'une variable. On cherche donc une variable qui approxime la vraie variable de sorte à diminuer le biais. Une proxy peut être une combinaison linéaire de une ou plusieurs variables qui misent ensemble ont un effet sur Y car elles résument l'information d'une variables que je n'ai pas¹⁸.

C. TESTER UNE RESTRICTION LINÉAIRE

Nous allons effectuer un test de F¹⁹. Il a été suggéré que l'impact de l'éducation parentale était le même pour les deux parents, c'est-à-dire que β_2 et β_3 pourraient être égaux. Si c'est le cas alors nous avons SP qui est la somme de SM et de SF. Notons que nous ne rajoutons qu'un paramètre au modèle (SP). Attention l'hypothèse nulle est qu'il n'y a aucune différence entre le modèle complet et restreint.

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + u \quad \leftarrow \text{Modèle complet: aucune restriction}$$

$$\beta_4 = \beta_3$$

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 (SM + SF) + u$$

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SP + u \quad \leftarrow \text{Modèle restreint: j'ajoute une restriction}$$

$$SP = SM + SF \quad \beta_4 = \beta_3$$

$$H_0: \beta_4 = \beta_3 \quad H_1: \beta_4 \neq \beta_3$$

(4-3)

$$F(1, n - k) = \frac{(RSS_R - RSS_U)/1}{RSS_U/(n-k)} = \frac{2177.51 - 2176.01}{2176.01 / 566} = 0.39$$

On ne rejette pas H0 car < à la valeur critique.

! On ne rejette jamais quand la valeur est < à 1.

Modèle complet				Modèle restreint			
Source	SS	df	MS	Source	SS	df	MS
Model	1278.24153	3	426.080508	Model	1276.73764	2	638.368819
Residual	2176.00584	566	3.84453329	Residual	2177.50973	567	3.84040517
Total	3454.24737	569	6.07073351	Total	3454.24737	569	6.07073351

Des restrictions linéaires peuvent également être testées en utilisant un test t. Pour ce faire, on soustrait le modèle complet du modèle restreint.

$$1 \rightarrow S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + u$$

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SP + u$$

$$0 = \beta_3 SM + \beta_4 SF - \beta_3 SP$$

$$0 = \beta_3 SM + \beta_4 SF - \beta_3 (SM + SF)$$

$$0 = \beta_3 SM + \beta_4 SF - \beta_3 SM + \beta_3 SF$$

$$2 \rightarrow 0 = (\beta_3 - \beta_4) SF$$

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SP + (\beta_3 - \beta_4) SF + u$$

$$H_0: \beta_4 - \beta_3 = 0 \quad H_1: \beta_4 - \beta_3 \neq 0$$

$$S - S = 0$$

$$\beta_1 - \beta_1 = 0$$

$$\beta_2 ASVABC - \beta_2 ASVABC = 0$$

$$\beta_3 SM - \beta_3 SP = \beta_3 SM - \beta_3 SP$$

$$\beta_4 SF - 0 = \beta_4 SF$$

$$u - u = 0$$

On réécrit SP en termes de SM et SF

Se sera = à 0 quand $\beta_4 = \beta_3$. Donc les modèles seront les mêmes quand $\beta_4 = \beta_3$.

1 + 2 = au modèle de départ (voir feuille jointe dans le syllabus).

On rejette pas H0 car < à 2 (je ne rejette pas que les modèles sont les mêmes).

reg S ASVABC SP SF						
Source	SS	df	MS			
Model	1278.24153	3	426.080508			
Residual	2176.00584	566	3.84453329			
Total	3454.24737	569	6.07073351			

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ASVABC	.1295006	.0099544	13.009	0.000	.1099486 .1490527
SP	.069403	.0422974	1.641	0.101	-.013676 .152482
SF	.0408654	.0653386	0.625	0.532	-.0874704 .1692012
_cons	4.914654	.5063527	9.706	0.000	3.920094 5.909214

¹⁸ Voir exemple dans le syllabus.
¹⁹ Voir exemple complet dans le syllabus (je ne détaille que l'essentiel).

Partie 1 : économétrie

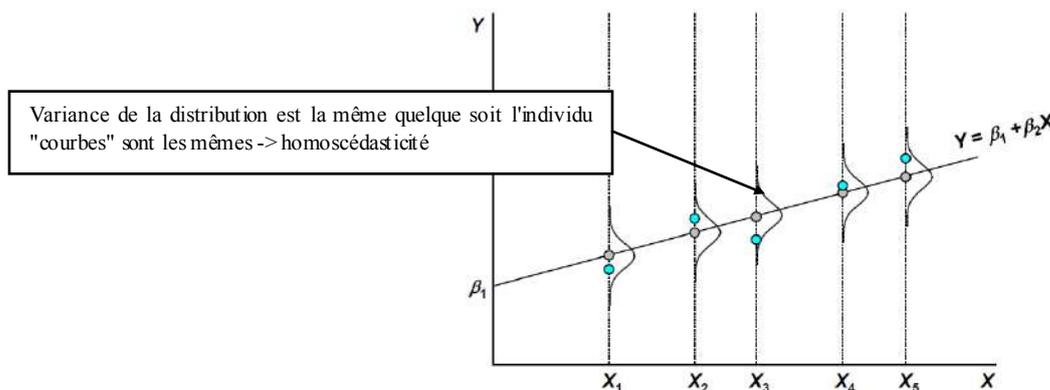
Chapitre 6 : Heteroscedasticité

A. HOMOSCEDASTICITÉ ET HETEROSCEDASTICITÉ

Cette séquence introduit le concept d'hétéroscédasticité qui fait référence à la distribution du terme d'erreur dans un modèle de régression.

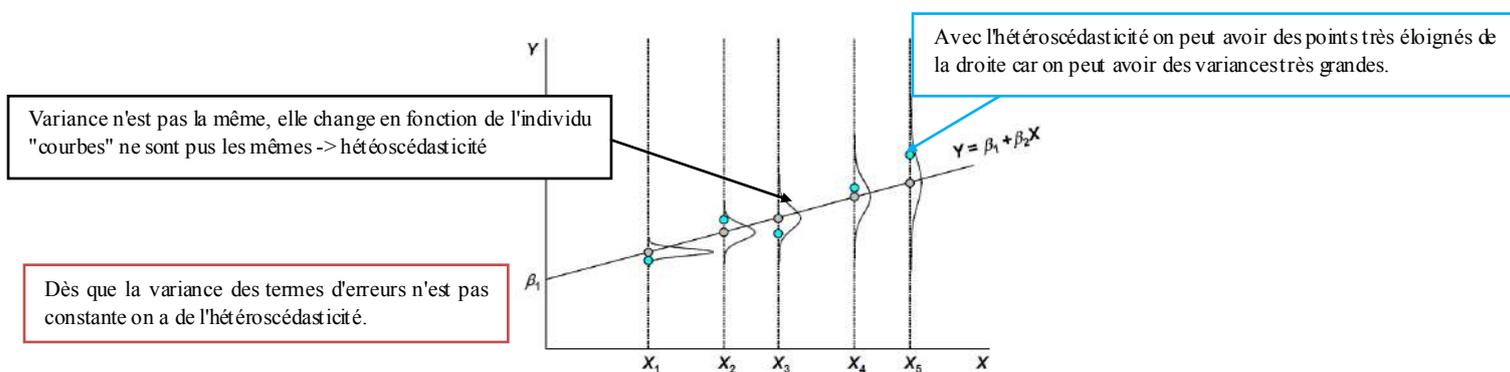
1. Homoscedasticité

On parle **d'homoscédasticité** lorsque la **variance** de la distribution est la **même** pour chaque observation (les distributions normales montrées ont la même variance). Si cette condition est satisfaite, le terme d'erreur est dit être "homoscédastique". Une fois que l'échantillon a été tirée, des observations seront plus proche que d'autres de la droite (mais on ne sait pas anticiper lesquelles).



2. Hétéroscédasticité

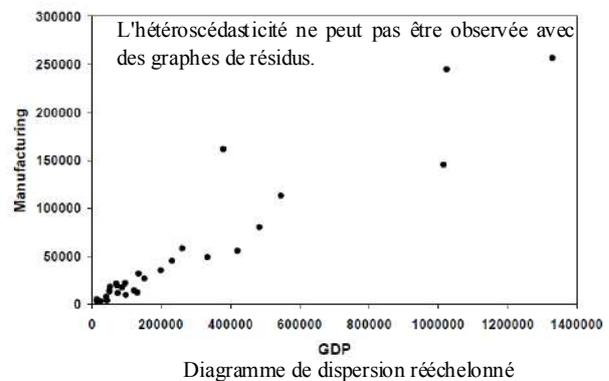
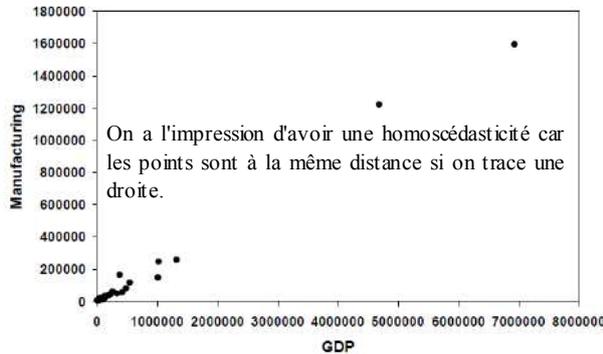
On parle **d'hétéroscédasticité** lorsque la **variance** de la distribution du terme d'erreur **varie** en fonction des observations (des individus)-> variance n'est plus constante. Quand la distribution n'est pas la même pour chaque observation, le terme d'erreur est dit être sujet à de l'hétéroscédasticité. La distribution du terme d'erreur variera donc avec chaque X.



Mais attention, l'hétéroscédasticité **ne cause pas de biais**, elle va simplement rendre **invalide l'inférence** (écart-types sont mal estimés et les tests (t-F) sont invalides).

Partie 1 : économétrie

Prenons un exemple : diagramme de dispersion du GDP en fonction de MANUFACTURING. Le diagramme de dispersion est dominé par le Japon et les USA et il est difficile de détecter quelque sorte de tendance. Cependant si ces 2 pays sont enlevés et que le diagramme de dispersion est rééchélonné, une image claire d'hétérosédasticité apparaît²⁰.

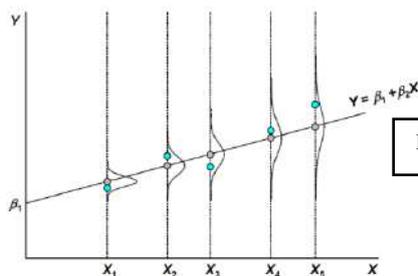


3. Comment détecter l'hétérosédasticité ?

a) Test Goldfeld-Quant²¹

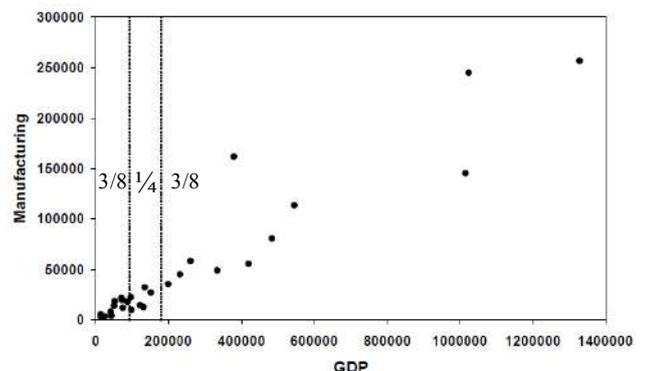
Les formes possibles d'hétérosédasticité sont infinies. Cependant, le test de Goldfeld-Quant teste une forme particulière commune d'hétérosédasticité: **une distribution des écart-types proportionnelle à la taille des variables explicatives.**

Illustration:



Écart-type de la distribution est proportionnel à X (quand X augmente, l'écart-type augmente).

Le test de Goldfeld-Quant consiste à diviser l'échantillon en 3 parties. Le plus souvent on prendra les 3/8 des observations avec les valeurs les plus petites de la variable X, les 3/8 des observations avec les valeurs les plus grandes et 1/4 au milieu.



²⁰ Voir exemple complet dans le syllabus (je détaille que l'essentiel).

²¹ Ibid.

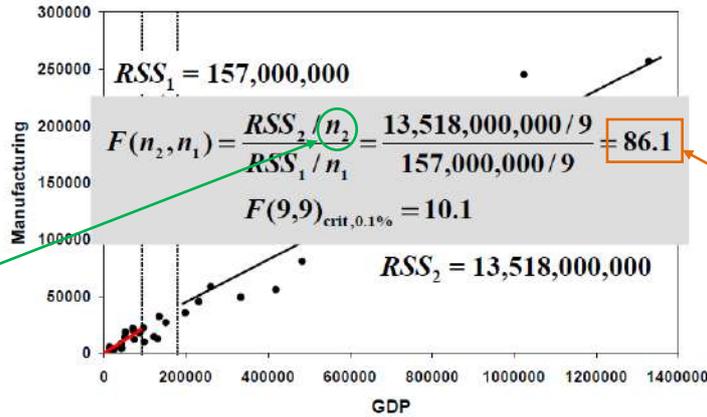
Partie 1 : économétrie

On ajuste ensuite les lignes de régression pour les colonnes extrêmes. On comparera la somme des carrés résiduels pour les deux régressions: RSS_1 et RSS_2 . Si le terme d'erreur est homoscedastique, il ne devrait pas y avoir de différence entre RSS_1 et RSS_2 . La statistique du test est distribuée selon une distribution de Fisher.

S'il y a homoscedasticité, la somme des carrés des résidus (distantes entre la droite) n'est pas différentes. Car la distribution est identique, les points se trouvent +/- à la même distante de la droite (la variance est constante).

H_0 est l'homoscedasticité (je teste si la dispersion des points à droite et à gauche est la même).

Ils seront les mêmes car ce sont le nombre de degrés de liberté dans la régression inférieure et supérieure.



Supérieure à 10.1 je rejette l'hypothèse d'homoscedasticité.
 ⇒ J'ai donc détecté l'hétéroscedasticité.

b) Test de white²²

Le test de White est une seconde façon de détecter l'hétéroscedasticité. L'idée du test de White est de régresser le **RSS** sur **les variables explicatives**, **leur carré** et tous **les produits croisés**. Si au moins un des paramètres est différent de 0 alors cela veut dire que les résidus dépendent d'un des X (hétéroscedasticité).

Je n'ai pas le terme d'erreur mais j'ai les résidus.

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$E(Y) = \hat{Y} = b_1 + b_2 X_2 + b_3 X_3 \quad Y = \hat{Y} + e$$

$$e^2 = \gamma_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_2^2 + \gamma_5 X_3^2 + \gamma_6 X_2 X_3 + \varepsilon$$

$$H_0 = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \gamma_6 \quad H_1 = \gamma_i \neq 0 \text{ pour au moins un } i$$

4. Comment résoudre l'hétéroscedasticité

Si on a de l'hétéroscedasticité, on a un problème dans les variances.

a) Matrice sandwich

Dans ce cas, la variance est correcte mais on ne sait pas modéliser, on ne règle pas l'hétéroscedasticité (exemple de la pluie et du parapluie).

$$Var(\hat{\beta}) = (X' - X)^{-1} X' \varepsilon \varepsilon' X (X' - X)^{-1} \Rightarrow Var(\hat{\beta}) = \sigma^2 (X' - X)^{-1}$$

Je ne connais pas dont je l'estime avec le terme d'erreur.

²² Voir syllabus pour exemple complet (je détaille que l'essentiel).

Partie 1 : économétrie

b) $u_i = \sigma_i^2$

La connaissance de σ_i dans chaque observation, permettrait de déduire un modèle homosédastique en divisant l'équation par σ_i . La variance du terme d'erreur *dans la population* dans le modèle révisé est maintenant égale à 1 dans toutes les observations, le terme d'erreur est donc homosédastique.

$$Y = \beta_1 + \beta_2 X + u$$

variance de la population de $u_i = \sigma_i^2$

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

variance de la population de $\left\{ \frac{u_i}{\sigma_i} \right\} = \frac{1}{\sigma_i^2}$ variance de la population de u_i

Je le sors et je l'éleve au carré.

Homosédasticité car le terme d'erreur à une variance constante (1) pour tous les individus.

$$Y' = \beta_1 H + \beta_2 X' + u' \quad Y' = \frac{Y_i}{\sigma_i}, \quad H = \frac{1}{\sigma_i}, \quad X' = \frac{X_i}{\sigma_i}, \quad u' = \frac{u_i}{\sigma_i}$$

c) Variance est proportionnelle à une variable qui n'est pas dans le modèle

En pratique nous ne connaissons pas la valeur de la variance de population du terme d'erreur de chaque observation i . Cependant il peut être raisonnable de supposer qu'elle est proportionnelle à une certaine variable mesurable Z_i . Si c'est le cas, nous pouvons rendre le modèle homosédastique en divisant par Z_i . Le terme d'erreur dans le modèle révisé a une variance constante λ^2 .

$$Y = \beta_1 + \beta_2 X + u$$

variance de la population de $u_i = \sigma_i^2$

$$\sigma_i = \lambda Z_i$$

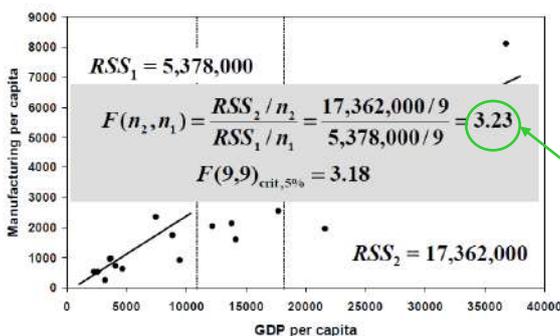
Variance proportionnelle à une variable qui n'est pas dans le modèle.

$$\frac{Y_i}{Z_i} = \beta_1 \frac{1}{Z_i} + \beta_2 \frac{X_i}{Z_i} + \frac{u_i}{Z_i}$$

Nouveau terme d'erreur

variance de la population de $\left\{ \frac{u_i}{Z_i} \right\} = \frac{1}{Z_i^2} \sigma_i^2 = \frac{\sigma_i^2}{Z_i^2} = \frac{\sigma_i^2}{\lambda^2 Z_i^2} = \lambda^2$

Si je normalise par une variable proportionnelle qui n'est pas dans le modèle (variance du terme d'erreur constante).

$$Y' = \beta_1 H + \beta_2 X' + u' \quad Y' = \frac{Y_i}{Z_i}, \quad H = \frac{1}{Z_i}, \quad X' = \frac{X_i}{Z_i}, \quad u' = \frac{u_i}{Z_i}$$


Dans notre exemple, nous avons divisé par la variable population, suspectée d'être à l'origine de notre hétérosédasticité. Cependant, l'hypothèse nulle est rejetée à un niveau de 5 % (pas de 1 %).

Supérieure à 3.18 on rejette l'hypothèse d'homosédasticité.

Partie 1 : économétrie

d) La variance est proportionnelle à une variable de mon modèle

$$Y = \beta_1 + \beta_2 X + u$$

variance de la population de $u_i = \sigma_i^2$

$$\sigma_i = \lambda X_i$$

Variance proportionnelle à une variable de mon modèle.

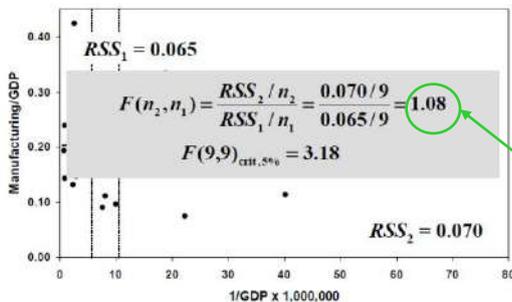
$$\frac{Y_i}{X_i} = \beta_1 \frac{1}{X_i} + \beta_2 + \frac{u_i}{X_i}$$

Nouveau terme d'erreur

variance de la population de $\left\{ \frac{u_i}{X_i} \right\} = \frac{1}{X_i^2} \sigma_i^2 = \frac{\sigma_i^2}{\lambda^2 X_i^2} = \lambda^2$

Si je normalise par une variable proportionnelle qui est dans le modèle (variance du terme d'erreur constante). Permet de résoudre le problème d'hétéroscédasticité.

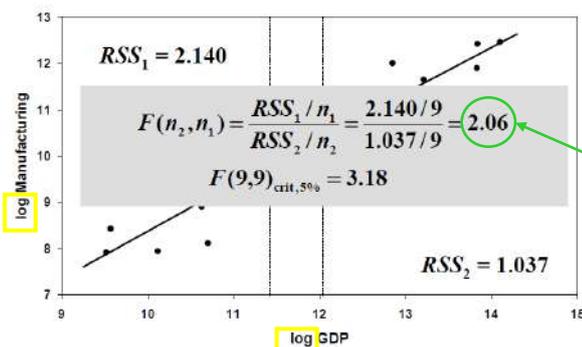
$$Y' = \beta_1 H + \beta_2 + u' \quad Y' = \frac{Y_i}{X_i}, \quad H = \frac{1}{X_i}, \quad u' = \frac{u_i}{X_i}$$



Dans notre exemple en divisant les deux variables par GDP suspecté d'être à l'origine de l'hétéroscédasticité. On peut voir que cette dernière a été éliminée par la preuve du test F.

Inférieure à 3.18 on ne rejette pas l'hypothèse d'homoscédasticité.

e) Régression logarithmique



Afin d'éliminer l'hétéroscédasticité il est possible de mettre le modèle sous forme logarithmique. On peut le confirmer avec un test de F.

Inférieure à 3.18 on ne rejette pas l'hypothèse d'homoscédasticité.

Partie 1 : économétrie

5. Résumé des régressions avec les 4 spécifications alternatives du modèle²³

Nous pouvons remarquer que les différentes manières de résoudre l'hétéroscédasticité n'ont pas affectées les paramètres estimés.

$$\begin{aligned} \hat{MANU} &= 604 + 0.194GDP & R^2 &= 0.89 \\ & (5700) (0.013) \\ \frac{\hat{MANU}}{POP} &= 612 \frac{1}{POP} + 0.182 \frac{GDP}{POP} & R^2 &= 0.70 \\ & (1371) (0.016) \\ \left(\begin{aligned} \frac{\hat{MANU}}{GDP} &= 0.189 + 533 \frac{1}{GDP} & R^2 &= 0.02 \\ & (0.019) (841) \\ \log \hat{MANU} &= -1.694 + 0.999 \log GDP & R^2 &= 0.90 \\ & (0.785) (0.066) \end{aligned} \right. \end{aligned}$$

J'ai résolu le problème d'hétéroscédasticité.

a) Quel modèle choisir ?

Nous choisirions probablement le modèle logarithmique étant donné son R^2 élevé. Mais attention ne **JAMAIS** utiliser le R^2 pour comparer des modèles (il est **régression invariant** !). C'est-à-dire qu'il change en fonction de la pente de la droite. De ce fait, il ne sert pas à choisir un modèle!!!

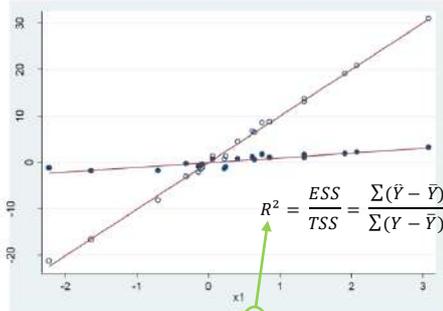
6. Problème avec le R^2

Nous allons montrer par un exemple que le R^2 est de régression invariant (il ne sert pas à choisir un modèle).

```
clear (effacer la mémoire)
set obs 20 (je travaille avec 20 obs)
drawnorm x1 e (générer à partir d'une normal (je génère X1 et e de manière indépendante)).
gen y=x1 +e
gen y2 = 10*x1+e
reg y x1
reg y2 x1
twoway (scatter yx1, mcolor(navy)) ///
scatter y2 x1, mcolor(navy) msymbol(circle_hollow) ///
(lfit y x1, lcolor(maroon)) (lfit y2 x1, lcolor(maroon)) ///
, legend(off)
```

Permet d'avoir les graphes.

On s'attend à ce que la dispersion des points autour de la droite soit la même car le terme d'erreur est le même.



$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} = 1 - \frac{\sum e^2}{\sum(Y - \bar{Y})^2}$$

Comme le R^2 est construit avec le e^2 => on s'attend à avoir le même R^2 (mais le R^2 est sensible à la pente).

Résidu est le même.

Source	SS	df	MS
Model	30.3225246	1	30.3225246
Residual	9.29389008	18	.516327227
Total	39.6164147	19	2.08507446

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	1	0.000	7.66	0.000	-.7512267 1.318699
_cons	0	0.678	-0.42	0.678	-.4296231 .2859011

La variance des résidus est la même (dispersion identique).

Source	SS	df	MS
Model	2850.6732	1	2850.6732
Residual	9.2938909	18	.516327272
Total	2859.96709	19	150.524584

t	P> t	[95% Conf. Interval]
10.30	0.000	9.751226 10.3187
-0.42	0.678	-.429623 .2859012

Le fait de changer la pente le R^2 a changé. Les 2 modèles sont identiques en termes de qualité d'ajustement car même X.

²³ Voir syllabus pour plus de détail sur toutes les alternatives.

Partie 1 : économétrie

Chapitre 7 : conditions de Gauss-Markov, exogénéité, erreurs de mesure, variables instrumentales et spécification du test de Durbin-Wu-Hausman

A. CONDITIONS DE GAUSS-MARKOV

- 1) Bonne spécification: $E(u_i) = 0$ sinon problème d'inférence et de convergence.
- 2) Normalité: $u_i \sim \text{Normal}$ normal sinon perte d'efficacité et problème avec l'inférence lorsque l'échantillon est petit (lorsque l'échantillon est grand ça n'est pas grave).
- 3) Homosédasticité: $\text{Var}(u_i) = \sigma^2, \text{constant}$ sinon on a hétéroscédasticité (problème de variance).
- 4) Exogénéité: $\text{Cov}(u_i, X_i) = 0$ sinon coefficients biaisés, problème d'inférence et de convergence.
- 5) Indépendance sérielle: $\text{Cov}(u_i, u_j) = 0 \quad i \neq j$

B. EXOGÉNITÉ

1. Régresseurs stochastiques

On parle **d'exogénéité** lorsque la **covariance** entre **le terme d'erreur et les variables explicatives** est égale à 0. Si ce n'est pas le cas j'ai des estimateurs biaisés, non convergents et j'estime n'importe quoi.

Jusqu'à présent nous avons fait l'hypothèse que les variables explicatives n'avaient pas de composantes aléatoires. Les estimateurs restent non biaisés, efficaces et convergents à condition que les composantes aléatoires des variables explicatives soient distribuées indépendamment du terme d'erreur (**u et X soit indépendant**) et que le modèle soit correctement spécifié.

$$y = \alpha + \beta x + u$$

$$b = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = \frac{\text{Cov}(x, [\alpha + \beta x + u])}{\text{Var}(x)} = \beta + \frac{\text{Cov}(x, u)}{\text{Var}(x)}$$

donc

$$E(b) = \beta \text{ seulement si } E\left(\frac{\text{Cov}(x, u)}{\text{Var}(x)}\right) = 0$$

Doit-être égal à 0 aussi non j'ai β + quelque chose et dans ce cas j'ai un biais

Rappel: estimateur non biaisé quand l'espérance est = à la vraie valeur. On a un estimateur non biaisé quand $b = \beta$.
Pour voir si un estimateur est non biaisé on prend l'espérance.

Partie 1 : économétrie

C. CAS DE FIGURES QUI VIOLENT CETTE HYPOTHÈSE

1. Endogénéité: biais de variables omises

On parle **d'endogénéité** lorsque le terme d'erreur (u) n'est pas distribué indépendamment de X (X est lié à u)

Le premier cas où l'on peut avoir de l'endogénéité est **OVB (biais de variables omises)**: oublie une variable importante et elle est liée à au moins une variable.

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

Manque une variable \Rightarrow la mauvaise spécification mène à un problème d'endogénéité.

On remplace toujours par la vraie relation.

$$\hat{Y} = b_1 + b_2 X_2$$

$$\begin{aligned} b_2 &= \frac{\text{Cov}(X_2, Y)}{\text{Var}(X_2)} = \frac{\text{Cov}(X_2, [\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u])}{\text{Var}(X_2)} \\ &= \frac{\text{Cov}(X_2, \beta_1) + \text{Cov}(X_2, \beta_2 X_2) + \text{Cov}(X_2, \beta_3 X_3) + \text{Cov}(X_2, u)}{\text{Var}(X_2)} \\ &= \frac{0 + \beta_2 \text{Cov}(X_2, X_2) + \beta_3 \text{Cov}(X_2, X_3) + \text{Cov}(X_2, u)}{\text{Var}(X_2)} \\ &= \beta_2 + \beta_3 \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)} + \frac{\text{Cov}(X_2, u)}{\text{Var}(X_2)} = \beta_2 + \beta_3 \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)} * \end{aligned}$$

$$\Rightarrow E(b_2) = \beta_2 \text{ seulement si } \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)} = 0$$

Quand j'omets une variable, elle va dans le terme d'erreur. X_2 est liée au terme d'erreur. Quand une variable est liée au terme d'erreur (u) on a un biais.

Si j'oublie une variable qui est liée à mon modèle, j'aurais un biais. Le biais sera d'autant plus grand que X_2 et X_3 sont corrélées. *

2. Erreurs de mesure

a) Erreurs de mesure dans une variable explicative

Supposons que nous avons : $y = \alpha + \beta z + v$. Supposons que Z ne peut pas être mesuré avec précision et que nous devons utiliser X pour désigner sa valeur mesurée. X est égal à la vraie valeur Z plus l'erreur de mesure w .

Nous substituons ensuite dans l'équation ci-dessous. L'équation aura alors 2 composantes aléatoires: le terme d'erreur original (v) et l'erreur de mesure multipliée par β (w). Ensemble ils forment un terme d'erreur composite que nous allons appeler u . Mais on se rend compte qu'à la fois X et u dépendent de w . **Donc u n'est pas distribué indépendamment de X (ils sont liés)**. On appelle ça un biais d'endogénéité. C'est très embêtant car on va avoir des **coefficients biaisés** et non **convergent**, une **inférence incorrecte**.

Partie 1 : économétrie

Je n'arrive pas à mesurer Z, donc je mesure X (je sais mesurer).

$$y = \alpha + \beta z + v \quad \leftarrow \text{Erreur}$$

$$x = z + w \quad \leftarrow \text{Erreur de mesure}$$

$$y = \alpha + \beta(x - w) + v$$

$$= \alpha + \beta x + v - \beta w$$

$$= \alpha + \beta x + u$$

\uparrow \uparrow
 w w

$$u = v - \beta w$$

On ne le connaît pas donc il va dans le terme d'erreur.

b) Erreur de mesure dans la variable dépendante

Une erreur de mesure dans la variable dépendante a des conséquences moins graves. La seule conséquence sera que les écarts types et les tests resteront valides mais les écart-types auront tendance à être plus grands.

Je ne sais pas mesurer. $q = \alpha + \beta x + v$

Je sais mesurer. $y = q + r$

X ne dépend pas de r

$$y - r = \alpha + \beta x + v$$

$$y = \alpha + \beta x + v + r$$

$$= \alpha + \beta x + u$$

$$u = v + r$$

Quelle serait la variance du terme d'erreur?

$$\frac{\sigma^2}{n\sigma^2}$$

Quand on ajoute de l'incertitude dans le terme d'erreur l'écart-type gonfle.

$$\sigma_b^2 = \frac{\sigma_u^2}{n\sigma_x^2} = \frac{\sigma_v^2 + \sigma_r^2}{n\sigma_x^2}$$

Somme des variances des deux termes d'erreurs

Variance de b_1 : $\sigma^2 (X'X)^{-1}$

3. Biais de simultanéité

On parle de **biais de simultanéité** lorsque (dans l'exemple) Y dépend de Z mais Z dépend de Y. Par exemple: les prix dépendent des quantités et les quantités dépendent du prix. Ou de causalité inverse, c'est-à-dire que l'on se trompe de relation.

$$y = \beta_1 X_1 + \gamma_1 z + u$$

$$z = \beta_2 X_1 + \gamma_2 y + v$$

Y dépend de Z et U. Z dépend de Y. Donc, comme Z dépend de Y qui lui-même dépend de U, Z est lié à u.

Z et u ne sont pas indépendants (ils sont liés).

$$y = \beta_1 X_1 + \gamma_1 z + u$$

$$z = \frac{\beta_2 + \gamma_2 \beta_1}{1 - \gamma_1 \gamma_2} X_1 + \frac{1}{1 - \gamma_1 \gamma_2} v + \frac{\gamma_2}{1 - \gamma_1 \gamma_2} u$$

* voir développement feuille joint syllabus.

$$E(z, u) = \frac{\gamma_2}{1 - \gamma_1 \gamma_2} E(u, u) \neq 0$$

Partie 1 : économétrie

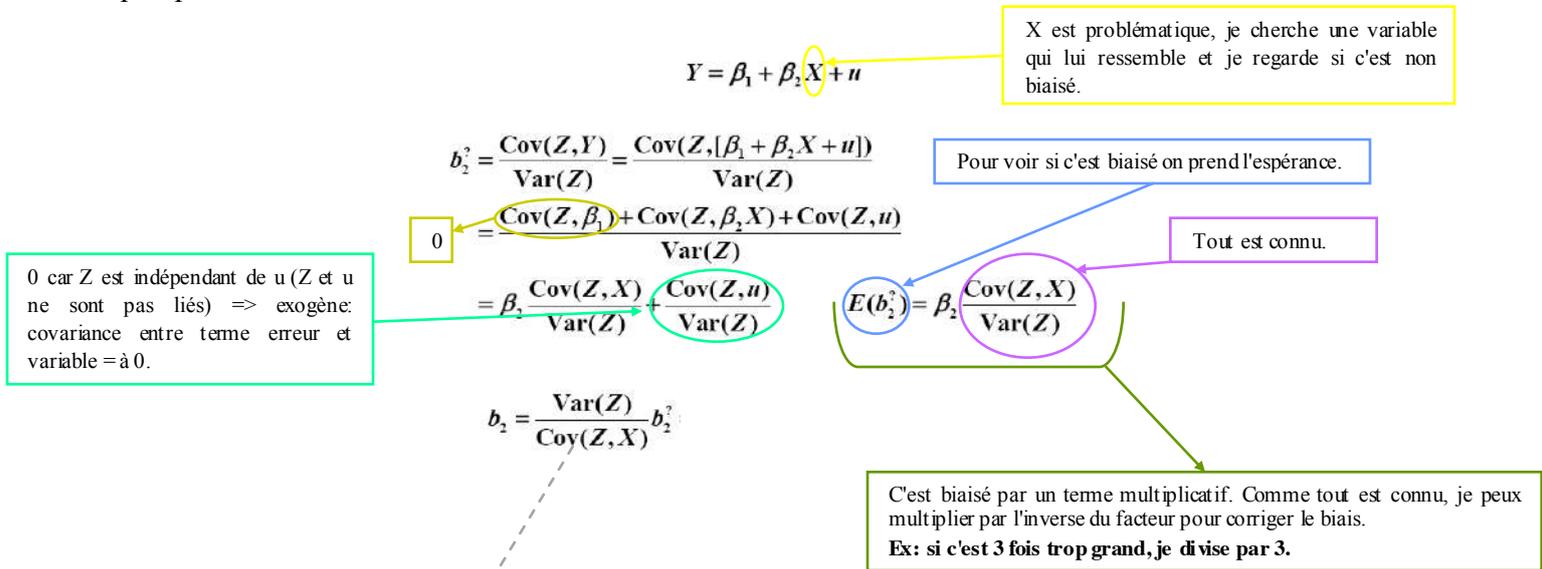
Résumé des 3 cas de biais d'endogénéité

- 1) Nous dirons qu'il y a un biais d'endogénéité lorsque :
 - a. On a un biais de variable omise (voir chapitre 5)
 - b. Erreur de mesure et dans ce cas $Cov(u_i, X_i) \neq 0$ (voir supra)
 - c. Biais de simultanéité ou de causalité inverse

D. VARIABLES INSTRUMENTALES

Par variable **instrumentales** nous entendons une variable qui ressemble à une variable qui nous intéresse mais qui n'est pas liée au terme d'erreur.

Supposons qu'une variable Z soit liée à X mais pas à u . Z agit comme un proxy de X . Nous démontrons que les estimations résultantes seront biaisées mais nous sommes capables d'y apporter quelques améliorations.



En effet nous pouvons neutraliser le biais en multipliant b_2^OLS par $\frac{Var(Z)}{Cov(Z, X)}$.

$$= \beta_2 \frac{Cov(Z, X)}{Var(Z)} + \frac{Cov(Z, u)}{Var(Z)} \quad E(b_2^OLS) = \beta_2 \frac{Cov(Z, X)}{Var(Z)}$$

$$b_2 = \frac{Var(Z)}{Cov(Z, X)} b_2^OLS = \frac{Var(Z)}{Cov(Z, X)} \frac{Cov(Z, Y)}{Var(Z)} = \frac{Cov(Z, Y)}{Cov(Z, X)}$$

$\frac{Cov(Z, Y)}{Cov(Z, X)}$ est le nouveau **estimateur de variable instrumentale** de (b_2) . Il permet de résoudre certains problèmes associés à l'endogénéité. Cette estimateur est biaisé mais convergent.

Partie 1 : économétrie

1. Démonstration des propriétés de l'estimateur de variable instrumentale

$$Y = \beta_1 + \beta_2 X + u$$

Plus n augmente, plus le biais tend vers 0: la distribution se resserre – le biais devient plus petit et on se rapproche de la vraie valeur.

$$\begin{aligned} b_2^{IV} &= \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)} = \frac{\text{Cov}(Z, [\beta_1 + \beta_2 X + u])}{\text{Cov}(Z, X)} && \text{Biaisé mais convergent.} \\ &= \frac{\text{Cov}(Z, \beta_1) + \text{Cov}(Z, \beta_2 X) + \text{Cov}(Z, u)}{\text{Cov}(Z, X)} \\ &= \frac{\beta_2 \text{Cov}(Z, X) + \text{Cov}(Z, u)}{\text{Cov}(Z, X)} = \beta_2 + \frac{\text{Cov}(Z, u)}{\text{Cov}(Z, X)} \end{aligned}$$

Ne fonctionne que quand on a 1 variable problématique (X) et 1 instrument (Z) -> la covariance fonctionne one to one.

Si u, Z et u, X sont indépendants, on peut montrer que la valeur du terme d'erreur est égale à zéro. Mais u, X ne sont pas indépendants (ils sont liés). Ce qui fait qu'il y aura toujours un biais. Nous démontrons que b_2^{IV} est un estimateur convergent de β_2 . Si u est indépendant de Z , la $\text{Cov}(Z, u)$ tendra vers 0 en grand échantillon et b_2^{IV} tendra vers β_2 .

De plus nous voudrions que la variance de la population soit la plus petite possible. Cela signifie que nous voulons que la corrélation entre X et Z soit la plus grande possible.

$$Y = \beta_1 + \beta_2 X + u$$

$$b_2^{IV} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)}$$

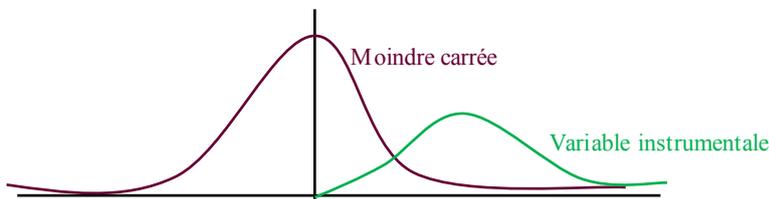
Si ce n'est pas endogène l'estimateur des variables instrumentales n'est pas efficace car: écart-type gonfle et l'inférence est nulle.

$$\text{variance de population de } b_2^{IV} = \sigma_{b_2^{IV}}^2 = \frac{\sigma_u^2}{n \text{Var}(X)} \times \frac{1}{\rho_{X,Z}^2}$$

Nous voudrions que la variance de la population soit la plus petite possible. Pour ce faire, la corrélation entre X et Z doit être grande.

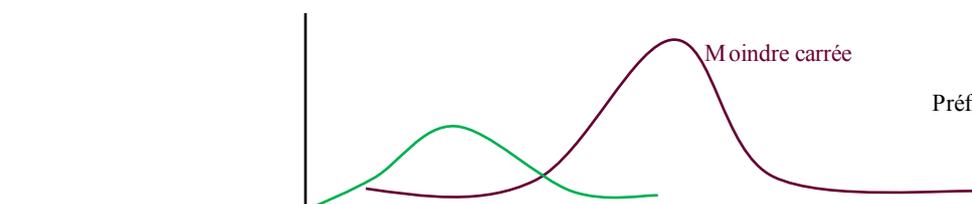
On utilise seulement l'estimateur de variable instrumentale seulement s'il y a de l'endogénéité.

Si pas d'endogénéité



Préfère le moindre carrée

Si endogénéité

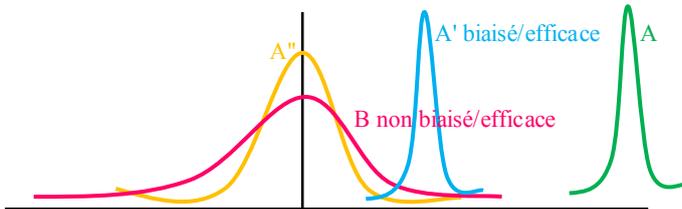


Préfère variable instrumentale

Partie 1 : économétrie

E. SPÉCIFICATION DU TEST DE DURBIN-WU-HAUSMAN

Exemple: Modélisation



Entre B et A'': je préfère A''.

Entre B et A: je préfère B car A est trop biaisé

Entre B et A': ce n'est pas évident, il va falloir trancher entre biais ou efficacité

Le test de Durbin-Wu-Hausman²⁴ va nous permettre de trancher entre biais ou efficacité.

L'idée du test est donc de comparer les coefficients. Si les:

- β sont les mêmes (sont pas différents) => je préfère moindres carrés.
- β sont différents => je préfère variable instrumentale.

La première étape consiste à estimer le modèle avec des estimateurs convergent (variables instrumentales) et à le sauver. En effet, il faut sauver le modèle puisqu'il faut comparer les coefficients.

La deuxième étape consiste à estimer l'autre modèle et à le sauver.

Modèle convergent

Modèle efficace et potentiellement non convergent. On suppose qu'il est endogène car s'il était exogène nous n'aurions pas besoin de la variable instrumentale, on choisirait directement le moindres carrés.

Différence dans les écart-types (donne une idée des différences dans l'efficacité).

--- Coefficients ---				
	(b) Prior	(B) Current	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
S	.1026815	.0618848	-.0407967	.0362637
ASVABC	.0025508	.0093287	-.0067779	.0060248
MALE	.2280404	.2130222	-.0150182	.0133495
ETHBLACK	-.15289	-.1019355	-.0509546	.045293
ETHHISP	.0463734	.0537519	-.0073784	.0065586
_cons	.7939315	1.009459	-.2155273	.1915801

Si toute la colonne = 0 alors je préfère l'efficace. (Moindres carrés).

b = less efficient estimates obtained previously from ivreg.
B = more efficient estimates obtained from regress.

Test: Ho: difference in coefficients not systematic

$\chi^2(1) = \frac{(b-B)'(V_b-V_B)^{-1}(b-B)}{1} = 1.27$

Prob>chi2 = 0.2606

$< 1.96^2 (3.94) \Rightarrow$ je compare la valeur avec la valeur critique d'une K_i^2 avec un degré de liberté (toujours un degré de liberté).

Je ne rejette pas l'H0 car $P > 0.05$: je suis prêts à renoncé à un peu de biais pour de l'efficacité.

L'H₀ nous dit que les β ne sont pas différents. Il n'y a pas violation de la 4^e condition de Gauss-Markov (exogénéité). Si l'H₀ est vraie, il ne faut utiliser moindres carrés car variable instrumentale est moins efficace.

Distribué comme une K_i^2 avec un degré de liberté. Comme β est une normale, la différence entre deux normales est une normale et de moyenne différence moyenne. De variance, somme des variances moins $2 * Cov.$ => tombe dans la différence des Var B.

²⁴ Voir syllabus pour l'exemple en entier, je détaille que l'essentiel.

Partie 1 : économétrie

Test de Durbin

Est-ce que le biais est compensé par l'efficacité:

- Biais > efficacité je préfère instrumentale
- Biais < efficacité je préfère moindres carrés.

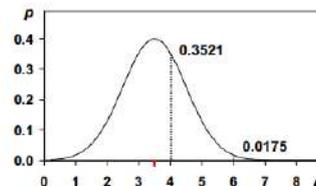
Chapitre 8: Introduction à l'estimation par maximum de vraisemblance²⁵

A. MAXIMUM DE VRAISEMBLANCE

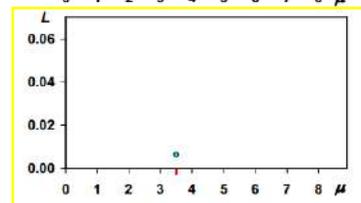
1. Graphiquement

Supposons que nous ayons une variable aléatoire X distribuée normalement avec une moyenne de population inconnue μ et un écart type σ , et que nous ayons un échantillon de 2 observations, 4 et 6. Pour l'instant, nous ferons l'hypothèse que σ est égal à 1.

Hypothèse initiale $\mu = 3.5$. Sous cette hypothèse la **densité de probabilité** à 4 serait 0.3521 et à 6, 0.0175. **La densité jointe (L)** en est le produit, 0.0062 (0.3521*0.0175).



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062

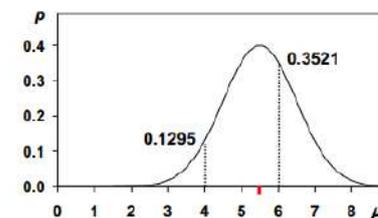


Graphique de densité de probabilité jointe

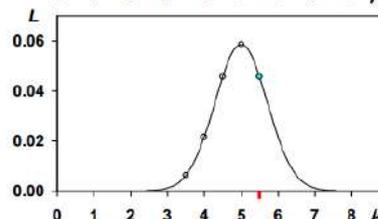
La probabilité jointe va nous décrire la vraisemblance d'observer notre échantillon s'il y a indépendance. Le but du jeu va être ici de regarder à la distribution (ou plus particulièrement à la valeur du paramètre) qui a la plus grande probabilité de générer mon échantillon.

La vraisemblance consiste donc à chercher les caractéristiques de la distribution de la population qui avec la plus grande probabilité a généré mon échantillon.

Maintenant supposons que $\mu=4$ et nous refaisons la même chose...jusqu'à $\mu=5.5$ et on finit par avoir une fonction complète de probabilité jointe. On voit que le pic se situe à $\mu=5$. C'est donc la valeur moyenne de la population qui aurait pu, avec la plus grande vraisemblance, générer les valeurs de 4 et 6 pour X .



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456
5.0	0.2420	0.2420	0.0585
5.5	0.1295	0.3521	0.0456



Je regarde le paramètre μ pour lequel la vraisemblance est plus grande.

²⁵ C'est une autre façon d'estimer les paramètres.

Partie 1 : économétrie

2. Mathématiquement

Prenons maintenant la fonction de densité de X. X est normalement distribué avec une moyenne μ et un écart type σ . L'hypothèse actuelle est que $\sigma=1$. Reprenons notre exemple d'échantillon avec 2 observations: 4 et 6. Nous allons prendre leur **probabilité jointe (produit de leur densité individuelle)**.

Fonction de densité d'une normale (permet de trouver le point => calculé la hauteur des points).

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} = 0.3521$$

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2} \rightarrow \text{Densité observée.}$$

$$f(4) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \quad f(6) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}$$

$$\text{densité jointe} = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

μ doit être trouvé afin qu'il maximise la densité de probabilité jointe. Pour cela, il faut dériver par rapport à μ et je recherche la valeur pour laquelle la vraisemblance est maximum.

Dans notre estimation du maximum de vraisemblance, nous choisissons comme estimation de μ la valeur qui donne la plus grande densité jointe pour les observations de notre échantillon. Cette valeur nous donne la plus grande probabilité, au maximum de vraisemblance, d'obtenir les observations dans l'échantillon.

Nous allons maintenant chercher μ qui maximise l'expression de la vraisemblance. Nous écrivons la fonction de vraisemblance de μ : $L(\mu|4,6)$. Nous allons transformer l'expression en somme pour que cela soit plus simple à dériver (pour cela il faut passer par le log).

Note: Log L est fonction monotonique croissante de L (c'est-à-dire que Log L augmente si L augmente et diminue si L décroît). De ce fait, la valeur de μ qui maximise le Log L est la même que celle qui maximise L.

$$L(\mu | 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right) \right]$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

Somme des Log

$$= \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(6-\mu)^2} \right)$$

Quand je travail avec le Log, la valeur change mais l'endroit où le maximum à lieu ne change pas.

Ex: 1000 est > que 100

Log 1000 = 3 > Log 100 = 2

Partie 1 : économétrie

$$\log a^b = b \log a$$

$$\begin{aligned} \log e^{-\frac{1}{2}(X-4)^2} &= -\frac{1}{2}(X-4)^2 \log e \\ &= -\frac{1}{2}(X-4)^2 \end{aligned}$$

$$\begin{aligned} &= \log\left(\frac{1}{\sqrt{2\pi}}\right) + \log\left(e^{-\frac{1}{2}(4-\mu)^2}\right) + \log\left(\frac{1}{\sqrt{2\pi}}\right) + \log\left(e^{-\frac{1}{2}(6-\mu)^2}\right) \\ &= 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2 \end{aligned}$$

$$-\frac{1}{2}(a-\mu)^2 = -\frac{1}{2}(a^2 - 2a\mu + \mu^2) = -\frac{1}{2}a^2 + a\mu - \frac{1}{2}\mu^2$$

$$\frac{d}{d\mu} \left\{ -\frac{1}{2}(a-\mu)^2 \right\} = a - \mu$$

$$\frac{d \log L}{d\mu} = (4-\mu) + (6-\mu)$$

$$\frac{d \log L}{d\mu} = 0 \Rightarrow \hat{\mu} = 5$$

CPO confirme que 5 est la valeur de μ qui maximise le Log L et donc la fonction de vraisemblance elle-même. Nous parlons désormais d'estimation de μ et non plus de sa vraie valeur.

3. Généralisation

Nous allons maintenant généraliser ce résultat à un échantillon de n observations X_1, \dots, X_n .

$$f(X_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i-\mu)^2}$$

$$L(\mu | X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_1-\mu)^2} \right) \times \dots \times \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_n-\mu)^2} \right)$$

Probabilité jointe d'observer les observations de $X_1 \dots X_n$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_1-\mu)^2} \right) \times \dots \times \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_n-\mu)^2} \right) \right]$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_1-\mu)^2} \right) + \dots + \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_n-\mu)^2} \right)$$

$$n * \frac{1}{\sqrt{2\pi}}$$

$$= n \log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2}(X_1 - \mu)^2 - \dots - \frac{1}{2}(X_n - \mu)^2$$

Log transforme la valeur de mes fonctions mais pas l'endroit du maximum de μ

Partie 1 : économétrie

$$\log L = n \log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2}(X_1 - \mu)^2 - \dots - \frac{1}{2}(X_n - \mu)^2$$

Valeur théorique de μ

$$\frac{d \log L}{d\mu} = (X_1 - \mu) + \dots + (X_n - \mu)$$

Valeur estimée de μ

C'est à ce moment là que le chapeau apparaît car on cherche le maximum.

$$\frac{d \log L}{d\mu} = 0 \Rightarrow \sum X_i - n\hat{\mu} = 0$$

$$\therefore \hat{\mu} = \frac{1}{n} \sum X_i = \bar{X}$$

Nous voyons que l'estimateur du maximum de vraisemblance de la moyenne population est la moyenne de l'échantillon. Si la distribution de la population est normalement distribuée et d'écart-type égal à 1.

4. Estimateur du maximum de vraisemblance quand μ et σ sont inconnus

Relâchons à présent l'hypothèse que $\sigma=1$. Donc, la variance n'est pas connue, on devra estimer μ et σ . Nous avons tout de même une distribution normale.

$$f(X_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

$$L(\mu, \sigma | X_1, \dots, X_n) = \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1 - \mu}{\sigma}\right)^2} \right) \times \dots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n - \mu}{\sigma}\right)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1 - \mu}{\sigma}\right)^2} \right) \times \dots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n - \mu}{\sigma}\right)^2} \right) \right]$$

$$= \log \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1 - \mu}{\sigma}\right)^2} \right) + \dots + \log \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n - \mu}{\sigma}\right)^2} \right)$$

$$= n \log \left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{1}{2} \left(\frac{X_1 - \mu}{\sigma} \right)^2 - \dots - \frac{1}{2} \left(\frac{X_n - \mu}{\sigma} \right)^2$$

$$= n \log \left(\frac{1}{\sigma} \right) + n \log \left(\frac{1}{\sqrt{2\pi}} \right) + \frac{1}{\sigma^2} \left(-\frac{1}{2}(X_1 - \mu)^2 - \dots - \frac{1}{2}(X_n - \mu)^2 \right)$$

Partie 1 : économétrie

Dérivé par rapport à μ

$$\log L = n \log\left(\frac{1}{\sigma}\right) + n \log\left(\frac{1}{\sqrt{2\pi}}\right) + \frac{1}{\sigma^2} \left(-\frac{1}{2}(X_1 - \mu)^2 - \dots - \frac{1}{2}(X_n - \mu)^2\right)$$

$$\begin{aligned} \frac{\partial \log L}{\partial \mu} &= \frac{1}{\sigma^2} \frac{\partial}{\partial \mu} \left(-\frac{1}{2}(X_1 - \mu)^2 - \dots - \frac{1}{2}(X_n - \mu)^2\right) \\ &= \frac{1}{\sigma^2} [(X_1 - \mu) + \dots + (X_n - \mu)] \end{aligned}$$

Je multiplie par σ^2 pour m'en débarrasser.

$$= \frac{1}{\sigma^2} (\sum X_i - n\mu)$$

$$\frac{\partial \log L}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \bar{X}$$

Nous voyons que l'estimateur du maximum de vraisemblance de la moyenne population est la moyenne de l'échantillon. Quelle que soit la valeur de σ quand la distribution de la population est normalement distribuée.

Dérivé par rapport à σ

$$\begin{aligned} \log L &= n \log\left(\frac{1}{\sigma}\right) + n \log\left(\frac{1}{\sqrt{2\pi}}\right) + \frac{1}{\sigma^2} \left(-\frac{1}{2}(X_1 - \mu)^2 - \dots - \frac{1}{2}(X_n - \mu)^2\right) \\ &= -n \log \sigma + n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2} \sum (X_i - \mu)^2 \end{aligned}$$

$$\log a^b = b \log a$$

$$\log \frac{1}{\sigma} = \log \sigma^{-1} = (-1) \log \sigma = -\log \sigma$$

$$\begin{aligned} \log L &= n \log\left(\frac{1}{\sigma}\right) + n \log\left(\frac{1}{\sqrt{2\pi}}\right) + \frac{1}{\sigma^2} \left(-\frac{1}{2}(X_1 - \mu)^2 - \dots - \frac{1}{2}(X_n - \mu)^2\right) \\ &= -n \log \sigma + n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2} \sum (X_i - \mu)^2 \end{aligned}$$

Valeur théorique de σ

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \sigma^{-3} \sum (X_i - \mu)^2$$

Valeur estimée de

$\sigma \Rightarrow \hat{\sigma}$

$$\frac{\partial \log L}{\partial \sigma} = 0 \Rightarrow -\frac{n}{\hat{\sigma}} + \hat{\sigma}^{-3} \sum (X_i - \hat{\mu})^2 = 0$$

$$\therefore -n\hat{\sigma}^2 + \sum (X_i - \bar{X})^2 = 0$$

Sous l'hypothèse de normalité

$$\therefore \hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2 = \text{Var}(X)$$

Moyenne d'échantillon

Reste un estimateur biaisé.

Partie 1 : économétrie

Nous observons que l'estimateur du maximum de vraisemblance pour la variance de la population est la variance d'échantillon.

L'estimateur obtenu est biaisé. L'estimateur non biaisé est obtenu en divisant par $(n-1)$ au lieu de n . Notons que le biais s'atténue au fur et à mesure que l'échantillon s'agrandit.

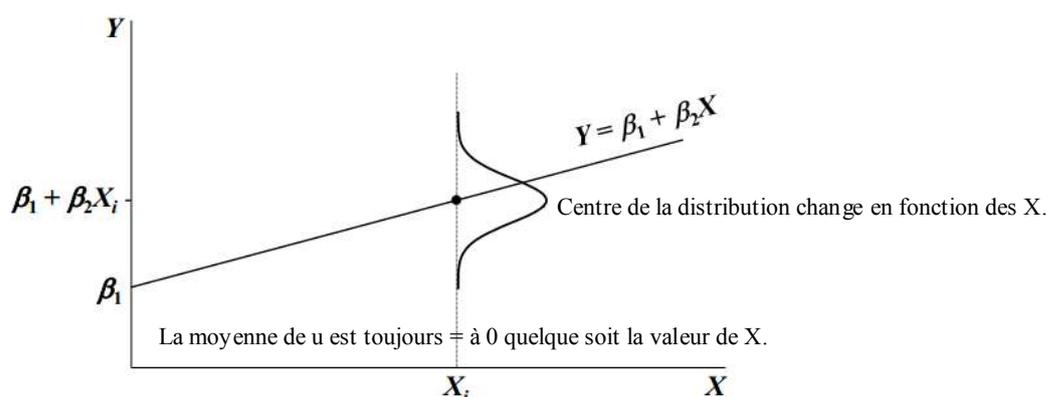
5. Estimation des coefficients de régression par maximum de vraisemblance

Revenons à notre équation de régression simple $Y = \beta_1 + \beta_2 X + u$ et supposons que le terme d'erreur soit **normalement distribué, de moyenne 0** (quand le point est sur la droite) et que sa **variance soit identique** partout (homoscédasticité).

Par rapport au point, la courbe représente la distribution de u , sa distribution potentielle avant que l'observation ne soit générée. Ex post, bien entendu, elle est fixée à une certaine valeur.

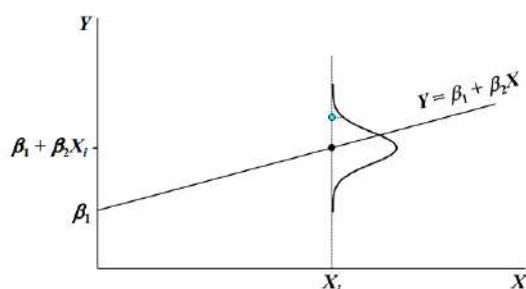
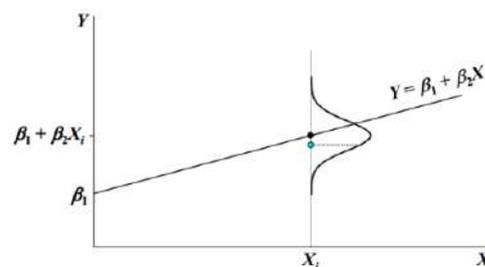
Par rapport à l'axe horizontal, la courbe représente aussi la **distribution de Y** pour cette observation, c'est-à-dire, **conditionnellement à $X=X_i$** .

La distribution de Y dépend de X: moyenne conditionnelle car elle dépend de la valeur des X. Et non inconditionnelle: ce n'est pas un bête chiffre.



Dire que u est normalement distribué de moyenne 0 c'est la même chose de dire que Y est normalement distribué de moyenne $\beta_1 + \beta_2 X_i$.

Les valeurs potentielles de Y proches de $\beta_1 + \beta_2 X_i$ auront une densité relativement grande.

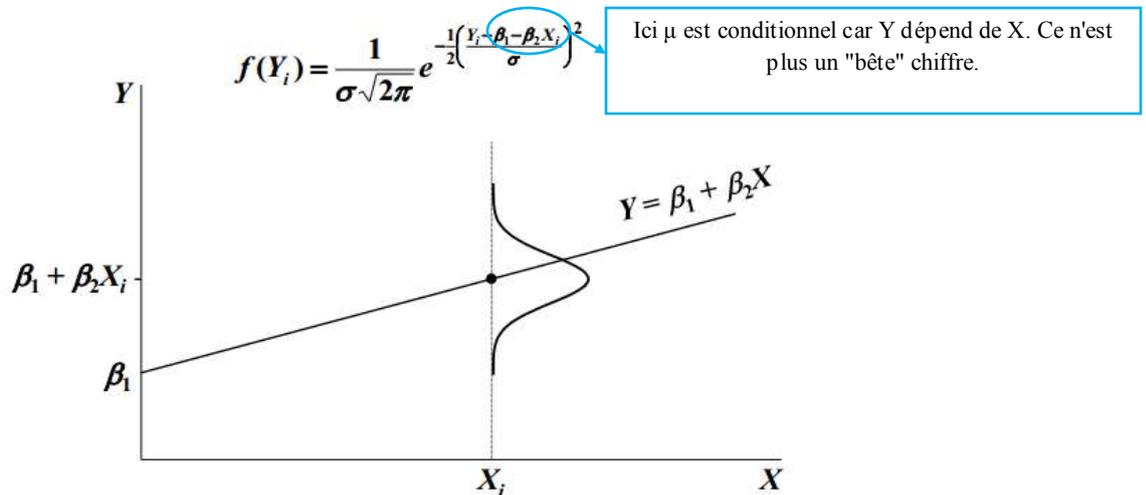


Tandis que les valeurs potentielles de Y relativement éloignées de $\beta_1 + \beta_2 X_i$ auront une densité petite.

Partie 1 : économétrie

La valeur moyenne de la distribution de Y_i est $\beta_1 + \beta_2 X_i$. Son écart type est σ , l'écart type du terme d'erreur.

Nous pouvons donc dire que de par la distribution normale du terme d'erreur, on sait qu' Y aura une distribution normale de moyenne $\beta_1 + \beta_2 X_i$ (moyenne du terme d'erreur = 0) et conditionnelle à X_i .



a) Mathématiquement

Cherchons à présent les estimateurs du maximum de vraisemblance pour les coefficients de régression. Pour cela nous allons prendre la fonction de densité jointe pour les observations de Y qui n'est autre que le produit de leur densité individuelle.

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_i - \beta_1 - \beta_2 X_i}{\sigma}\right)^2}$$

Paramètres inconnus.

$$f(Y_1) \times \dots \times f(Y_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2} \times \dots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2}$$

Je calcule la densité associée à chaque $Y \Rightarrow$ J'estime la valeur de Y pour chaque X .

$$L(\beta_1, \beta_2, \sigma | Y_1, \dots, Y_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2} \times \dots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2}$$

$$\log L = \log \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2} \times \dots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2} \right)$$

Partie 1 : économétrie

$$\begin{aligned} &= \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2}\right) + \dots + \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2}\right) \\ &= n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2 - \dots - \frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2 \\ &= n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2} Z \\ &\text{où } Z = \left[(Y_1 - \beta_1 - \beta_2 X_1)^2 + \dots + (Y_n - \beta_1 - \beta_2 X_n)^2 \right] \end{aligned}$$

Max la vraisemblance et minimiser le carré des erreurs c'est la même chose.

Nous voyons que pour maximiser le logarithme de la vraisemblance, nous devons minimiser Z . Mais choisir des estimateurs de β_1 et β_2 qui minimisent Z , c'est ce que nous avons fait lorsque nous dérivions des coefficients de régression dans la technique des moindres carrés. Donc, pour ce modèle de régression, les estimateurs du maximum de vraisemblance sont identiques aux estimateurs des moindres carrés sous les hypothèses de Gauss-Markov.

! Sous quelle(s) condition(s) le maximum de vraisemblance est égal au minimum des moindres carrés?

Les hypothèses de Gauss-Markov doivent être respectées

- H_1 : erreur doit être égale à 0
- H_2 : erreur doit être normale car si u n'est pas normale y n'est pas normal
- H_3 : σ ne doit pas avoir d'indice
- H_5 : erreur ne doit pas être corrélée car la densité jointe est égale au produit des densités individuelles si on a indépendance aussi non la densité jointe n'est pas le produit des densités individuelles.

Partie 1 : économétrie

Chapitre 9: Modèle de choix binaire : analyses linéaire, logit et probit

A. MODÈLE DE CHOIX BINAIRE: MODÈLE DE PROBABILITÉ LINÉAIRE

1. Modèle de probabilité linéaire

Les variables indépendantes ne sont plus continues mais sont discrètes. Dans le modèle de choix binaire, la variable indépendante Y , peut prendre **deux valeurs (0 ou 1)**. Prenant la valeur 1 si l'événement se passe et 0 sinon.

Un modèle de choix binaire est le modèle de **probabilité linéaire** qui indique la probabilité que l'événement ait lieu, p , est supposé une **fonction linéaire d'un ensemble de variables explicatives**. En somme, nous observons **comment un changement dans les variables X affecte la probabilité d'observer $Y=1$ ou $Y=0$** ²⁶.

Illustration:

Construisons une variable GRAD qui est égal à 1 si l'individu a obtenu un diplôme supérieur, 0 sinon.

```
. g GRAD=0
```

```
. replace GRAD=1 if S>11  
(523 real changes made)
```

Comment une variation de X affecte la probabilité d'observer $Y=1$ ou $Y=0$.

```
. reg GRAD ASVABC
```

Source	SS	df	MS	Number of obs =	570
Model	7.13422753	1	7.13422753	F(1, 568) =	112.59
Residual	35.9903339	568	.063363264	Prob > F =	0.0000
Total	43.1245614	569	.07579009	R-squared =	0.1654
				Adj R-squared =	0.1640
				Root MSE =	.25172

Quand l'intelligence augmente de 1, la probabilité d'obtenir un diplôme supérieur est de 0.012 soit 1.2 %.

GRAD	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ASVABC	.0121518	.0011452	10.611	0.000	.0099024 .0144012
_cons	.3081194	.0583932	5.277	0.000	.1934264 .4228124

La constante n'a pas de signification. Elle indique qu'à intelligence 0, la probabilité d'obtenir un diplôme supérieur est de 30 %.

Comme d'habitude, la valeur de la variable dépendante, Y_i dans l'observation i a une composante non stochastique et une composante aléatoire. La composante non stochastique dépend de X_i et des paramètres. La composante aléatoire est le terme d'erreur.

La composante non stochastique dans l'observation i est sa valeur espérée pour cette observation. Elle est très simple à calculer car elle ne peut prendre que 2 valeurs. Elle sera 1 avec une probabilité p_i et 0 avec une probabilité $(1-p_i)$. La valeur attendue dans l'observation i est de ce fait $\beta_1 + \beta_2 X_i$.

²⁶ Voir exemple dans le syllabus. Je détaille que l'essentiel.

Partie 1 : économétrie

Y est composé: Y moyen "pour chaque valeur de X" + terme d'erreur.

$$E[Y_i | X_i] + u_i$$

$$p_i = p(Y_i = 1) = \beta_1 + \beta_2 X_i$$

$$Y_i = E(Y_i) + u_i$$

$$E(Y_i) = 1 \times p_i + 0 \times (1 - p_i) = p_i = \beta_1 + \beta_2 X_i$$

$$E(Y) = \sum_x Y \cdot p(Y) = p(Y_i = 1)$$

→ L'espérance de Y n'est rien d'autre que la probabilité que Y=1

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

a) Quelles sont les limites liées au modèle de probabilité linéaire

Premièrement, il y a un problème avec **le terme d'erreur** qui n'est pas distribué normalement (écart-type et testes statistique invalide) et de ce fait, Y n'on plus. **Mais si n augmente c'est gérable** (je peux estimer les β même si u n'est pas normalement distribué.

$$p_i = p(Y_i = 1) = \beta_1 + \beta_2 X_i$$

$$Y_i = E(Y_i) + u_i$$

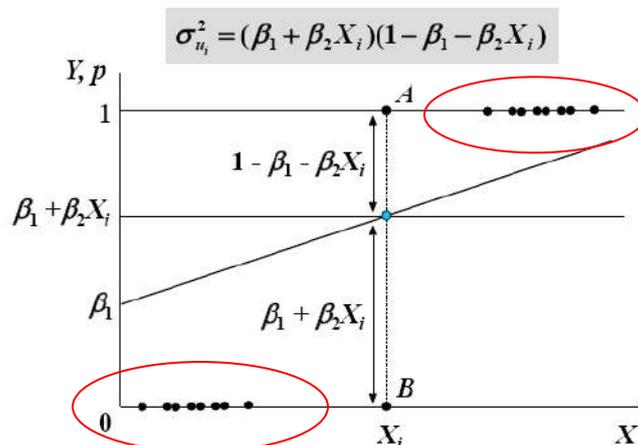
$$E(Y_i) = 1 \times p_i + 0 \times (1 - p_i) = p_i = \beta_1 + \beta_2 X_i$$

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

On aura deux blocs d'erreur. Dans l'observation i, pur que Y_i soit 1, u_i doit être $(1 - \beta_1 - \beta_2 X_i)$. Pour que Y_i soit 0, u_i doit être $(-\beta_1 - \beta_2 X_i)$.

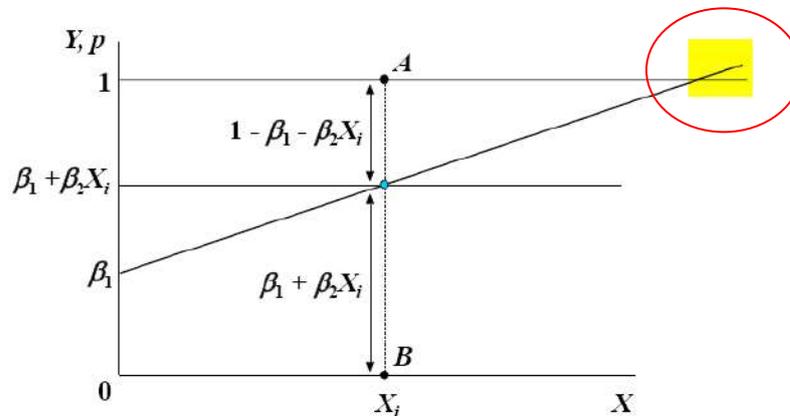
$$\left\{ \begin{array}{l} Y_i = 1 \Rightarrow u_i = 1 - \beta_1 - \beta_2 X_i \\ Y_i = 0 \Rightarrow u_i = -\beta_1 - \beta_2 X_i \end{array} \right.$$

Un second problème est que **la variance change avec X_i** (quand X augmente, la distribution augmente). Il y a donc de l'hétérosédasticité. Mais à nouveau **c'est gérable car on peut la modéliser ou la traiter** (exemple: matrice sandwich).



Partie 1 : économétrie

Un troisième et dernier problème apparaît. Celui-ci est beaucoup plus embêtant que les deux autres. Il consiste dans le fait que nous pourrions avoir des **probabilités supérieures à 1** ainsi que des **probabilités inférieures à 0**. C'est embêtant car on ne **sait pas résoudre**.



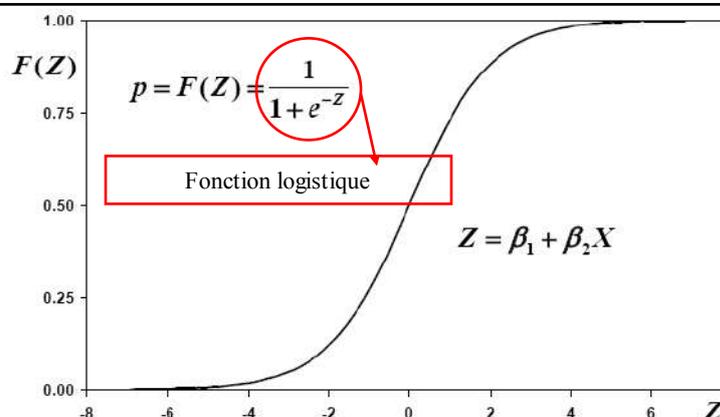
B. ANALYSE LOGIT

La manière habituelle d'éviter le 3^e problème vu supra est de faire l'hypothèse que **la probabilité est une fonction sigmoïde** de Z (de forme S), $F(Z)$, d'asymptotes horizontales en 1 et en 0, où Z est une fonction des variables explicatives. Nous utiliserons ici la fonction dite fonction logistique.

Une propriété est la suite: lorsque Z tend vers $+\infty$, e^{-Z} tend vers 0 et p tend vers 1 (**mais ne peut excéder 1**). Inversement, lorsque Z tend vers $-\infty$, e^{-Z} tend vers ∞ et p tend vers 0 (**mais ne peut être inférieur à 0**).

Le modèle implique que pour des valeurs de $Z < -2$, la **probabilité** que l'événement se produise est **faible** et **insensible** à des variations de Z . De la même manière, pour des **valeurs > 2** , la **probabilité** que l'événement se produise est **élevée** et **insensible** à des variations de Z .

Estime un modèle non linéaire (sigmoïde) avec une asymptote en 1 et 0 que je ne vais jamais dépasser.



Partie 1 : économétrie

Nous allons voir comment le modèle est estimé en utilisant des estimations par maximum de vraisemblance. Prenons $Z = \beta_1 + \beta_2 ASVABC$.

$$p = F(Z) = \frac{1}{1 + e^{-Z}} \quad Z = \beta_1 + \beta_2 ASVABC$$

$$= \frac{1}{1 + e^{-\beta_1 - \beta_2 ASVABC}}$$

Individus diplômés: le résultat de probabilité est

$$\frac{1}{1 + e^{-\beta_1 - \beta_2 ASVABC_i}} \quad F(Z) = \text{probabilité d'être diplômé.}$$

Individus non diplômés: le résultat de probabilité est

$$1 - \frac{1}{1 + e^{-\beta_1 - \beta_2 ASVABC_i}} \quad 1 - F(Z) = 1 - \text{probabilité d'être diplômé.}$$

Nous allons donc chercher les b_1 et b_2 de manière à maximiser la probabilité jointe. Nous allons donc prendre le produit des probabilités individuelles. Les b_1 et b_2 doivent être déterminé de manière numérique²⁷.

Maximiser $F(Z_1) \times \dots \times F(Z_s) \times [1 - F(Z_{s+1})] \times \dots \times [1 - F(Z_n)]$

$$\frac{1}{1 + e^{-b_1 - b_2 ASVABC_1}} \times \dots \times \frac{1}{1 + e^{-b_1 - b_2 ASVABC_s}}$$

$$\times \left(1 - \frac{1}{1 + e^{-b_1 - b_2 ASVABC_{s+1}}} \right) \times \dots \times \left(1 - \frac{1}{1 + e^{-b_1 - b_2 ASVABC_n}} \right)$$

Diplômé

$$\frac{1}{1 + e^{-\beta_1 - \beta_2 ASVABC_i}}$$

Non diplômé

$$1 - \frac{1}{1 + e^{-\beta_1 - \beta_2 ASVABC_i}}$$

Cherche les b_1 et b_2 qui maximise la probabilité jointe.

Note:

- Variable continue: on parle de densité
 - Maximum de vraisemblance: calcule la densité pour chaque Y
- Variable discrète: on parle de probabilité
 - Maximum de vraisemblance: calcule la probabilité Y=1 ou Y=0

²⁷ Voir syllabus pour déterminer les b_1 et b_2 (développement de Taylor)

Partie 1 : économétrie

a) Effet marginal

Pour obtenir une expression de la sensibilité, nous dérivons $F(Z)$ par rapport à Z .

$$p = F(Z) = \frac{1}{1 + e^{-Z}}$$

$$\begin{aligned} \frac{dp}{dZ} &= \frac{(1 + e^{-Z}) \times 0 - 1 \times (-e^{-Z})}{(1 + e^{-Z})^2} \\ &= \frac{e^{-Z}}{(1 + e^{-Z})^2} \end{aligned}$$

$$Y = \frac{U}{V}$$

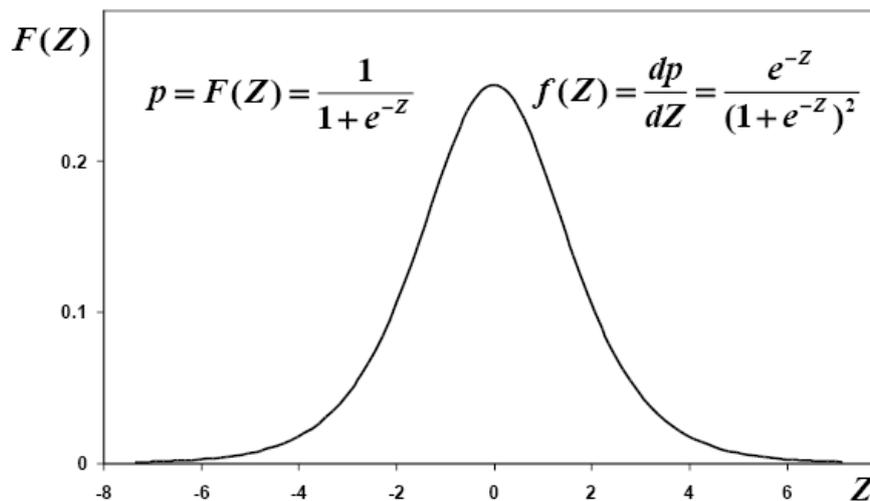
$$\frac{dY}{dZ} = \frac{V \frac{dU}{dZ} - U \frac{dV}{dZ}}{V^2}$$

$$U = 1 \Rightarrow \frac{dU}{dZ} = 0$$

$$V = (1 + e^{-Z})$$

$$\Rightarrow \frac{dV}{dZ} = -e^{-Z}$$

Suivie à la dérivée, nous pouvons observer comment la probabilité d'obtenir $Y=1$ change en fonction de Z . On peut voir que la sensibilité (pente) est la plus élevée lorsque $Z=0$. La fonction marginale, $f(Z)$ atteint un maximum à ce point.



Rappel! Dérivé: comment un changement dans X implique un changement dans Y .

Partie 1 : économétrie

Étant donnée que l'effet n'est pas le même partout le **coefficient ne peut être lu!** Seul son signe et l'inférence est interprétable.

```
. logit GRAD ASVABC
```

```
Iteration 0: Log Likelihood =-162.29468
Iteration 1: Log Likelihood =-132.97646
Iteration 2: Log Likelihood =-117.99291
Iteration 3: Log Likelihood =-117.36084
Iteration 4: Log Likelihood =-117.35136
Iteration 5: Log Likelihood =-117.35135
```

On ne sait pas lire les coefficients car maintenant cela dépend des X (exemple: si je fais rien pendant l'année, et que je travaille une heure de plus, cela ne changera rien dans mes résultats et inversement).

Logit Estimates

Si on est plus intelligent, on augmente le risque de réussir (signe positif).

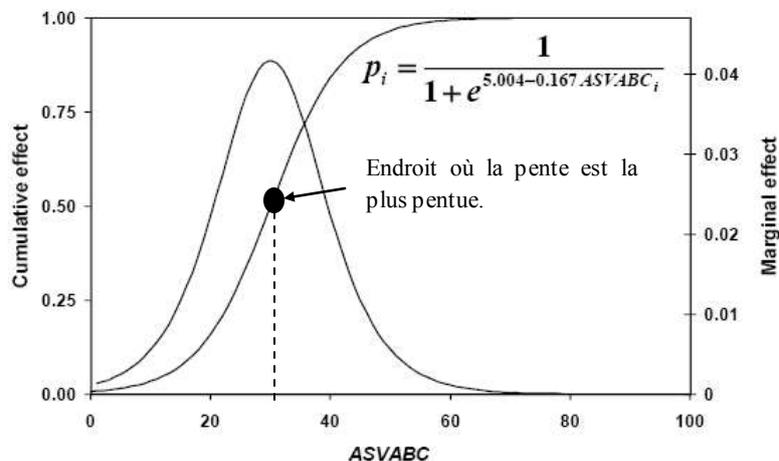
Log Likelihood = -117.35135

Number of obs = 570
chi2(1) = 89.89
Prob > chi2 = 0.0000
Pseudo R2 = 0.2769

GRAD	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
ASVABC	.1666022	.0211265	7.886	0.000	.1251951 .2080094
_cons	-5.003779	.8649213	-5.785	0.000	-6.698993 -3.308564

$$\hat{Z} = -5.004 + 0.167 ASVABC$$

Dans notre exemple, l'individu qui a une intelligence de 30 est celui dont l'effet d'une augmentation marginale de son intelligence aura le plus d'impact sur sa probabilité de réussir. De même un individu avec un score au-dessus de la moyenne (50) est presque certain d'être diplômé.



$$\hat{Z} = -5.004 + 0.167 ASVABC$$

Partie 1 : économétrie

Donc, les coefficients de la fonction Z n'ont pas d'interprétation intuitive directe. Pour pouvoir lire le coefficient il faut passer par **les effets marginaux**. Notons que p est une fonction de Z, et Z est une fonction des variables explicatives (X). L'effet marginal de X_i sur p peut être décrit comme le produit de l'effet marginal de Z sur p et l'effet marginal de X_i sur Z.

$$p = F(Z) = \frac{1}{1 + e^{-Z}}$$

$$Z = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Pour voir comment la probabilité d'obtenir $Y=1$ change en fonction de X, on regarde l'effet marginal: $f(Z) * \beta_i$. Là nous pouvons lire les coefficients.

Effet marginal de X_i sur Z est donné par son coefficient β

$$f(Z) = \frac{dp}{dZ} = \frac{e^{-Z}}{(1 + e^{-Z})^2}$$

Comment la probabilité d'obtenir $Y=1$ change en fonction de X.

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ} \frac{\partial Z}{\partial X_i} = f(Z) \beta_i = \frac{e^{-Z}}{(1 + e^{-Z})^2} \beta_i$$

Bêta associé à la dérivé.

On peut voir qu'au final **l'effet marginal n'est pas constant** puisqu'il dépend de la valeur de Z, qui en retour dépend des valeurs des variables explicatives (X). Notons que **f(Z) dépend de l'individu** (car X change d'un individu à un autre).

Une première façon est de l'évaluer à partir des moyennes d'échantillon des variables explicatives (**prendre un individu de référence**). On obtiendra donc un effet marginal évalué à la moyenne. Nous obtenons un effet très faible (0.005). La raison est qu'il est fort probable que n'importe qui ayant une intelligence moyenne soit diplômé. Donc une augmentation dans l'intelligence a peu d'effet.

. sum GRAD ASVABC

Variable	Obs	Mean	Std. Dev.	Min	Max
GRAD	570	.9175439	.2753	0	1
ASVABC	570	50.15088	9.214589	22	65

$$Z = \beta_1 + \beta_2 \bar{X} = -5.004 + 0.167 \times 50.15 = 3.371$$

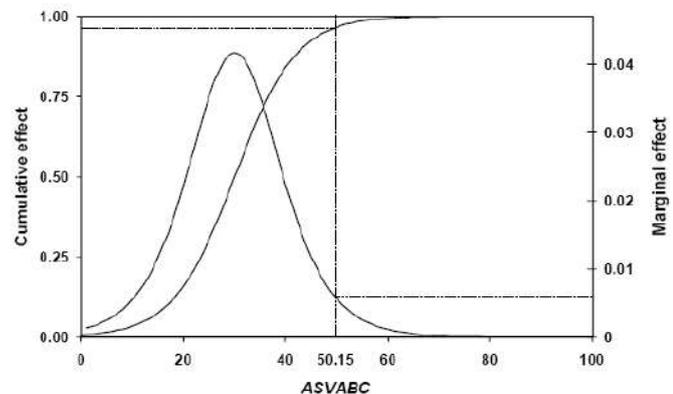
$$e^{-Z} = e^{-3.371} = 0.034$$

$$f(Z) = \frac{dp}{dZ} = \frac{e^{-Z}}{(1 + e^{-Z})^2} = \frac{0.034}{(1 + 0.034)^2} = 0.032$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ} \frac{\partial Z}{\partial X_i} = f(Z) \beta_i = 0.032 \times 0.167 = 0.005$$

Individu de référence (intelligence moyenne).

Pour l'individu de référence, augmenter l'intelligence de 1 unité, augmente la probabilité de réussir de 0.5 %



Partie 1 : économétrie

Pour montrer que l'effet marginal varie, nous le calculerons pour $ASVABC = 30$. Nous voyons via le qu'une augmentation d'un point de $ASVABC$ augmente la probabilité de réussir de 4.2 %. De plus on peut voir qu'avec une intelligence de 30, un individu a seulement une probabilité de 50 % d'être diplômé et une augmentation dans l'intelligence a un impact relativement grand.

```
. sum COLLEGE ASVABC
```

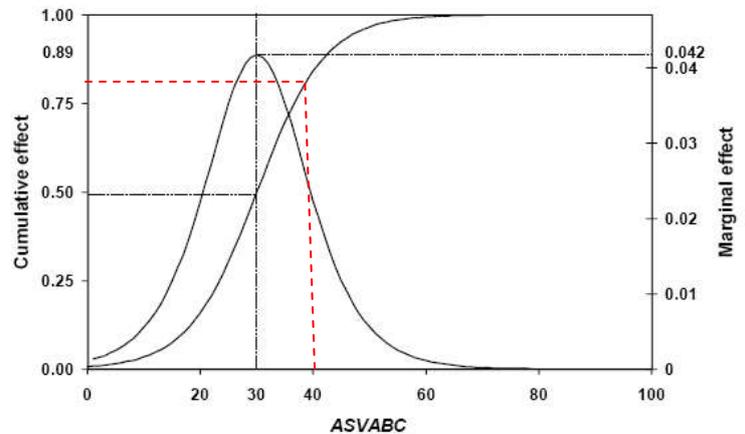
Variable	Obs	Mean	Std. Dev.	Min	Max
COLLEGE	570	.4964912	.5004269	0	1
ASVABC	570	50.15088	9.214589	22	65

$$Z = \beta_1 + \beta_2 X = -5.004 + 0.167 \times 30 = 0.006$$

$$e^{-Z} = e^{-0.006} = 0.994$$

$$f(Z) = \frac{dp}{dZ} = \frac{e^{-Z}}{(1+e^{-Z})^2} = \frac{0.994}{(1+0.994)^2} = 0.250$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ} \frac{\partial Z}{\partial X_i} = f(Z) \beta_i = 0.250 \times 0.167 = 0.042$$



Généralisation a plus d'une variable explicative

Quand on prend en compte plusieurs variables explicatives, nous pouvons remarquer que l'individu de référence n'existe pas.

```
. sum GRAD ASVABC SM SF MALE
```

Variable	Obs	Mean	Std. Dev.	Min	Max
GRAD	570	.9175439	.2753	0	1
ASVABC	570	50.15088	9.214589	22	65
SM	570	11.65263	2.561449	0	20
SF	570	11.81754	3.533178	0	20
MALE	570	.5701754	.4954857	0	1

Une deuxième façon de faire est de calculer l'effet marginal de chaque individu (qui existe) et de prendre la moyenne des effets marginaux.

Logit: Marginal Effects

	mean	b	product	f(Z)	f(Z)b
ASVABC	50.15	0.156	7.839	0.033	0.005
SM	11.65	0.065	0.753	0.033	0.002
SF	11.82	0.006	0.065	0.033	0.000
MALE	0.57	-0.279	-0.159	0.033	-0.009
Constant	1.00	-5.159	-5.159		
Total			-3.338		

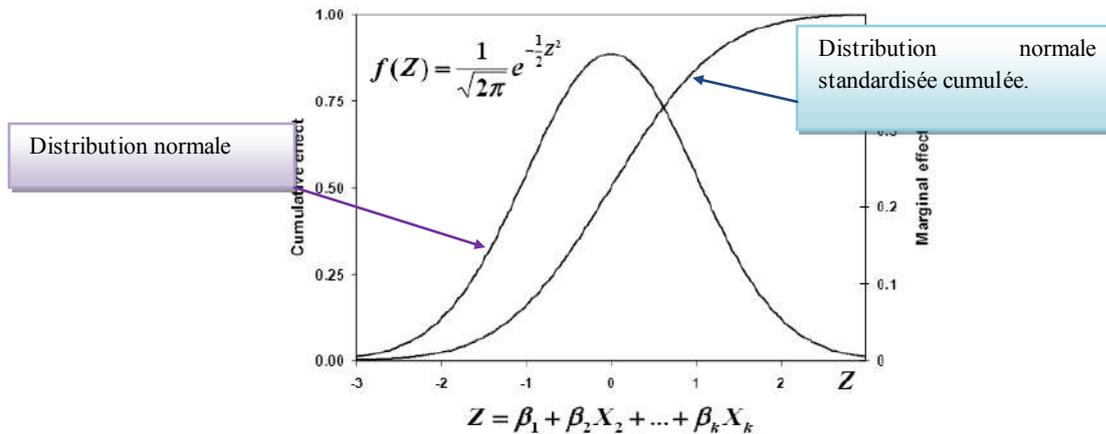
$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ} \frac{\partial Z}{\partial X_i} = f(Z) \beta_i$$

Effet marginal

Partie 1 : économétrie

C. ANALYSE PROBIT

Dans le cas d'analyse probit, la fonction sigmoïde est la distribution normale standardisée cumulée.



Le raisonnement est le même que pour l'analyse logit.

$$p = F(Z)$$

$$Z = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

$$f(Z) = \frac{dp}{dZ} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ} \frac{\partial Z}{\partial X_i} = f(Z) \beta_i = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \right) \beta_i$$

Effet marginal qui permet de lire les coefficients.

Nous pouvons tout comme pour l'analyse probit²⁸

- Prendre un individu de référence
- Calculer l'effet marginal de chaque individu et de prendre la moyenne des effets marginaux.

	Logit	Probit	Linear
	$f(Z)b$	$f(Z)b$	b
ASVABC	0.005	0.006	0.012
SM	0.002	0.003	0.007
SF	0.000	0.000	-0.001
MALE	-0.009	-0.014	-0.040

Lorsqu'on compare les résultats logit et probit, les coefficients dans les régressions sont très différents car la forme fonctionnelle est différente $F(Z)$. Néanmoins les effets marginaux sont habituellement similaires. L'avantage du modèle linéaire est qu'il est plus facile à estimer (il est souvent utilisé pour une première exploration des données).

²⁸ Voir syllabus pour exemple.

Partie 1 : économétrie

Chapitre 10 : Indépendance sérielle : autocorrelation, variable laguée (série chronologique), test de facteur commun et spécification de modèle dynamique

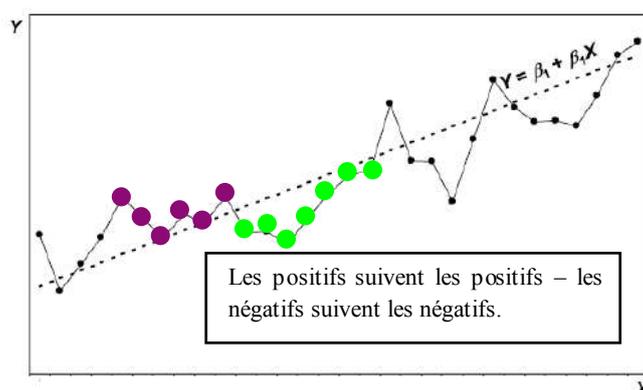
Afin d'analyser ce chapitre nous avons d'abord besoin de rappeler le concept d'indépendance sérielle qui est l'une des conditions de Gauss-Markov.

A. AUTOCORRELATION

L'indépendance sérielle se définit comme suit : $Cov(u_i, u_j) = 0$. Autrement dit, les termes d'erreurs doivent être générés indépendamment les uns des autres aussi non, nous avons un problème d'inférence.

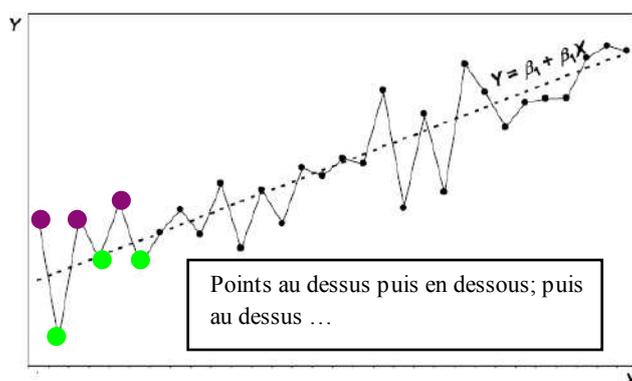
a) Dépendance sérielle (autocorrélation) positive

On viole l'hypothèse: les valeurs positives ont tendance à être suivies par des positives et les valeurs négatives par des négatives.



b) Dépendance sérielle (autocorrélation) négative

On viole l'hypothèse: les valeurs positives sont suivies par des négatifs qui sont suivies par des positifs.... On observe un phénomène d'alternance directe.



Partie 1 : économétrie

Un cas particulier d'autocorrélation est l'autocorrélation autorégressive de premier ordre, notée **AR(1)** (**On se concentre sur cette catégorie**). On dit qu'elle est autorégressive car, **u_t dépend de ses propres valeurs passées**, et de premier ordre car elle **dépend seulement de sa valeur précédente** (erreur est liée à celle d'avant). u_t dépend aussi de ε_t , une injection d'un **caractère aléatoire** nouvellement **apparue au temps t**, souvent appelée comme **l'innovation au temps t**.

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

→ Temps et non plus d'individus.

Autocorrélation autorégressive d'ordre 1: AR(1)

Car on régresse u_t sur u_{t-1} . Elle dépend d'elle-même (l'erreur d'aujourd'hui est = à l'erreur d'hier).

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Car u_t dépend seulement de u_{t-1} (son erreur d'hier).

Autocorrélation autorégressive d'ordre 5: AR(5)

Erreur d'aujourd'hui dépend: erreur d'hier, avant-hier....

On régresse u_t sur lui-même et 5 retards!

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \rho_5 u_{t-5} + \varepsilon_t$$

Peut être positif ou négatif

Équivalent à l'erreur. Il apparaît directement. Il est lié à rien.

L'autre principal type d'autocorrélation est la moyenne mobile où le terme d'erreur est une combinaison linéaire de l'innovation actuelle et un nombre fini d'innovations antérieures. Nous ne nous y attarderons pas ici!

Moyenne mobile d'ordre 3: MA(3)

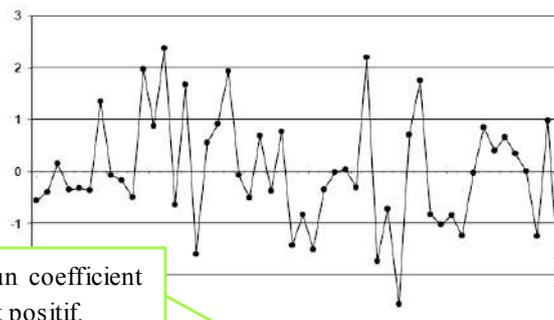
$$u_t = \lambda_0 \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \lambda_3 \varepsilon_{t-3}$$

u_t dépend uniquement de l'innovation.

Exemples lorsque le terme d'erreur est sujet à de l'autocorrélation AR(1)

Prenons 50 valeurs de ε indépendantes, prises d'une distribution normale ayant une moyenne de 0. Séries générées pour u en utilisant différentes valeurs de ρ .

Nous commençons avec un $\rho = 0$, il n'y a donc pas d'autocorrélation. Nous augmentons ρ progressivement.

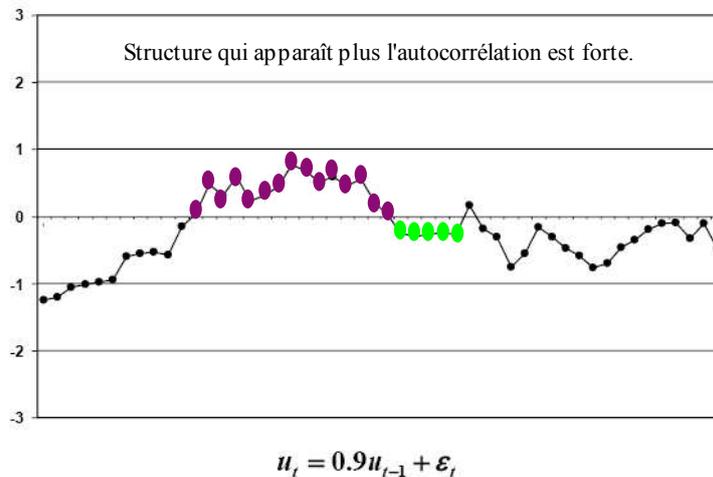


Erreur est liée à un coefficient de 0. Ici, le rho est positif.

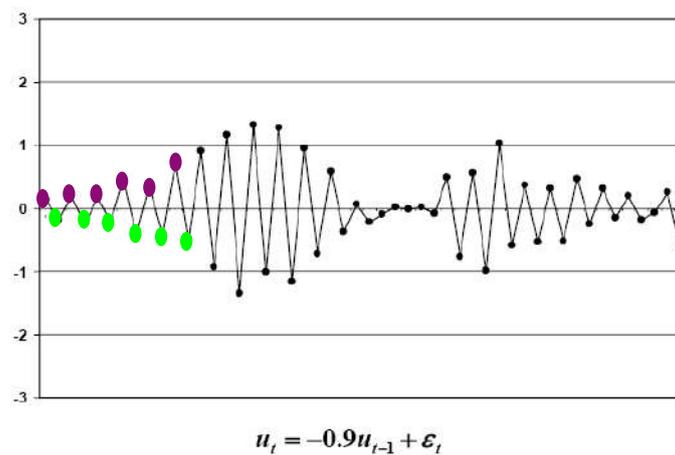
$$u_t = 0.0u_{t-1} + \varepsilon_t$$

Partie 1 : économétrie

En augmentant le ρ nous observons qu'une tendance d'autocorrélation positive commence à être apparente. Avec $\rho = 0.9$, des longues séquences de valeurs ayant le même signe sont apparentes et la tendance à retourner à 0 est faible



Même constat peut être fait dans l'autre sens pour de l'autocorrélation négative.



Note:

- Quand $\rho = 0$ il n'y a pas d'autocorrélation
- Quand $\rho \neq 0$ il y a autocorrélation

! L'autocorrélation ne biaise pas les coefficients c'est l'inférence qu'on ne sait pas lire. !

Partie 1 : économétrie

B. DÉTECTION DE L' AUTOCORRÉLATION

1. Test d'autocorrélation AR(1) de Durbin-Watson

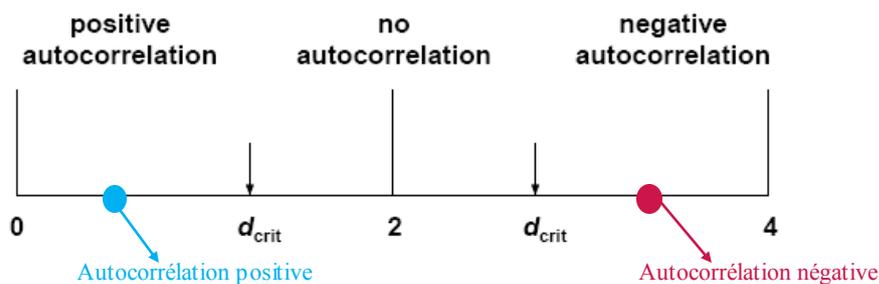
La statistique d de Durbin Watson est calculée pour les résidus.

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

On utilise les résidus car l'erreur on ne l'a pas étant donné que c'est la distance entre le point et la vraie droite (qu'on n'a pas).

Il peut être montré qu'en grand échantillon, d tend vers $2-2\rho$, où ρ est le paramètre dans la relation AR(1) : $u_t = \rho u_{t-1} + \varepsilon_t$

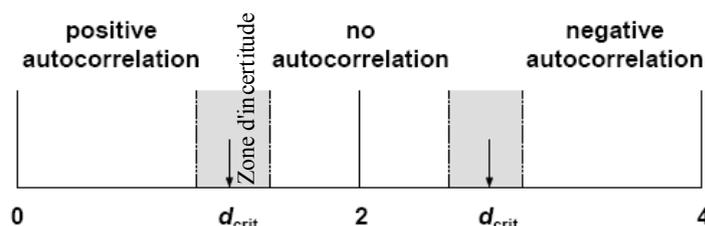
- Si il n'y a pas d'autocorrélation, $\rho = 0$ et d doit être distribué de manière aléatoire autour de 2.
- Si autocorrélation positive sévère, ρ sera proche de 1 et d proche de 0.
- S'il y a de l'autocorrélation négative sévère, ρ sera proche de -1 et d sera proche de 4.



No autocorrelation $d \rightarrow 2$
Severe positive autocorrelation $d \rightarrow 0$
Severe negative autocorrelation $d \rightarrow 4$

Afin d'effectuer le test de Durbin Watson il convient de définir des valeurs critiques pour d avec $H_0 : \rho = 0$ (pas de autocorrélation). Les valeurs critiques (d_{crit}) **dépendent du nombre d'observations** dans l'échantillon et du **nombre de variables explicatives**.

Une complication vient du fait que la valeur critique dépend aussi des valeurs particulières prises par les variables explicatives et donc, varie d'un échantillon à échantillon. Il est donc impossible d'établir une table donnant les valeurs critiques exactes pour tous les échantillons possibles.



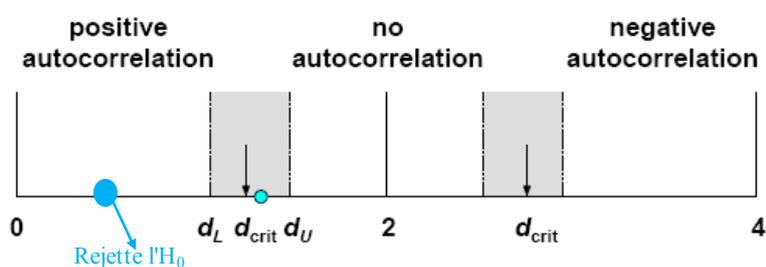
La valeur critique dépend de l'échantillon, donc on détermine une zone (borne supérieure et inférieure). Le problème est que si on se trouve dans la zone grise, on ne sait pas où se trouve la valeur critique.

Partie 1 : économétrie

Cependant D-W ont déterminé des limites inférieure et supérieures, d_L et d_U , pour les valeurs critiques et celles-ci sont présentées dans les tables standards.

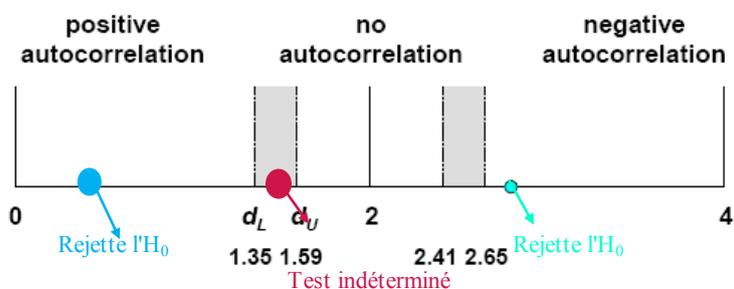
Ainsi:

- Si $d < d_L$, on **rejetera** H_0 . Il y a donc autocorrélation positive.
- Si $d > d_U$, nous ne **rejeterons pas** H_0 . Il n'y a pas d'autocorrélation. Bien entendu, si d était > 2 , nous devrions tester plutôt pour l'autocorrélation négative.
- Si d est entre d_L et d_U , nous ne **pouvons pas dire** si d est supérieur ou inférieur à la valeur critique donc le test est indéterminé (mais dans l'incertitude mieux vaut rejeter H_0).



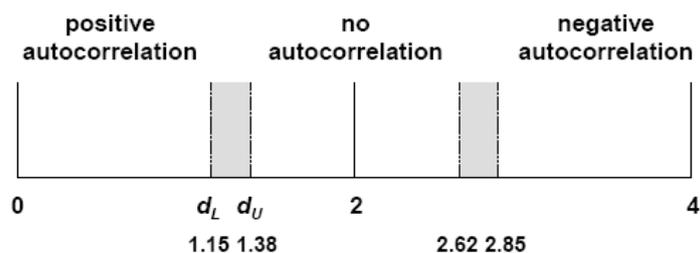
Regardons les bornes supérieures et inférieures pour un test à 5 % et un test à 1 % pour par exemple 36 observations et 2 variables explicatives.

Bornes pour un test de 5 %



Les limites pour les valeurs critiques en cas d'autocorrélation négative ne sont pas données dans les tables standards mais il est facile de les calculer car elles sont localisées symétriquement.

Bornes pour un test de 1 %



Partie 1 : économétrie

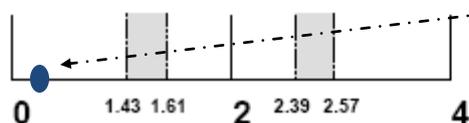
```
. reg LGFOOD LGDPI LGPRFOOD
```

Source	SS	df	MS	Number of obs = 45		
Model	2.17339266	2	1.08669633	F(2, 42) = 4347.62	Prob > F = 0.0000	
Residual	.010497979	42	.000249952	R-squared = 0.9952	Adj R-squared = 0.9950	
Total	2.18389064	44	.049633878	Root MSE = .01581		

LGFOOD	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
LGDPI	.6437701	.0259776	24.78	0.000	.5913451	.696195
LGPRFOOD	-.1107087	.0203594	-5.44	0.000	-.1517956	-.0696217
_cons	1.137597	.1372463	8.29	0.000	.8606231	1.414571

```
. dwstat
```

Durbin-watson d-statistic(3, 45) = .8294176



À un test de 5 %, nous rejetons l' H_0 car $0.82 < 1.43$ (l'inférence est fautive). Je ne sais pas lire la t-stat.

2. Le test de Breush-Godfrey

Le test de D-W ne permet que d'identifier l'éventuelle présence d'une autocorrélation de type AR(1). Le test de B-G est plus général. Si la valeur calculée est supérieure à la valeur critique, on rejette l'hypothèse d'absence d'autocorrélation (d'ordre p).

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$\hat{u}_t = \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \dots + \rho_p \hat{u}_{t-p} + \varepsilon_t$$

Prend les résidus et on les régresse en t-1.

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$$

H_0 : pas d'autocorrélation

$$(n-p)R^2 \sim \chi_p^2$$

Degré de liberté

```
. estat bgodfrey, lags(1/2)
```

Retards

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
t-1 1	14.264	1	0.0002
t-2 2	16.157	2	0.0003

Rejette l' H_0 car < 0.05

H_0 : no serial correlation

Compare avec la valeur critique d'une χ^2 .

Partie 1 : économétrie

C. ÉLIMINER L'AUTOCORRÉLATION AR(1)

Nous analyserons le cas d'une régression simple. Si le modèle est valide au temps t , il est aussi valide au temps $t-1$. L'équation a été multipliée par ρ .

$$\begin{aligned}
 & Y_t = \beta_1 + \beta_2 X_t + u_t & u_t = \rho u_{t-1} + \varepsilon_t \\
 & \rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{t-1} + \rho u_{t-1} \\
 & Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + u_t - \rho u_{t-1} \\
 & Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + \varepsilon_t
 \end{aligned}$$

Si l'erreur vaut aujourd'hui, elle vaut pour hier donc je peux exprimer par $t-1$.

$u_t = \rho u_{t-1} + \varepsilon_t$
 $\varepsilon_t = u_t - \rho u_{t-1}$

ε_t est l'innovation qui apparaît au temps t et il ne dépend pas de l'erreur passé, il est indépendamment distribué. **Nous avons donc un modèle sans autocorrélation.**

Cependant cette méthode a un problème mineur: il y a une non linéarité dans les paramètres. Le coefficient de X_{t-1} est égal au produit des coefficients de X_t et d' Y_{t-1} .

Cela signifie que si OLS était utilisé pour ajuster ce modèle, il ne prendrait pas en compte la restriction et nous aurions un conflit entre l'estimation des paramètres. Nous pourrions obtenir des estimations de 0.5 pour ρ et 0.8 β_2 pour. Mais ces nombres sont incompatibles avec l'estimation de 0.6 pour $\beta_2 \rho$.

$$\hat{Y}_t = 100 + 0.5Y_{t-1} + 0.8X_t - 0.6X_{t-1}$$

Nous avons donc besoin d'une technique d'estimation non linéaire (Moindre carrée non linéaire). Avant cela nous étendons le modèle à une régression multiple avec deux variables explicatives.

$$\begin{aligned}
 & Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t & u_t = \rho u_{t-1} + \varepsilon_t \\
 & \rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{2t-1} + \beta_3 \rho X_{3t-1} + \rho u_{t-1} \\
 & Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3t-1} + u_t - \rho u_{t-1} \\
 & Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3t-1} + \varepsilon_t
 \end{aligned}$$

$u_t = \rho u_{t-1} + \varepsilon_t$
 $\varepsilon_t = u_t - \rho u_{t-1}$

Nous avons deux restrictions:

- Les coefficients Y_{t-1} , X_{2t} et X_{2t-1}
- Les coefficients Y_{t-1} , X_{3t} et X_{3t-1}

Partie 1 : économétrie

$$Y_t = \beta_1(1 - \rho) + \rho Y_{t-1} + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3t-1} + \varepsilon_t$$

Source	SS	df	MS			
Model	7.64990646	3	2.54996882	Number of obs =	44	
Residual	.002663912	40	.000066598	R-squared =	0.9997	
Total	7.65257037	43	.177966753	Adj R-squared =	0.9996	
				Root MSE =	.0081607	
				Res. dev. =	-302.468	

LGHOUS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/b1	6.724528	1.147409	5.86	0.000	4.405528	9.043528
/rho	.9777322	.0045654	214.16	0.000	.9685051	.9869592
/b2	.1568757	.0827032	1.90	0.065	-.0102736	.3240251
/b3	-.0706955	.0664954	-1.06	0.294	-.2050878	.0636967

nl) (LGHOUS = {b1}*(1-{rho}) + {rho}*L.LGHOUS + {b2}*LGDPI - {b2}*{rho}*L.LGDPI + {b3}*LGPRHOUS - {b3}*{rho}*L.LGPRHOUS) in 2/l

nl: non linéaire

l. un retard

Paramètres estimés dans le modèle de départ.

Quand la variable b_2 augmente d'une unité, LGHOUS augmente de 0.15 %.

2/l car t-1 n'existe pas (ex: si j'ai 5 retards, je pars de la 6^e. L'ordre à de l'importance, il n'est pas aléatoire.)

Les bêtas et rho sont estimés à partir du modèle de base, on peut donc lire l'inférence car ε est une innovation indépendamment distribuée (pas d'autocorrélation sérielle).

D. AUTOCORRÉLATION DANS UN MODÈLE AVEC UNE VARIABLE DÉPENDANTE LAGUÉE

Lorsqu'une **variable dépendante laguée** (ex: le PIB d'aujourd'hui est lié à celui d'hier) est utilisée comme **variable explicative**, les estimations OLS seront sujettes à certain **biais en petits échantillons**. Même si tu satisfais les conditions de Gauss-Markov. En grand échantillon ce biais disparaît.

Cependant, si le **terme d'erreur** est sujet à de **l'autocorrélation**, OLS mènera à des estimations non **consistantes**.

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + u_{t-1}$$

Si le modèle est valide au temps t , il doit aussi être valide au temps $t-1$.

Donc, Y_{t-1} contient une composante aléatoire u_{t-1} , et donc u_t aussi. Y_{t-1} dépend de u_{t-1} mais u_t dépend de u_t . Par conséquent, Y_t dépend de u_t . Il y a donc violation de la 4^e hypothèse de Gauss-Markov (exogénéité). Il y a donc un biais d'endogénéité. Cela signifie que OLS mène à des estimations non consistantes et que les tests t et F sont non valides.

Partie 1 : économétrie

Lorsqu'une variable dépendante laguée est utilisée comme variable explicative, il faut utiliser la statistique h de Durbin pour l'autocorrélation car la statistique d est biaisée vers 2.

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + u_{t-1}$$

Statistique alternative de Durbin Watson.

$$h = \hat{\rho} \sqrt{\frac{n}{1 - ns^2_{b_{Y(-1)}}}}$$

Distribué normalement donc les valeurs critiques sont celles de la normale. Ce qui signifie que pour un niveau de 5 %, la valeur critique est de 1.96.

$$d \approx 2 - 2\rho$$

$$\hat{\rho} = 1 - 0.5d$$

Trois éléments sont requis pour ce test:

- Estimation de ρ , le paramètre dans le processus AR(1)
- n le nombre de régression
- Estimation de la variance du coefficient de la variable indépendante laguée.

```
. reg LGHOUS LGDPI LGPRHOUS 1.LGHOUS if DATE>=1964&DATE<=1994
```

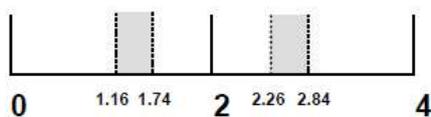
Source	SS	df	MS	Number of obs = 31		
Model	2.89802489	3	.966008296	F(3, 27)	=	18859.89
Residual	.001382947	27	.00005122	Prob > F	=	0.0000
				R-squared	=	0.9995
				Adj R-squared	=	0.9995
Total	2.89940784	30	.096646928	Root MSE	=	.00716

LGHOUS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
LGDP	.2625971	.0573576	4.58	0.000	.144909	.3802853
LGPR	-.0391084	.0136851	-2.86	0.008	-.0671879	-.0110289
LGHOUS						
L1.	.7875359	.0462763	17.02	0.000	.6925848	.882487
_cons	-.642853	.1954479	-3.29	0.003	-1.043879	-.241827

```
. dwstat
```

```
Durbin-watson d-statistic( 4, 31) = 1.915793
```

Est-ce qu'il y a autocorrélation? Je ne peux pas regarder la d (1.91) car Y dépend de Y_{t-1} (LGHOUS dépend de LGHOUS).



```
. Durbin h=0.2426
```

On doit calculer la h :

$$1 - 0.5 \times 1.91 \sqrt{\frac{31}{1 - 31 \times (0.46)^2}} = 0.2426 < 2$$

Valeur critique d'une

normale à un niveau de 5 % => on rejette pas l' H_0 . (il n'y a pas d'autocorrélation).

Partie 1 : économétrie

E. TEST DE FACTEUR COMMUN ET SPÉCIFICATION DE MODÈLE DYNAMIQUE

Dans les séquences précédentes, il a été montré que, si on a un modèle de régression simple avec Y dépendant de X et un terme d'erreur u sujet à un processus AR(1) ; le modèle peut être ré-écrit avec Y_t dépendant de X_t , Y_{t-1} , X_{t-1} et un terme d'erreur ε_t (non sujet à l'autocorrélation).

Quand on a autocorrélation dans le terme d'erreur :

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

- Problème de corrélation sérielle (pas trop grave).
- Modèle dynamique (donne les mêmes symptômes que la corrélation sérielle).

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + \varepsilon_t$$

Dans ce modèle les paramètres sont non linéaires étant donné que le coefficient de X_{t-1} est égal au produit des coefficients de X_t et Y_{t-1} . Ce modèle est une version restreinte d'un modèle plus général. Dans ce cas, la version restreinte incorpore deux restrictions. **Le nombre de restrictions dans le modèle (AR1) étant normalement égal au nombre de variables explicatives.**

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Modèle Restreint

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + \varepsilon_t$$

Modèle Non Restreint

$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_t + \lambda_3 X_{t-1} + \varepsilon_t$$

Restriction incorporée dans le processus AR(1)

$$\lambda_3 = -\lambda_1 \lambda_2$$

Considérons le cas où la spécification originale est un modèle de régression multiple avec deux variables explicatives et un terme d'erreur sujet à un processus AR(1). Le cas spécial AR(1) est une fois de plus une version restreinte d'un modèle plus général. Dans ce cas, la version incorpore deux restrictions.

Partie 1 : économétrie

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Modèle Restreint

Modèle statique $Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_{2t} - \beta_2 \rho X_{2,t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3,t-1} + \varepsilon_t$

Modèle Non Restreint

Modèle dynamique $Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} + \lambda_3 X_{2,t-1} + \lambda_4 X_{3t} + \lambda_5 X_{3,t-1} + \varepsilon_t$

Restrictions incorporées dans le processus AR(1)

$$\lambda_3 = -\lambda_1 \lambda_2 \quad \lambda_5 = -\lambda_1 \lambda_4$$

Normalement le nombre de restrictions dans le modèle (AR1) est égal au nombre de variables explicatives.

Nous devons tester la validité de ces restrictions: test du facteur commun. L' H_0 est la suivante: **les restrictions sont valides**. Le test usuel **F n'est pas approprié** car il s'agit d'un **modèle non linéaire**. Nous utiliserons une statistique alternative impliquant une comparaison de **RSS_R** et **RSS_U**. Pour ce faire, nous utilisons le test statistique ci-dessous (où n est le nombre d'observation dans la régression):

$$\text{Test statistic: } n \log \frac{RSS_R}{RSS_U}$$

Distribué comme une χ^2 avec come degré de liberté le nombre de contrainte.

Notons que RSS_R ne pourra jamais être plus petit que RSS_U et il sera en pratique plus grand, car imposer une restriction en général mène à des pertes de validité d'approximation. La question qui se pose est sur la significativité de la perte de validité d'approximation. Si c'est le cas, c'est une indication qu'imposer une restriction a causé une distorsion, et donc nous devrions conclure que les restrictions sont non valides.

Exemple:

```

=====
LS // Dependent Variable is LGHOUS
Echantillon(adjusted): 1960 1994
Included observations: 35 after adjusting endpoints
Convergence achieved after 24 iterations
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.131573	0.727241	8.431276	0.0000
LGDP1	0.275879	0.078318	3.522534	0.0013
LGPRHOUS	-0.303387	0.085802	-3.535896	0.0013
AR(1)	0.972488	0.004167	233.3540	0.0000

```

=====
R-squared          0.999695    Mean dependent var 6.017555
Adjusted R-squared 0.999665    S.D. dependent var 0.362063
S.E. of regression 0.006622    Akaike info criter -9.927360
Sum squared resid  0.001360    Schwarz criterion  -9.749606
Log likelihood     128.0660    F-statistic         33865.14
Durbin-Watson stat 1.423031    Prob(F-statistic)  0.000000
=====

```

Méthode d'estimation par AR(1).

Partie 1 : économétrie

```

=====
LS // Dependent Variable is LGHOUS
Echantillon(adjusted): 1960 1994
Included observations: 35 after adjusting endpoints
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.386286	0.177312	-2.178563	0.0376
LGDP	0.301400	0.066582	4.526717	0.0001
LGPRHOUS	-0.192404	0.078085	-2.464038	0.0199
LGHOUS(-1)	0.726714	0.064719	11.22884	0.0000
LGDP(-1)	-0.014868	0.092493	-0.160748	0.8734
LGPRHOUS(-1)	0.138894	0.084324	1.647143	0.1103

```

=====
R-squared          0.999797    Mean dependent var  6.017555
Adjusted R-squared 0.999762    S.D. dependent var  0.362063
S.E. of regression 0.005589    Akaike info criter -10.21915
Sum squared resid  0.000906    Schwarz criterion   -9.952519
Log likelihood     135.1723    F-statistic         28532.58
Durbin-Watson stat 1.517562    Prob(F-statistic)  0.000000
=====

```

Spécification non restreinte

$$n \log \left(\frac{RSS_R}{RSS_U} \right) = 35 \log \left(\frac{0.001360}{0.000906} \right) = 14.2$$

$$\chi^2_{crit} = 13.8 \quad (2, 0.1\%)$$

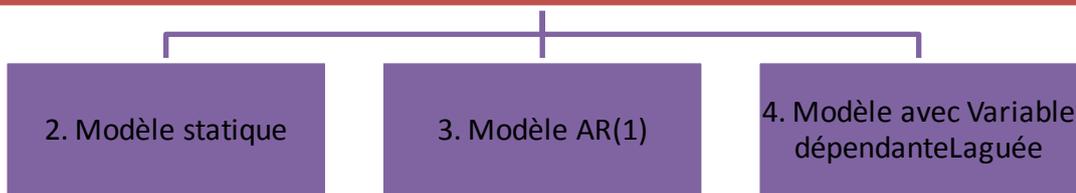
Comme $14.2 > 13.8$ (la valeur d'une χ^2 à un niveau de 0.1 % avec deux degré de liberté), nous rejetons H_0 : les restrictions sont valides.

Nous rejetons donc les restrictions. Nous devrions choisir le modèle dynamique (modèle plus général) plutôt que de supposer que le terme d'erreur est sujet à un processus AR(1).

Méthodologiquement, nous avons choisis une procédure allant d'un modèle spécifique à un modèle plus général afin de choisir le modèle adéquat. En principe, il est préférable de commencer avec un modèle qui soit suffisamment général pour éviter des problèmes potentiels de "sous spécification" et dès lors voir s'il est possible de le simplifier.

Partir d'un modèle dynamique complet pour le simplifier

1. Modèle général avec des variables Laguées



1. Dans ce cas, le point de départ est le modèle avec toutes les variables laquées:

$$Y_t \text{ dépend de } Y_{t-1} X_t X_{t-1} \quad Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} + \lambda_3 X_{2t-1} + \lambda_4 X_{3t} + \lambda_5 X_{3t-1} + \varepsilon_t$$

2. Après l'avoir approximé, nous devrions être capables de le simplifier en modèle statique.

- Si on ne rejette pas l' H_0 : modèle est statique et j'utilise OLS
- Si on rejette l' H_0 : modèle AR(1) avec autocorrélation dans le terme d'erreur.

$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} + \lambda_3 X_{2t-1} + \lambda_4 X_{3t} + \lambda_5 X_{3t-1} + \varepsilon_t$$

$$\lambda_1 = \lambda_3 = \lambda_5 = 0 \quad \text{Pour que le modèle soit purement statique.}$$

$$Y_t = \lambda_0 + \lambda_2 X_{2t} + \lambda_4 X_{3t} + \varepsilon_t$$

3. Peut faire un test de facteur commun et voir si nous pouvons simplifier le modèle en une spécification AR(1)

- Si on ne rejette pas l' H_0 : garde le modèle AR(1) et on corrige le terme d'erreur.
- Si on rejette l' H_0 : prend le modèle dynamique où Y_t dépend de Y_{t-1} mais pas de X_{t-1}

$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} + \lambda_3 X_{2t-1} + \lambda_4 X_{3t} + \lambda_5 X_{3t-1} + \varepsilon_t$$

$$\lambda_3 = -\lambda_1 \lambda_2 \quad \lambda_5 = -\lambda_1 \lambda_4 \quad \text{Contrainte non linéaire}$$

$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} - \lambda_1 \lambda_2 X_{2t-1} + \lambda_4 X_{3t} - \lambda_1 \lambda_4 X_{3t-1} + \varepsilon_t$$

4. Le modèle avec une variable laquée est une spécification dynamique adéquate, si les autres variables laquées manquent de pouvoir explicatif significatif.

- Si on ne rejette pas l' H_0 : garde le modèle.
- Si on rejette l' H_0 : on garde le modèle général.

$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} + \lambda_3 X_{2t-1} + \lambda_4 X_{3t} + \lambda_5 X_{3t-1} + \varepsilon_t$$

$$\lambda_3 = \lambda_5 = 0$$

Retard dépend de Y_{t-1} mais pas de retard dans les X.

$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} + \lambda_4 X_{3t} + \varepsilon_t$$