Ecole Doctorale Thématique
en Sciences de Gestion

Méthodologie quantitative

Prof. François Rycx (frycx@ulb.ac.be)

Homepage : http://homepages.ulb.ac.be/~frycx/

Examen à cahier fermé (tte la matière vue au cours + sections 1.10 & 1.11, càd formes fct. & dummies) + articles à lire (cf homepage)

Ouvrages de référence :

a) Baltagi, B.H. (2002), Econometrics, 3rd. ed., Springer, NY.

b) Greene, W.H (2005), Econométrie, 5e éd., Pearson Education, Paris.

c) Gujarati, D.N. (2003), Basic Econometrics, 4th. ed., McGraw-Hill Higher Education, NY.

d) Hamilton, J.D. (1994), Time Series Analysis, Princeton University Press, Princeton.

e) Kennedy, P. (1998), A Guide to Econometrics, 4th ed., MIT Press, Cambridge (Mass.).

f) Sevestre, P. (2002), Econométrie des données de panel, Dunod, Paris.

g) Wooldridge, J.M. (2002), Econometric Analysis of Cross-section and Panel Data, MIT Press, Cambridge (Mass.).

Table des matières

1. Introduction

⚠

Aussi à étudier pour l'examen :

a) Les formes fonctionnelles des modèles de régression (section 1.10. dans slides : 1.28 – 1.40)

b) Les modèles de régression à variables qualitatives indépendantes (section 1.11 dans slides : 1.41 – 1.57)

1. Relâchement des hyp. du modèle classique

2.1. La multicolinéarité

2.2. L'hétéroscédasticité

2.3. L'autocorrélation

2.4. Modélisat° économétrique : spécification du modèle et tests de diagnostique (pas à étudier pour l'examen)

Pour 2.1, 2.2, 2.3, questions :

a) Nature du pblme

b) Estimation en présence du pblme

c) Conséquences, méthodes de détection, remèdes

3. Thèmes d'économétrie

3.1. Les modèles de régression avec données de panel

↳ effets de fixité vs. effets aléatoires

3.2. L'économétrie des séries chronologiques

↳ processus stochastiques stationnaires, non stationnaires, à racine unitaire, intégrés

↳ la fausse régression, tests de stationnarité, de racine unitaire, la cointégration

3.3. Les modèles de régression à variable dpdte qualitative

↳ le modèle de probabilité linéaire (MPL)

↳ les modèles logit, probit, Tobit

1. Introduction

1.1. Le modèle de régression linéaire classique

a) Pourquoi ce modèle est-il intéressant ?

* Modèle de base en économétrie
  (cf. modèle en conc. pfte en économie)

* Lorsque les hypothèses (sous-jacentes) de ce modèle sont satisfaites, les estimateurs des MCO possèdent des propriétés statistiques intéressantes.

b) Les 10 hyp. du MRLC

Hypothèse 1 : Le modèle de régression est linéaire dans les paramètres :

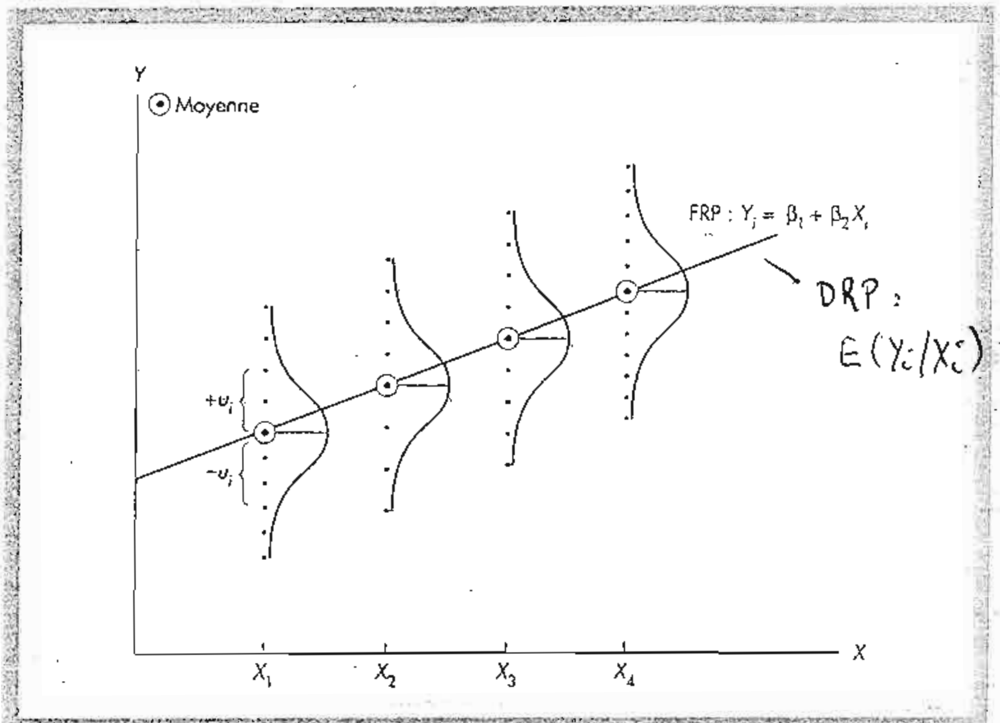
$Y_i = \beta_1 + \beta_2 X_i + u_i$ (p. ex.)

! Modèle peut ne pas être linéaire dans les variables.

Rem : Une fction est linéaire dans un paramètre β ssi ce paramètre apparaît uniqmt à la puissance 1 et β n'est ni divisé, ni multiplié par un autre paramètre.

Hypothèse 2 : Les valeurs de X sont fixes lorsqu'on procède à des tirages répétés d'échantillons. X est une variable non stochastique.

Hypothèse 3 : La moyenne des erreurs / des perturbations u_i est nulle. Autrement dit, la moyenne conditionnelle de l'erreur aléatoire u_i vaut zéro : $E(u_i / X_i) = 0$
⇒ Les variables qui ne sont pas prises en compte dans le modèle de régression (et qui sont en qque sorte comprises ds le terme d'erreur) n'ont pas d'impact syst. sur Y.



DRP : $E(Y_i/X_i) = \beta_1 + \beta_2 X_i$

Figure 3.3
Distribution conditionnelle des termes d'erreur u_i

Remarque : L'hyp. selon laquelle $E(u_i / X_i) = 0$ implique que la fction de régression pop. s'écrit comme suit : $E(Y_i / X_i) = \beta_1 + \beta_2 X_i$ → les 2 hyp. sont équivalentes. (Hyp. que droite de rég. passe par les esp. cond. de Y implique que les espérances cond. de u_i sont égales à zéro).

Hypothèse 4 : Homoscédasticité du terme d'erreur.

La variance conditionnelle des u_i est constante.

$$Var(u_i / X_i) = E(u_i / X_i - E(u_i / X_i))^2$$
$$= E(u_i^2 / X_i) \quad \text{car } E(u_i / X_i) = 0 \quad (\text{Hyp } 3).$$
$$= \sigma^2$$

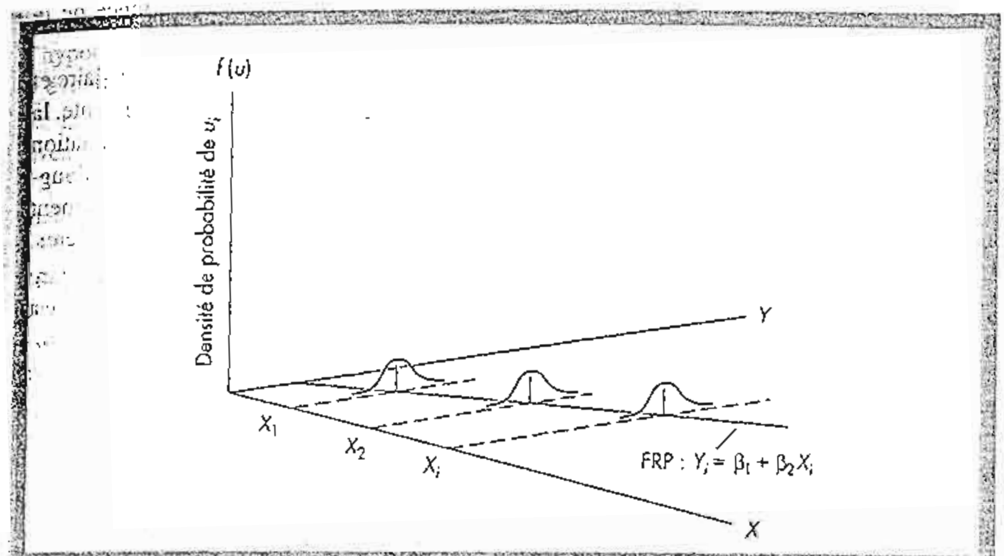


Figure 3.4
L'homoscédasticité

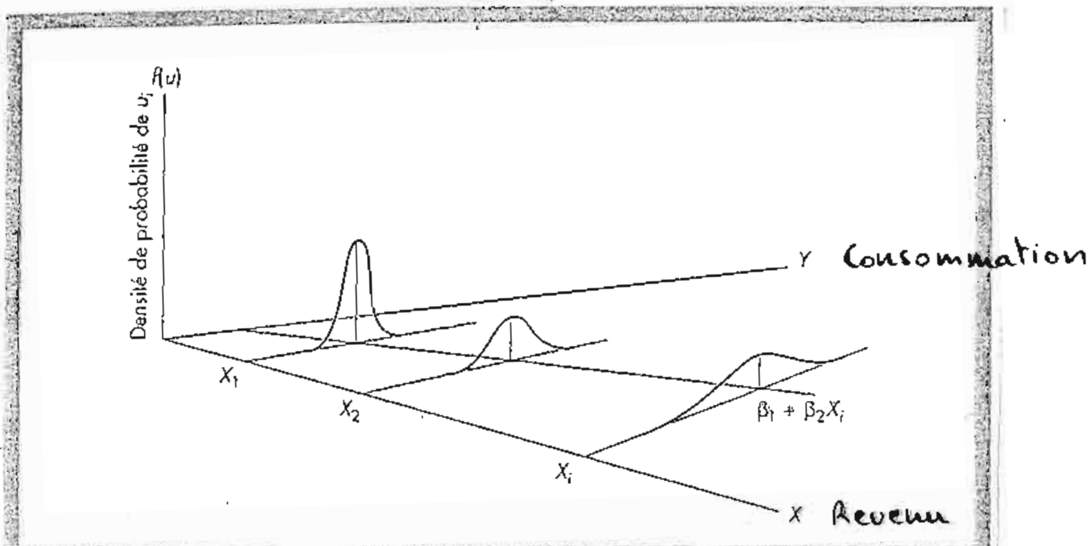


Figure 3.5
L'hétéroscédasticité

Lorsque la variance conditionnelle des u_i (ou des Y_i) varie en fct° de la valeur prise par X, on parle d'hétéroscédasticité : $var(u_i / X_i) = \sigma_i^2$

<u>Hypothèse 4</u> : variance conditionnelle de Y étant donné X est constante $\longrightarrow$ $var(Y/X) = \sigma^2$

<u>Intuition</u> :

En présence d'hétéroscéd., $var(u_i / X_i) = \sigma_i^2$

Figure 3.5 :

$var(u/X_1) < var(u/X_2) < \ldots < var(u/X_i)$

$\longrightarrow$ il est probable que les obs. relatives à la variable Y pour $X = X_1$ soient + proches de la régr. pop. que les obs. qui correspondent à d'autres valeurs de X (p.ex $X_2, X_3, \ldots$)

$\longrightarrow$ ttes les obs. ne bénéficient pas du m̂ degré de „fiabilité" („fiabilité" = distance |e| l'obs. Y_i et son espérance cond.).

En présence d'hétérosc., on pourrait vouloir restreindre l'échantillon aux valeurs de Y qui sont proches de leur moyenne conditionnelle ... mais ↓ de la variance, biais de sélect°, etc.

En imposant hyp. 4, on garantit que les Y correspondant aux ≠ valeurs de X ont la même importance.

Hypothèse 5 :

Pas d'autocorrélation du terme d'erreur.

Pour des valeurs de X, X_i et X_j $(i \neq j)$, la corrélation [el] u_i et u_j $(i \neq j)$ vaut zéro :

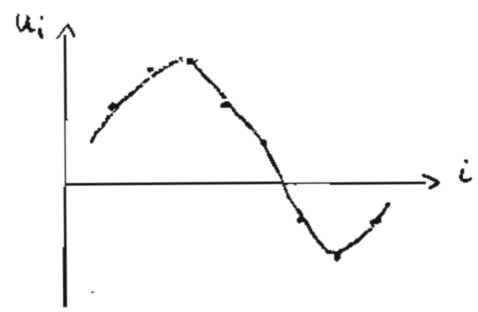
$$\mathrm{cov}(u_i, u_j / X_i, X_j) = E\{[u_i - E(u_i)]/X_i\}\{[u_j - E(u_j)]/X_j\}$$

$$= E\{(u_i/X_i)(u_j/X_j)\}$$

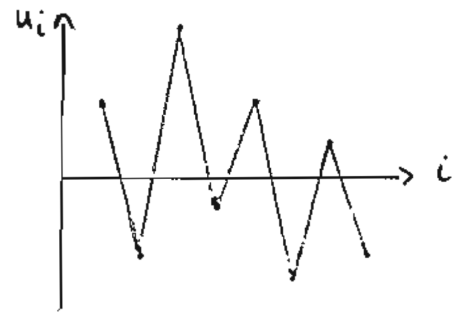
$$(\text{car } E(u_i/X_i) = 0\text{, hyp. 3})$$

$$= 0$$

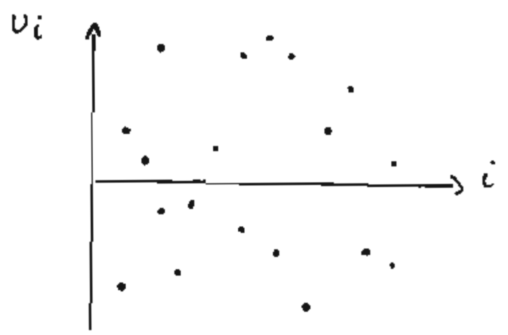
Exemples de corrélation des erreurs :



(a) corrélat° positive



(b) corrélat° négative



(c) aléatoire

Intuition :

Soit une fct° de rég. pop. :

$Y_t = \beta_1 + \beta_2 X_t + u_t$

où termes d'erreur u_t et u_{t-1} sont positivmt corrélés.

Dans ce cas, Y_t dpd non seulement de X_t mais aussi de u_{t-1} (car par hyp. u_{t-1} exerce influence positive sur u_t).

Sous hyp. 5, on peut étudier impact systématique de X_t sur Y_t sans se soucier de l'influence de terme d'erreur sur Y_t.

Hypothèse 6 :

La covariance lel u_i et X_i vaut zéro : $E(u_i X_i) = 0$

En effet :
$$\begin{aligned} cov(u_i X_i) &= E\{(u_i - E(u_i))(X_i - E(X_i))\} \\ &= E\{u_i (X_i - E(X_i)\} \quad \text{car } E(u_i) = 0 \\ &= E(u_i X_i) - E(X_i) E(u_i) \\ &= E(u_i X_i) \quad \text{car } E(u_i) = 0 \\ &= 0 \quad \text{(par hyp.)} \end{aligned}$$

$\Rightarrow$ Le terme d'erreur et la var. expl. X sont non corrélés.

Intuition :

En spécifiant forme de la fct° de rég. pop ($Y_i = \beta_1 + \beta_2 X_i + u_i$), nous avons supposé que X et u avaient une influence distincte et additionnelle sur Y.

Or, si X et u sont corrélés, il devient difficile d'isoler l'influence de X sur Y de celle de u sur Y.

Hypothèse 7 :

Le nbre d'obs. n doit être supérieur au nbre de paramètres à estimer → nbre d'obs. n doit être supérieur au nbre de var. expl.
(cf. pblme de micronumérosité)

Hypothèse 8 :

La variance de X doit être un nbre positif fini → valeurs prises par la variable X ne peuvent pas ttes être identiques.
(cf. il est difficile d'expliquer les variations ds la variable Y par la variable X si X varie peu).

Hypothèse 9 :

Le modèle de régression doit être correctmt spécifié.

→ Quelles variables faut-il inclure ds le modèle ?
Quelle est la forme fctionnelle du modèle (quadratique, exponentielle, etc.) ?
Le modèle est-il linéaire ds les paramètres ?
Quelle est la fct° de distribut° du terme d'erreur ?
...

Hypothèse 10 : Il n'y a pas de multicolinéarité parfaite
→ ∄ de relation linéaire exacte (e) les variables explicatives → une variable explicative ne peut donc pas s'écrire comme une combinaison linéaire exacte des autres variables explicatives.

(Si deux var. explic. sont fortmt corrélées, il est difficile d'isoler leurs impacts respectifs sur la var. dpdte Y)

c) La précision ou l'erreur standard des estimateurs

Soit $Y_i = \beta_1 + \beta_2 X_i + u_i$ (FRP)

Dans ce cas, nous savons que : $\hat{\beta}_2 = \dfrac{\Sigma x_i y_i}{\Sigma x_i^2}$, $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

Les estimateurs des MCO dépendent des données de l'échantillon. Comme les données varient d'un échant. à l'autre, les estimateurs varient d'un échant. à l'autre.

Il nous faut un indicateur mesurant la précision ou la fiabilité des estimateurs $\hat{\beta}_1$ et $\hat{\beta}_2$.

La précision d'un estimateur se mesure par son erreur standard ("standard error").

Erreur standard d'un estimateur = écart-type de la distrib. d'échant. de l'estimateur.

Distribution d'échant. de l'estimateur = distribut° de l'ensemble des valeurs prises par l'estimateur lorsqu'on procède à des tirages répétés d'échant. (ensemble des échant. d'une m̂ taille et d'une m̂ population).

• Sous les hyp. du modèle de régression linéaire classique :

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \Rightarrow \boxed{\text{s.e.}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2 \Rightarrow \boxed{\text{s.e.}(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \cdot \sigma}$$

où
- var = variance , $x_i = (X_i - \overline{X})$, $y_i = (Y_i - \overline{Y})$
- s.e. = erreur standard
- σ^2 = variance homoscédastique du terme d'erreur u_i ($= \text{var}(u_i / X_i)$)

L'estimateur des MCO de σ^2 (paramètre pop. inconnu) est donné par :

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}}$$

où
- $(n-2)$ = nbre de degrés de liberté.
- $\sum \hat{u}_i^2$ = somme des carrés des résidus.

Degrés de liberté ? Nbre d'obs. moins nbre de param à estimer
Nbre de valeurs que l'on peut choisir arbitrairement.
Comme pour calculer $\sum \hat{u}_i^2$, il faut connaître (estimer) au préalable $\hat{\beta}_1$ et $\hat{\beta}_2$, on perd 2 degrés de liberté ... il en reste $(n-2)$.

$$\boxed{\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}}} \longrightarrow \begin{cases} \text{Erreur standard de la régression} \\ \text{Standard error of the estimate} \\ \text{Standard error of the regression} \end{cases}$$

$\hat{\sigma}$ est l'estimateur de l'écart type des erreurs, càd. de l'écart-type des observations Y autour de la droite de régression. Souvent utilisé pour juger de la qualité de l'ajustmt d'un modèle de régression.

Rappel : $\left.\begin{aligned} \sigma^2 &= var\ (u_i / X_i) \\ &= var\ (Y_i / X_i) \end{aligned}\right\} \longrightarrow$ on appelle $\hat{\sigma}$ aussi l'écart-type cond. des erreurs (u_i) et des obs. (Y_i)

- <u>Propriétés de la variance des estimateurs $\hat{\beta}_1$ et $\hat{\beta}_2$:</u>

$$\boxed{var\ (\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}} \quad , \quad \boxed{var\ (\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2}$$

a) $var\ (\hat{\beta}_2)$ est proportionnelle à la variance (homosc.) de l'erreur u_i (σ^2) <u>et</u> inversement proportionnelle à la variation des X $(\sum x_i^2)$ et à la taille de l'échantillon (n).

b) $var\ (\hat{\beta}_1)$ est proportionnelle à σ^2 et à $\sum X_i^2$ <u>et</u> inversement proportionnelle à $\sum x_i^2$ et à la taille de l'échantillon (n).

c) $\boxed{cov\ (\hat{\beta}_1, \hat{\beta}_2) =} - \bar{X} \cdot var\ (\hat{\beta}_2) \boxed{= - \bar{X} \left(\frac{\sigma^2}{\sum x_i^2}\right)}$

Comme par déf. $var\ (\hat{\beta}_2) \geq 0$, si $\bar{X} > 0$:
$cov\ (\hat{\beta}_1, \hat{\beta}_2) \leq 0$. Dans ce cas, si β_2 est sur-estimé, β_1 sera sous-estimé.

1.2. Les propriétés des estimateurs des MCO : le théorème de Gauss – Markov

- Théorème de Gauss – Markov :

  „ Sous les hyp. du modèle de rég. linéaire classique, les estimateurs des MCO sont BLUE, càd. que ds la classe des estimateurs linéaires non biaisés, ils ont la variance minimum."

  → Les estimateurs des MCO sont les meilleurs estimateurs linéaires non biaisés : BLUE (Best Linear Unbiased Estimators)

- Un estimateur (p. ex. $\hat{\beta}_2$) est BLUE si :

  a) $\hat{\beta}_2$ est un estimateur linéaire.

  b) $\hat{\beta}_2$ est un estimateur sans biais : $E(\hat{\beta}_2) = \beta_2$.

  c) $\hat{\beta}_2$ a la variance minimale ds l'ensemble de la classe des estimateurs linéaires sans biais.

- Remarque :

  Estimateur efficient = estimateur sans biais dont la variance est minimale.

• Graphiquement :

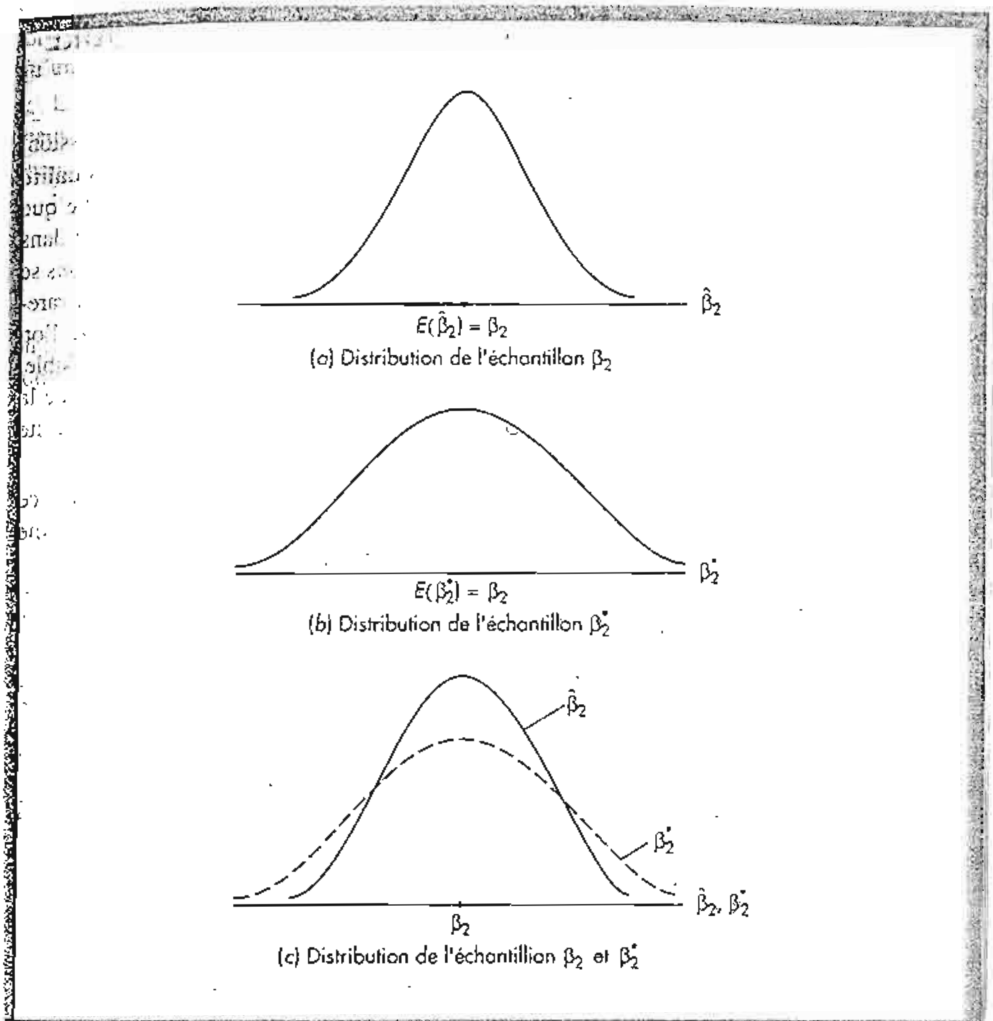


Figure 3.8
Distribution de l'estimateur MCO β_2 et de l'estimateur alternatif β_2^*

• Remarques :

a) Théorème de Gauss - Markov ne fait aucune hyp. par rapport à la distribution du terme d'erreur et à la taille de l'échantillon.

b) Propriétés asymptotiques = propriétés qui sont uniq. vérifiées lorsque la taille de l'échantillon est très grande.

1.3. Le modèle de régression linéaire classique normal

Pour faire de l'inférence statistique, il est nécessaire de spécifier la distribution de probabilité du terme d'erreur u_i.

Pq ?

Les estimateurs des MCO $\hat{\beta}_1$ et $\hat{\beta}_2$ sont des fonctions linéaires du terme d'erreur u_i

En effet, nous savons :

a) La variable Y est une fction linéaire du terme d'erreur u_i (qui est une variable aléatoire).

b) Les estimateurs des MCO $\hat{\beta}_1$ et $\hat{\beta}_2$ sont des fctions linéaires de la variable dpdte Y.

La distribution d'échantillonnage des estimateurs des MCO dépend des hyp. // à la distribution des erreurs u_i.

Comme distribut° d'échant. des estimateurs est nécess. pour faire de l'inférence stat. // paramètres pop., la nature de la distribut° de prob. du terme d'erreur u_i occupe une place centrale dans la théorie des tests d'hyp.

En ajoutant l'hyp. de normalité de u_i, on obtient le modèle de régr. linéaire classique normal.

- On a donc que les u_i suivent une loi normale et :

  a) La moyenne des u_i est égale à zéro : $E(u_i) = 0$

  b) Homoscédasticité : $E(u_i - E(u_i))^2 = E(u_i^2) = \sigma^2$

  c) Covariance ou autocorrélation des u_i est nulle :
  $E\{(u_i - E(u_i))(u_j - E(u_j))\} = E(u_i u_j) = 0 \quad , i \neq j.$

  $\Rightarrow \quad u_i \sim N(0, \sigma^2)$

- Comme 2 variables suivant une loi normale qui sont non corrélées sont indpdtes :

$$\boxed{u_i \sim \text{ i.i.d. } N(0, \sigma^2)}$$

- <u>Pourquoi suppose-t-on généralmt que le terme d'erreur u_i suit une loi normale</u> ?

  a) Thérorème limite central : « Lorsqu'on a un <u>grand</u> nbre de var. aléat. indpdtes et identiqmt distribuées, la distribution de leur somme tend (presque tjs) vers une distribution normale" $\longrightarrow$ asymptotiqmt !

  b) Toute fction linéaire de variables normales est elle-m̂ normale. Par conséquent, si les erreurs suivent une loi normale, il en sera de m̂ pour les estimateurs des MCO $\hat{\beta}_1$ et $\hat{\beta}_2$.

  c) Distribution simple entièrement caractérisée par sa moyenne et sa variance. Ses propriétés théoriques sont bien connues.

• <u>Propriétés des estimateurs des MCO sous l'hypothèse de normalité du terme d'erreur</u> :

a) Ils sont sans biais
b) Ils sont de variance minimale } => efficients

c) Ils sont consistants (convergents) :
Lorsque la taille de l'échant. n tend vers ∞, les estimateurs convergent vers leurs valeurs pop.

$$\boxed{\lim_{n \to \infty} P\{|\hat{\theta} - \theta| < \delta\} = 1}, \ \forall \delta > 0 \ (*)$$

$$\boxed{\underset{n \to \infty}{\text{plim}} \ \hat{\theta} = \theta}$$ (plim = probabilité limite)

<u>Attention</u> : Ne pas confondre un estimateur consistant avec un estimateur asymptotiqmt non biaisé. Asymptotiqmt non biaisé : $\lim_{n \to \infty} E(\hat{\theta}) = \theta$.

d) $\hat{\beta}_1$ suit une loi normale (car c'est une fct° linéaire de u_i) avec les moyenne et variance suivantes :
Moyenne : $E(\hat{\beta}_1) = \beta_1$
Variance : $\text{var}(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \dfrac{\Sigma X_i^2}{n \Sigma x_i^2} \cdot \sigma^2$

$$\Rightarrow \boxed{\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})}$$

Par définit°, variable $Z = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}}$ suit distribution normale centrée réduite :
$Z \sim N(0,1)$.

(*) La distance let $\hat{\theta}$ et θ peut être rendue aussi petite que l'on veut en prenant un échant. suff. gd.

e) $\hat{\beta}_2$ suit une loi normale (car c'est une fct° linéaire de u_i) avec les moyenne et variance suivantes :

Moyenne : $E(\hat{\beta}_2) = \beta_2$

Variance : $var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \dfrac{\sigma^2}{\sum x_i^2}$

$$\Rightarrow \boxed{\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})}$$

Par définition, variable $Z = \dfrac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}}$ suit une distrib. normale centrée réduite :

$Z \sim N(0, 1)$

<u>Géométriquement</u> :

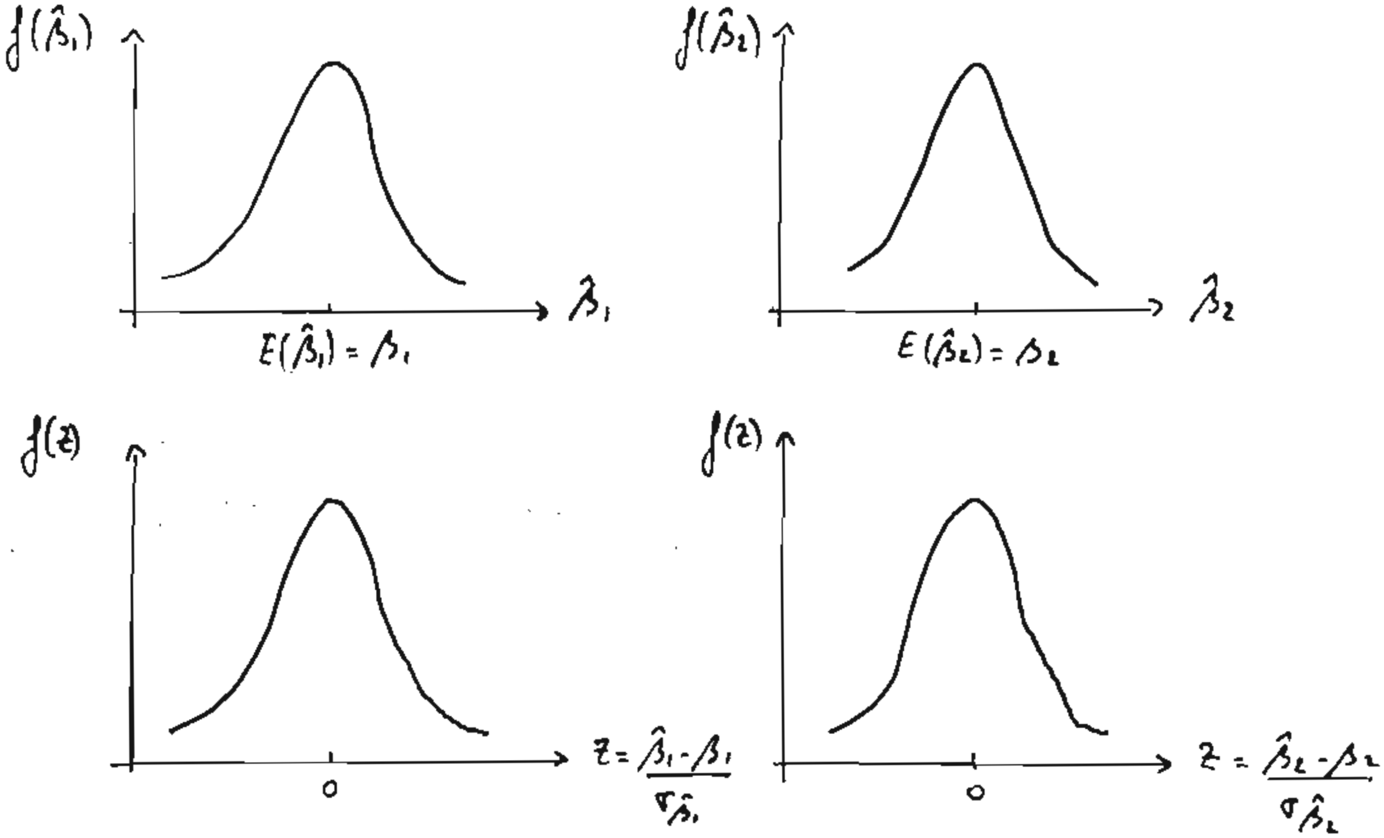


Fig. Distrib. de prob de $\hat{\beta}_1$ et $\hat{\beta}_2$.

f) $(n-2)\,(\hat{\sigma}^2/\sigma^2)$ suit une distrib. de χ^2 à $(n-2)$ degrés de liberté.

g) $\hat{\beta}_1$ et $\hat{\beta}_2$ sont distribués indépdt de $\hat{\sigma}^2$.

h) $\hat{\beta}_1$ et $\hat{\beta}_2$ ont la variance minimale ds l'ensemble de la classe des estimateurs non biaisés
$\Rightarrow$ ils sont BUE
(Best Unbiaised Estimators)

En résumé :

Sous l'hyp. de normalité de u_i :
a) $\hat{\beta}_1$ et $\hat{\beta}_2$ suivent une loi normale.
b) $\hat{\sigma}^2$ suit une loi de Chi-carré.
$\Rightarrow$ permet de construire des intervalles de confiance et de testes des hyp. stat.

Remarque :

Si $u_i \sim N(0, \sigma^2)$, alors Y_i (qui est une fct° linéaire de u_i) suit une distribution normale avec les moyenne et variance suivantes :

Moyenne : $E(Y_i) = \beta_1 + \beta_2 X_i$
Variance : $var(Y_i) = \sigma^2$

$$\Rightarrow \boxed{Y_i \sim i.i.d.\ N(\beta_1 + \beta_2 X_i\,,\ \sigma^2)}$$

1.4. Les tests de normalité

Comment peut-on s'assurer que les perturbations suivent une distribution normale ?

a) Histogramme des résidus

b) Test de Jarque - Bera (test asymptotique)

Basé sur résidus des MCO

$$\boxed{JB = n \cdot \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right]}$$

où n = taille de l'échantillon

S = coeff. d'asymétrie („skewness")

K = coeff. d'aplatissement („kurtosis")

avec $S = \dfrac{E(X-\mu)^3}{\sigma^3}$ et $K = \dfrac{E(X-\mu)^4}{[E(X-\mu)^2]^2}$

Sous l'hypothèse de normalité : $S = 0$ et $K = 3$

→ Test de Jarque - Bera teste hyp. jointe que $S = 0$ et $K = 3$

→ Lorsque résidus suivent une distribut° normale : $JB = 0$

Statistique $JB \sim \chi^2$ à 2 degrés de liberté (asympt.) sous Ho de normalité du terme d'erreur.

Règle de décision : RHo si p-value est faible.

Illustration :

$$\widehat{FoodExp}_i = 94{,}2087 + 0{,}4368 \; TotalExp_i$$
$$s.e. = (50{,}8563) \quad (0{,}0783)$$
$$t = (1{,}8524) \quad (5{,}5770)$$
$$p = (0{,}0695) \quad (0{,}0000)$$
$$R^2 = 0{,}3698 \qquad df = 53$$
$$F_{1,53} = 31{,}1034 \quad (p\text{-value} = 0{,}0000)$$

$H_0 : \beta_2 = 0$
Règle :
RH_0 si $|t| > 2$
($\alpha = 5\%$)

où FoodExp = dépenses d'alimentation
TotalExp = proxy du revenu
$n = 55$ (ménages indiens)
unités = rupees

- $H_0 : \beta_2 = 0{,}5$? $\longrightarrow$ sous H_0 : $t = \dfrac{(0{,}4368 - 0{,}5)}{0{,}0783} = -0{,}8071$

p-value > 20% $\longrightarrow$ $\cancel{R}H_0$

- Terme d'erreur suit une distribution normale ?

$JB = 0{,}257587$ (p-value = 0,879) $\longrightarrow$ $\cancel{R}H_0$

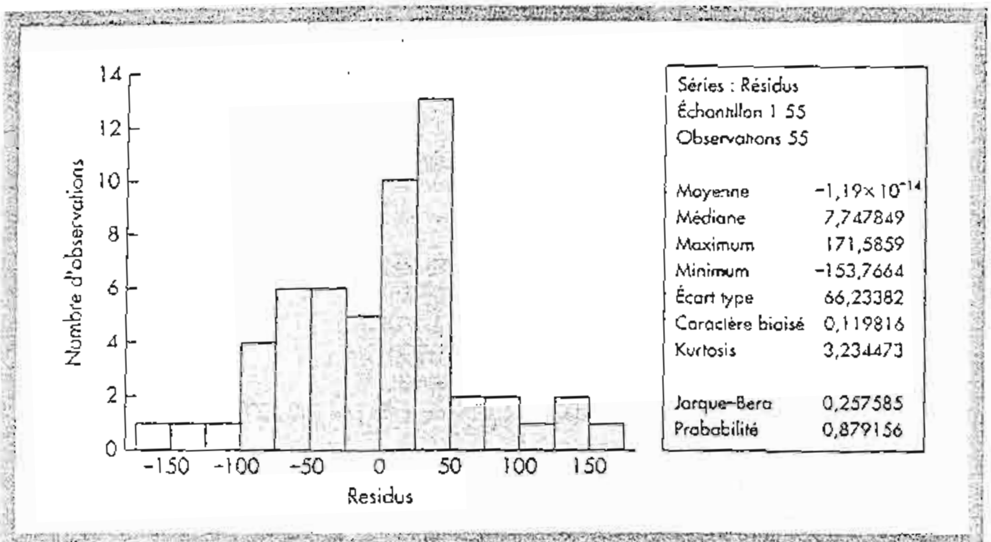


Figure 5.8
Résidus de la régression de la dépense d'alimentation

2. Relâchement des hypothèses du modèle classique

- Rappel : Hypothèses du MRLCN

1. Le modèle est linéaire dans les paramètres.
2. Les variables explicatives sont non stochastiques.
3. Pour des valeurs fixes de X, la valeur moyenne des erreurs u_i est nulle.
4. Homoscédasticité.
5. Absence d'autocorrélation des erreurs u_i.
6. Si les variables X sont stochastiques, le terme d'erreur u_i et les variables X sont indépendants ou du moins non corrélés.
7. Le nbre d'observations > nbre de variables explicatives.
8. $\exists$ suffisamment de variabilité ds les valeurs prises par les variables explicatives.
9. Le modèle est correctement spécifié.
10. $\nexists$ de relation linéaire exacte ent les variables explicatives (pas de multicolinéarité).
11. Le terme d'erreur u_i suit une loi normale.

- Quid si hyp. 8 pas satisfaite ?

Les estimateurs des MCO sont BLUE mais paramètres pop. ne sont pas estimés avec une gde précision.

- Quid si hyp. 7 ne pas satisfaite ?

Micronumérosité : estimateurs tjs BLUE mais s.e. gds par rapport aux coeff. estimés.

• Hypothèse 2 :

On suppose que pour pblme particulier, les valeurs prises par les variables expl. sont fixes, m̂ si les variables expl. elles-m̂ sont intrinsèqmt stochastiques.

Si variable est stochastique : vérifier si var. expl. est indpdte du terme d'erreur ou du moins si les valeurs de la var. expl. sont non corrélées au terme d'erreur.

Si hypothèse 6 est satisfaite : estimateurs des MCO pvt être biaisés mais ils sont consistants.

Si hypothèse 6 n'est pas satisfaite : estimateurs des MCO sont biaisés et inconsistants (variables instrumentales).

• Hypothèse 3 :

Espérance cond. des erreurs est nulle.
Si hyp. 3 non satisfaite ?

$$\begin{cases} Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \\ E(u_i / X_{2i}, X_{3i}, \dots, X_{ki}) = \underset{\sim}{\omega} \end{cases}$$

où ω = constante $(\neq 0)$

$$\Rightarrow E(Y_i / X_{2i}, X_{3i}, \dots, X_{ki}) = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \underset{\sim}{\omega}$$

$$= (\beta_1 + \omega) + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

$$= \alpha + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

où $\alpha = (\beta_1 + \omega)$

Estimation biaisée de β_1, autres coeff. sont BLUE

Rem : Si $E(u_i) = w_i$, alors coeff. de pente peuvent être biaisés et inconsistants.

- Hypothèse 11 :

Normalité du terme d'erreur.
Estimateurs des MCO sont BLUE indépdmt de l'hyp. 11.
Sous hyp. de normalité :
a) estimateurs des coeff. de rég. (MCO) suivent loi normale.
b) $(n-k)\,\hat{\sigma}^2/\sigma^2 \sim \chi^2$
c) test „t" et „F" peuvent être utilisés pour inférence statistique indépdmt taille de l'échantillon.

Quid si hyp. 11 n'est pas satisfaite ?

Théorème limite central : si termes d'erreur u_i sont indpts et suivent une m̂ distribution de moyenne „0" et de variance σ^2 (constante) et que variables expl. non stochast.
$\Rightarrow$ coeff. de régression des MCO suivent asympt. une distrib. normale dont moyenne est égale à la vraie. valeur des paramètres pop. $\Rightarrow$ tests „t" et „F" restent valables asympt.

Corollaire :

Si échant. est petit, tests „t" et „F" ne peuvent plus être utilisés pour faire de l'inférence stat.
(techniques d'estimation „non paramétriques").