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'observation > nbre de variables explicatives.
8. ∃ suffisamment de variabilité ds les valeurs prises par les variables explicatives.
9. Le modèle est correctement spécifié.
10. ∄ 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 : estimateur des MCO put 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 ?

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

où ω = constante $(\neq 0)$

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

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").

2.1. La multicolinéarité

Quelle est la nature de la multicolinéarité ?
Est-ce vraiment un problème ?
Quelles sont les csquences pratiques ?
Comment la détecter ?
Comment y remédier ?

(Hypothèse 10 – 7 – 8)

2.1.1. Nature de la multicolinéarité

Multicolinéarité = existence d'une relation „linéaire" „exacte" parmi certaines ou ttes les variables explicatives.

$$\boxed{\lambda_1 X_{1i} + \lambda_2 X_{2i} + \ldots + \lambda_k X_{ki} = 0} \quad (1) \equiv \text{relat° lin. exacte}$$

où $\lambda_1, \ldots, \lambda_k$ des constantes non toutes nulles
X_{1i} = intercepte ; $X_{2i} \ldots X_{ki}$ = variables explicatives.

Aujourd'hui, multicolinéarité désigne l'∃ d'une relation linéaire parfaite ou indique que les variables explicatives sont corrélées (m̂ impfmt).

$$\boxed{\lambda_1 X_{1i} + \lambda_2 X_{2i} + \ldots + \lambda_k X_{ki} + v_i = 0} \quad (2) \equiv \text{rel. lin. impfh}$$

où $\lambda_1, \ldots, \lambda_k$ des constantes non ttes nulles et v_i un terme d'erreur stochastique.

Soit, $\lambda_2 \neq 0$:

$$(1) : \boxed{X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \ldots - \frac{\lambda_k}{\lambda_2} X_{ki}}$$

$$(2) : \boxed{X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \ldots - \frac{\lambda_k}{\lambda_2} X_{ki} - \frac{1}{\lambda_2} v_i}$$

- Exemple :

X_{2i}	X_{3i}	X_{3i}^*
10	50	52
15	75	75
18	90	97
24	120	129
30	150	152

$X_{3i} = 5 X_{2i}$ $\quad (r_{23} = 1)$

X_3^* ne peut pas être écrite comme une combili de X_2

- Hyp. 10 ne concerne pas les relations non linéaires :

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

où $\begin{cases} Y = \text{coût total de production} \\ X = \text{niveau de production (= l'output)} \\ X^2 = \text{output au carré} \\ X^3 = \text{output au cube} \end{cases}$

Relation non linéaire tel X, X_i^2 et X^3 $\longrightarrow$ ne viole pas l'hyp. 10
Multicolinéarité imparfaite.

- Principales sources de multicolinéarité ?
  1. Méthode de collecte des données.
  2. Contraintes sur la population dont on tire l'échantillon.
  3. La spécification du modèle.
  4. Sur-identification du modèle.
  5. Les variables expl. ont une tendance commune.

2.1.2. Estimation en présence de multicolinéarité pfte

En présence de multicolinéarité pfte, coeff. de régr. des variables expl. sont indéterminés.

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \qquad (1)$$

$$\hat{\beta}_2 = \frac{(\Sigma y_i x_{2i})(\Sigma x_{3i}^2) - (\Sigma y_i x_{3i})(\Sigma x_{2i} x_{3i})}{(\Sigma x_{2i}^2)(\Sigma x_{3i}^2) - (\Sigma x_{2i} x_{3i})^2} \qquad (2)$$

$$\hat{\beta}_3 = \frac{(\Sigma y_i x_{3i})(\Sigma x_{2i}^2) - (\Sigma y_i x_{2i})(\Sigma x_{2i} x_{3i})}{(\Sigma x_{2i}^2)(\Sigma x_{3i}^2) - (\Sigma x_{2i} x_{3i})^2} \qquad (3)$$

Soit $X_{3i} = \lambda X_{2i}$

$$\Rightarrow \hat{\beta}_2 = \frac{(\Sigma y_i x_{2i})(\lambda^2 \Sigma x_{2i}^2) - (\lambda \Sigma y_i x_{2i})(\lambda \Sigma x_{2i}^2)}{(\Sigma x_{2i}^2)(\lambda^2 \Sigma x_{2i}^2) - \lambda^2 (\Sigma x_{2i}^2)^2}$$

$$= \frac{0}{0}$$

Impossible de distinguer l'influence qui est exercée sur Y par X_2 et X_3.

<u>Alternative</u> : substituer $X_{3i} = \lambda X_{2i}$ dans (1)

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 (\lambda x_{2i}) + \hat{u}_i$$
$$= (\hat{\beta}_2 + \lambda \hat{\beta}_3) x_{2i} + \hat{u}_i$$
$$= \hat{\alpha} x_{2i} + \hat{u}_i$$

où $\hat{\alpha} = (\hat{\beta}_2 + \lambda \hat{\beta}_3)$

$$\hat{\alpha} = (\hat{\beta}_2 + \lambda \hat{\beta}_3) = \frac{\Sigma x_{2i} y_i}{\Sigma x_{2i}^2}$$

Paramètre α peut être estimé de façon unique, il n'en va pas de même pour les paramètres β_2 et β_3.

2.1.3. Estimation en présence d'une forte multicolinéarité (imparfaite)

$$\boxed{x_{3i} = \lambda x_{2i} + v_i} \quad (1)$$

où $\lambda \neq 0$ et v_i terme d'erreur stochastique tq. $\boxed{\Sigma x_{2i} v_i = 0}$

Estimation unique des coeff. de régression β_1 et β_2. est possible.

En substituant (1) dans formule de $\hat{\beta}_2$:

$$\hat{\beta}_2 = \frac{(\Sigma y_i x_{2i})(\lambda^2 \Sigma x_{2i}^2 + \Sigma v_i^2) - (\lambda \Sigma y_i x_{2i} + \Sigma y_i v_i)(\lambda \Sigma x}{\Sigma x_{2i}^2 (\lambda^2 \Sigma x_{2i}^2 + \Sigma v_i^2) - (\lambda \Sigma x_{2i}^2)^2}$$

Quand $v_i \to 0$, on tend vers multicolinéarité pfte et indétermination.

2.1.4. La multicolinéarité : bcp de bruit pour rien

Multicolinéarité a pour seule cséquence d'augmenter l'erreur standard des estimateurs (càd. de rendre peu précise l'estimation des paramètres population).

(idem que lorsque pblme de micronumérosité ou peu de variabilité dans les variables explicatives)

Ne détruit pas propriétés stat. des estimateurs MCO (ils sont tjs BLUE)

→ pquoi en parle-t-on autant ?

1) Le fait qu'un estimateur soit non biaisé ne donne aucune garantie concernant valeur de l'estimation pour un échantillon particulier.

2) Estimateurs de variance minimale ds ensemble de classe des estimateurs non biaisés (efficients) ... mais ne garantit pas que la variance des estimateurs sera faible au regard des coeff. de régression.

1.1.5. Conséquences pratiques de la multicolinéarité impfte

1) Estimateurs des MCO sont tjs BLUE mais larges variances et covariance → estimation peu précise des paramètres pop.
2) Intervalles de confiance relativmt larges → on accepte trop svt. l'hypothèse nulle
3) t - stat. associées à un ou +ieurs coeff. de régression sont non significatives.
4) R^2 peut être très élevé
5) Estimateurs des MCO sensibles à de petits chgmts dans les données.

Estimateurs des MCO ont de larges variances et covariances

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_{2i}^2 (1 - r_{23}^2)}$$

$$var(\hat{\beta}_3) = \frac{\sigma^2}{\Sigma x_{3i}^2 (1 - r_{23}^2)}$$

$$cov(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\, \sigma^2}{(1 - r_{23}^2)\sqrt{\Sigma x_{2i}^2 \, \Sigma x_{3i}^2}}$$

où r_{23} = coeff. de corrél. lel X_2 et X_3

Lorsque $r_{23} \rightarrow 1$, $\left.\begin{matrix} var(\hat{\beta}_2) \\ var(\hat{\beta}_3) \end{matrix}\right\} \rightarrow \infty$

$$\boxed{VIF = \frac{1}{(1-r_{23}^2)}} \equiv \text{Variance inflating factor}$$

$r_{23} \longrightarrow 1$, $VIF \longrightarrow \infty$

$r_{23} = 0$, $VIF = 1$

$$\boxed{var(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_{2i}^2} VIF}$$

$$\boxed{var(\hat{\beta}_3) = \frac{\sigma^2}{\Sigma x_{3i}^2} VIF}$$

Table : The effect of increasing r_{23} on $var(\hat{\beta}_2)$ and $cov(\hat{\beta}_2, \hat{\beta}_3)$

Value of r_{23} (1)	VIF (2)	$var(\hat{\beta}_2)$ (3)	$\frac{var(\hat{\beta}_2)\ (r_{23} \neq 0)}{var(\hat{\beta}_2)\ (r_{23} = 0)}$ (4)	$cov(\hat{\beta}_2, \hat{\beta}_3)$ (5)
0,00	1,00	$\frac{\sigma^2}{\Sigma x_{2i}^2} = A$	—	0
0,50	1,33	1,33 × A	1,33	0,67 × B
0,70	1,96	1,96 × A	1,96	1,37 × B
0,80	2,78	2,78 × A	2,78	2,22 × B
0,90	5,26	5,26 × A	5,26	4,73 × B
0,95	10,26	10,26 × A	10,26	9,74 × B
0,97	16,92	16,92 × A	16,92	19,41 × B
0,99	50,25	50,25 × A	50,25	49,75 × B
0,995	100,00	100,00 × A	100,00	99,50 × B
0,999	500,00	500,00 × A	500,00	499,50 × B

Note : $A = \frac{\sigma^2}{\Sigma x_{2i}^2}$, $B = \frac{-\sigma^2}{\sqrt{\Sigma x_{2i}^2 \Sigma x_{3i}^2}}$

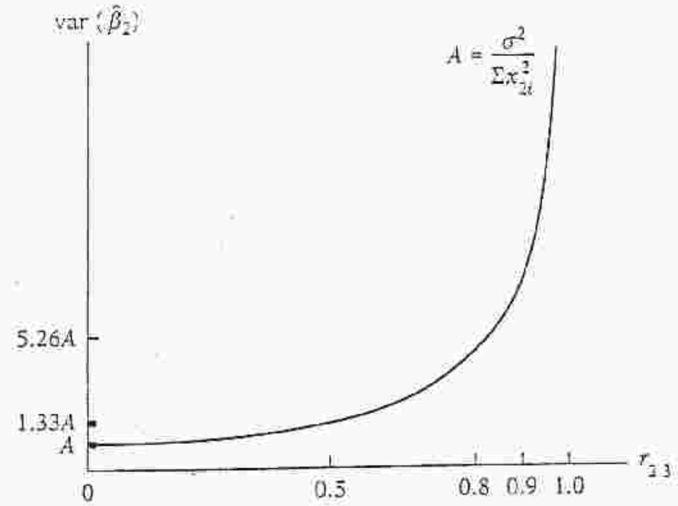


FIGURE 10.2 The behavior of var $(\hat{\beta}_2)$ as a function of r_{23}.

- <u>Extension pour modèle à k variables</u>

$$\boxed{var(\hat{\beta}_j) = \frac{\sigma^2}{\Sigma x_j^2}\left(\frac{1}{1-R_j^2}\right)}$$

où
$$\begin{cases} \hat{\beta}_j = \text{coeff. rég. partiel relatifs à } X_j \\ R_j^2 = R^2 \text{ de la régression de } X_j \text{ sur } (k-2) \text{ variables expl. restantes} \\ \Sigma x_j^2 = \Sigma (X_j - \bar{X}_j)^2 \end{cases}$$

$$\Longrightarrow \boxed{var(\hat{\beta}_j) = \frac{\sigma^2}{\Sigma x_j^2} \; VIF_j}$$

où $VIF_j = \dfrac{1}{1-R_j^2}$

De larges intervalles de confiance

Table : The effect of increasing collinearity on the 95% CI for β_2 $(\hat{\beta}_2 \pm 1{,}96\ se(\hat{\beta}_2))$

Value of r_{23}	95% confidence interval for β_2
0,00	$\hat{\beta}_2 \pm 1{,}96 \sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,50	$\hat{\beta}_2 \pm 1{,}96 . \sqrt{1{,}33} \sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,95	$\hat{\beta}_2 \pm 1{,}96 . \sqrt{10{,}26} \sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,995	$\hat{\beta}_2 \pm 1{,}96 \sqrt{100} \sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,999	$\hat{\beta}_2 \pm 1{,}96 \sqrt{500} \sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$

Rem : $\begin{cases} \text{hyp. } \sigma^2 \text{ est connue} \longrightarrow \text{on utilise distrib. normale} \\ \text{valeur critique à } 5\% = 1{,}96 \end{cases}$

Lorsque $r_{23} \uparrow \longrightarrow$ risque de second espèce $\uparrow \longrightarrow$ puissance du test $\downarrow$

Des t stat non significatives

Exemple : $H_0 : \beta_2 = 0 \longrightarrow t = \hat{\beta}_2 / se(\hat{\beta}_2)$

Un R^2 élevé mais peu de t stat significative

R^2 élevé , F stat significative

Sensibilité des estimateurs des MCO et de leurs se à de petits chgmts de données

Table A : Hypothetical data on Y, X_2 and X_3

Y	X_2	X_3
1	2	4
2	0	2
3	4	12
4	6	0
5	8	16

$$\hat{Y}_i = \underset{(0,7737)}{1,1939} + \underset{(0,1848)}{0,4463}\, X_{2i} + \underset{(0,0851)}{\underline{0,0030}}\, X_{3i}$$

$$t = (1,5431) \quad (\underline{2,4151}) \quad (0,0358)$$

$$R^2 = 0,8101 \ , \ \underline{r_{23} = 0,5523}$$

$$\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3) = \underline{-0,00868} \qquad df = 2$$

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0 \quad \Rightarrow \quad |t| \not> t_{\alpha/2},\ df \qquad (\alpha = 0,10)$$

valeur critique à 5% lorsque 2 df (Student)

$$(t_{0,025}\ ;\ 2df = 4,303 \ ;\ t_{0,005}\ ;\ 2df = 9,925)$$
$$t_{0,050}\ ;\ 2df = 2,920 \ ;\ t_{0,10},\ 2df = 1,886$$

$$H_0 : \beta_2 \leq 0$$
$$H_1 : \beta_2 > 0 \quad \Rightarrow \quad t > t_{\alpha},\ df \qquad (\alpha = 0,10)$$

$$(t_{0,010}\ ;\ 2df = 1,886)$$

Table B : Hypothetical data on Y, X_2 and X_3

Y	X_2	X_3
1	2	4
2	0	2
3	4	0
4	6	12
5	8	16

$$\hat{Y}_i = \underset{(0,7480)}{1,2108} + \underset{(0,2721)}{0,4014}\, X_{2i} + \underset{(0,1252)}{\underline{0,0270}}\, X_{3i}$$

$$t = (1,6187) \quad (\underline{1,4752}) \quad (0,2158)$$

$$R^2 = 0,8143 \ , \ \underline{r_{23} = 0,8285}$$

$$\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2) = \underline{-0,0282} \qquad df = 2$$

$$H_0 : \beta_2 \leq 0$$
$$H_1 : \beta_2 > 0 \quad \Rightarrow \quad t \not> t_{\alpha},\ df$$

(m̂ pblme en présence de micronumérosité)

Exemple : Relation lel dép. de cons., revenus du travail, richesse

$$\left[\begin{array}{llll} \hat{Y}_i = & 24,7747 & + \; 0,9415 \; X_{2i} & - \; 0,0424 \; X_{3i} \\ & (6,7525) & (0,8229) & (0,0807) \\ t = & (3,6690) & (1,1442) & (-0,5261) \end{array}\right.$$

$R^2 = 0,9635$, $\bar{R}^2_a = 0,9531$

$n = 10$, $df = 7$ $(= 10-3)$

ANOVA table for consumption - income - wealth example

Source of variation	SS	df	MSS
Due to regression	8,565 , 5541	2	4 282, 7770
Due to residual	324, 4459	7	46, 3494

$H_0 : \beta_2 = \beta_3 = 0$

$$\left.\begin{array}{l} F = \dfrac{4282,7770}{46,3494} = 92,4019 \\ \\ F_{0,05} = 4,74 \quad ; \quad F_{0,01} = 9,55 \end{array}\right\} \quad \begin{array}{c} F_{\text{calculée}} > F_\alpha(2,7) \\ (\text{pour } \alpha = 0,01) \end{array}$$

$$\left[\begin{array}{lll} \hat{X}_{3i} = & 7,5454 & + \; 10,1909 \; X_{2i} \\ & (29,4758) & (0,1643) \\ t = & (0,2560) & (62,0405) \end{array}\right. \qquad R^2 = 0,9979$$

$$\left[\begin{array}{lll} \hat{Y}_i = & 24,4545 & + \; 0,5091 \; X_{2i} \\ & (6,4138) & (0,0357) \\ t = & (3,8128) & (14,2432) \end{array}\right. \qquad R^2 = 0,9621$$

$$\left[\begin{array}{lll} \hat{Y}_i = & 24,411 & + \; 0,0498 \; X_{3i} \\ & (6,874) & (0,0037) \\ t = & (3,551) & (13,29) \end{array}\right. \qquad R^2 = 0,9567$$

2.1.6. Détection de la multicolinéarité

1) L'obtention d'un R^2 élevé et de peu de t stat signif.
Décèle uniq. multicol. aigue
2) Des corrélations bi-variées élevées. $(r > 0{,}8)$
Mais cdt° suff. mais non nécess.
3) Les régressions auxiliaires : calculer $R^2_{x_i}$
Exemple : $R^2_{x_2}$ obtenus à partir de $X_{2i} = a_1 + a_3 X_{3i} + \ldots + a_k X_k$

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

$\Rightarrow$ stat. de test : $\boxed{F_i = \frac{R^2_{x_i} \;/\; (k-2)}{(1-R^2_{x_i})/(n-k+1)}}$

où $\begin{cases} n = \text{taille de l'éch.} \\ k = \#\ \text{var expl.} + \text{intercepte} \\ R^2_{x_i} = \text{coeff. détermin. de } X_i \sim \hat{o}\ \text{var. expl. (intercpt compris)} \end{cases}$

F_i suit distribution de Fischer à $(k-2)$ et $(n-k+1)$ df
Règle de décision : $F_{i\,(\text{calculée})} > F_\alpha\,[(k-2), (n-k+1)]$

$\Rightarrow$ RH$_0$ (on rejette absence de multicol. tel X_i et ô var. expl.)

Règle de Klien (1962) : pblme de multicol. est sérieux seulemt si R^2 (rég. auxiliaire) > R^2 (modèle complet)
Incovénient : détecte multicol. uniq. si elle implique un gd nbre de var. expl.

4) Le VIF

$$VIF_j = \frac{1}{1-R^2_j}$$

où R^2_j est obtenu à partir de $X_j \sim (k-2)$ var. expl. restantes (sans intercepte)

Génér^lmt : multicolinéarité lorsque $VIF_j > 10$ ($R^2_j > 0{,}90$)

Inconvénient : $var(\hat{\beta}_j) = \frac{\sigma^2}{\Sigma x^2_j} VIF_j$

$\Rightarrow$ VIF_j élevé $\neq$ cdt° nécess. suff. pour obtention erreur standard élevée pour estimateur.

2.1.7. Remèdes à la multicolinéarité

1) Utiliser l'information disponible a priori

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

où $\begin{cases} Y = \text{dép. cons.} \\ X_2 = \text{revenu du travail} \\ X_3 = \text{richesse} \end{cases}$

$$\beta_3 = 0{,}10\ \beta_2 \Rightarrow Y_i = \beta_1 + \beta_2 X_{2i} + 0{,}10\ \beta_2 X_{3i} + u_i$$
$$= \beta_1 + \beta_2 X_i + u_i$$

où $X_i = X_{2i} + 0{,}10\ X_{3i}$

2) Combiner des données en coupes et des données longitud.
$\Rightarrow$ „pooling" des données.

Exemple : demande de pain aux USA („time series")

$$\boxed{\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t} \quad (1)$$

où $\begin{cases} Y = \text{nbre de pains vendus} \\ P = \text{prix moyen du pain} \\ I = \text{revenu moyen des consommateurs} \\ t = \text{temps} \end{cases}$

Solution de Tobin (1950) : utiliser un „cross-section" pour estimer l'élasticité-revenu β_3

$$\Rightarrow \boxed{\ln Y_i = \beta_1 + \beta_2 \ln P_i + \beta_3 \ln I_i + u_i} \quad (2)$$

où $\begin{cases} Y_i = \text{nbre de pains achetés par l'individu „}i\text{" au temps „}t\text{"} \\ P_i = \text{prix moyen des pains achetés par l'individu „}i\text{" au temps „}t\text{"} \\ I_i = \text{revenu moyen de l'individu „}i\text{" au temps „}t\text{"} \end{cases}$

Permet d'obtenir une estimation relativement précise de β_3
$\Rightarrow \hat{\beta}_3$

En utilisant cette estimation, (1) peut être réécrite comme :

$$\boxed{Y_t^* = \beta_1 + \beta_2 \ln P_t + u_t} \quad (3)$$

où $Y_t^* = \ln Y_t - \hat{\beta}_3 \ln I_t$

En estimant (3) on obtient estimation de β_2 en évitant pblme de colinéarité lel P_t et I_t.

! préférable de ne pas utiliser cette technique si estimation en cross-section varient fortement dans le temps

3). Exclure certaines variables en évitant un pblme de spécification.

4) Transformer les variables

$$\boxed{Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t} \qquad (1)$$

si valable en t $\longrightarrow$ valable en $(t-1)$

$$\boxed{Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1}} \qquad (2)$$

(1) - (2) :

$$\boxed{Y_t - Y_{t-1} = \beta_2 (X_{2,t} - X_{2,t-1}) + \beta_3 (X_{3t} - X_{3,t-1}) + v_t} \quad (3$$

où $v_t = u_t - u_{t-1}$

Ecriture du <u>modèle en différences premières</u> permet de réduire prblme de colinéarité, + (ds certains cas) permet de rendre une série stationnaire.

Autre transformation : „<u>transf. en ratio</u>"

$$\boxed{Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t}$$

où

$$\begin{cases} Y = \text{dép. de cons en \$ réels} \\ X_2 = \text{PIB} \\ X_3 = \text{pop. totale} \end{cases}$$

PIB et pop. ont une tendance commune $\Rightarrow$ colinéarité

$$\Rightarrow \boxed{\frac{Y_t}{X_{3t}} = \beta_1 \left(\frac{1}{X_{3t}}\right) + \beta_2 \left(\frac{X_{2t}}{X_{3t}}\right) + \beta_3 + \left(\frac{u_t}{X_{3t}}\right)}$$

où $\begin{cases} Y/X_3 = \text{dép. de cons par tête} \\ X_2/X_3 = \text{PIB par tête} \end{cases}$

⚠ i) Transformation en ≠ premières :

* (ds la plupart des cas) lorsque les erreurs originales u_t sont non corrélées, leurs ≠ prem. v_t le seront.
* perte d'1 df.
* inadaptée pour données en coupe.

ii) Transformation en ratio :

* lorsque terme d'erreur original u_t est homosced. nveau terme d'erreur (u_t / X_{3t}) sera hétéroscédast.

5) Utiliser un autre échantillon ou ajouter des observations

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_{2i}^2 (1 - r_{23}^2)}$$

si $n \uparrow \longrightarrow \Sigma x_{2i}^2 \uparrow$

Exemple : relation entre dép. de cons. (Y), revenu (X_2) et X_3 (richesse)

$n = 10 \Rightarrow \hat{Y}_i = 24,377 + 0,8716\, X_{2i} - 0,0349\, X_{3i}$
$t = (3,875) \quad (2,7726) \quad (-1,1595)$
$R^2 = 0,9682$

$n = 40 \Rightarrow \hat{Y}_i = 2,0907 + 0,7299\, X_{2i} + 0,0605\, X_{3i}$
$t = (0,8713) \quad (6,0014) \quad (2,0014)$
$R^2 = 0,9672$

6) Autres méthodes tq. l'analyse factorielle, ACP.

2.2. L'hétéroscédasticité

Quelle est la nature de l'hétéroscédasticité ?

Quelles en sont les csquences ?

Comment la détecter ?

Comment y remédier ?

2.2.1. Nature de l'hétéroscédasticité

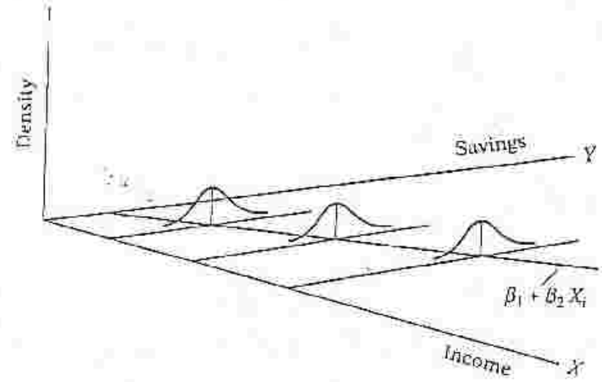


FIGURE 11.1 Homoscedastic disturbances.

$\boxed{E(u_i^2) = \sigma^2}$ $\forall i$

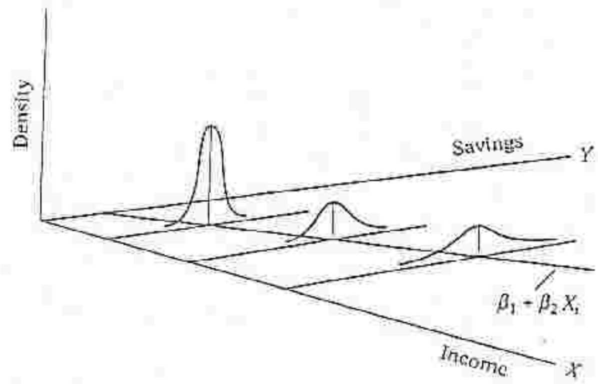


FIGURE 11.2 Heteroscedastic disturbances.

$\boxed{E(u_i^2) = \sigma_i^2}$

$Y_i = \beta_1 + \beta_2 X_i + u_i$

où $\begin{cases} Y = \text{l'épargne} \\ X = \text{revenu} \end{cases}$

- Causes de l'hétéroscédasticité

1) Existence d'un processus d'apprentissage

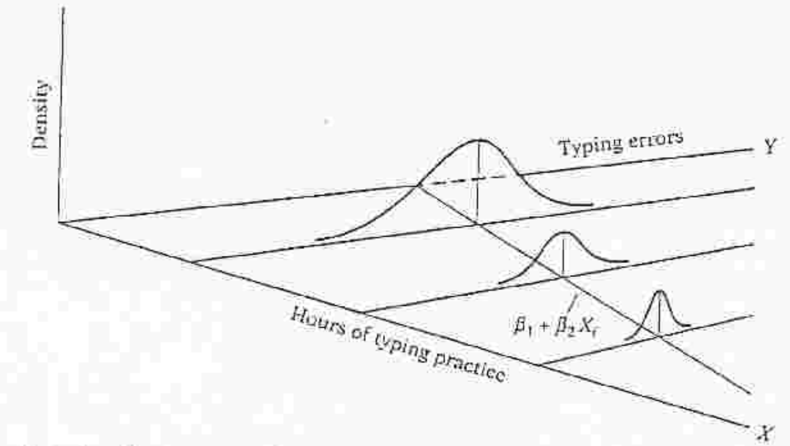


FIGURE 11.3 Illustration of heteroscedasticity.

2) Pouvoir discrétionnaire des individus ↑ avec leurs revenus

3) Existence de valeurs aberrantes (« outliers")

4) Mauvaise spécification du modèle de régression

5) Distribution d'une ou de +ieurs variables explicatives est asymétrique

6) Autres sources :

- transformation peu appropriée des données
- utilisation d'une forme fonctionnelle inadéquate

Concerne surtout les séries en „ cross-section".
qui portent génér^lmt sur les membres d'une population à un moment donné (tq consommateurs, familles, firmes, secteurs).

TABLE 11.1
COMPENSATION PER EMPLOYEE ($) IN NONDURABLE MANUFACTURING INDUSTRIES ACCORDING TO EMPLOYMENT SIZE OF ESTABLISHMENT, 1958

	Employment size (average number of employees)								
Industry	1–4	5–9	10–19	20–49	50–99	100–249	250–499	500–999	1000–2499
Food and kindred products	2994	3295	3565	3907	4189	4486	4676	4968	5342
Tobacco products	1721	2057	3336	3320	2980	2848	3072	2969	3822
Textile mill products	3600	3657	3674	3437	3340	3334	3225	3163	3168
Apparel and related products	3494	3787	3533	3215	3030	2834	2750	2967	3453
Paper and allied products	3498	3847	3913	4135	4445	4885	5132	5342	5326
Printing and publishing	3611	4206	4695	5083	5301	5269	5182	5395	5552
Chemicals and allied products	3875	4660	4930	5005	5114	5248	5630	5870	5876
Petroleum and coal products	4616	5181	5317	5337	5421	5710	6316	6455	6347
Rubber and plastic products	3538	3984	4014	4287	4221	4539	4721	4905	5481
Leather and leather products	3016	3196	3149	3317	3414	3254	3177	3346	4067
Average compensation	3396	3787	4013	4104	4146	4241	4388	4538	4843
Standard deviation	742.2	851.4	727.8	805.06	929.9	1080.6	1241.2	1307.7	1110.5
Average productivity	9355	8584	7962	8275	8389	9418	9795	10,281	11,750

Source: The Census of Manufacturers, U.S. Department of Commerce, 1958 (computed by author).

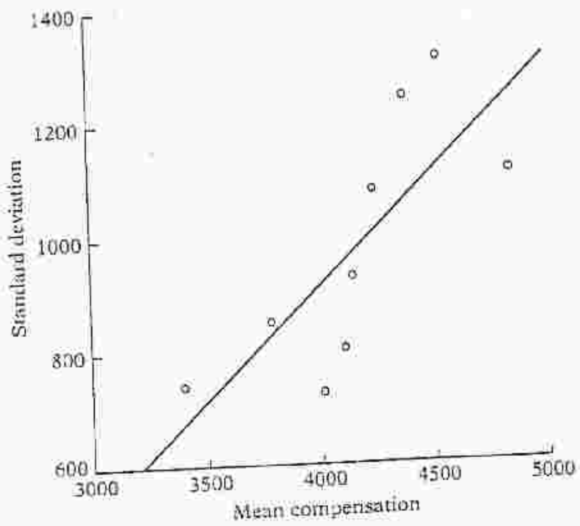


FIGURE 11.6 Standard deviation of compensation and mean compensation.

2.2.2. Estimation par MCO en présence d'hétéroscédasticité

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\boxed{var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}} \qquad \left(\text{si homoscéd. : } var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \right)$$

Sous hyp. modèle rég. lin. classique, $\hat{\beta}_2$ est BLUE
En présence d'hétéroscédasticité, $\hat{\beta}_2$ est linéaire, non biaisé et consistant. (pas efficient).

Idem pour $\hat{\beta}_1$

2.2.3. La méthode des moindres carrés généralisés (MCG)

Pq estimateur des MCO $\hat{\beta}_2$ n'est-il pas de variance minimale ?
On voudrait une procédure d'estimation qui accorderait un poids plus important aux observations provenant des classes de taille où la variabilité des salaires est + faible.
→ permettrait une estimation + précise de la fction de régression population.

MCO attribue un poids identique à ttes les observations.
MCG pondère les observations en fction de la variance conditionnelle du terme d'erreur → fournit des estimateurs BLUE en présence d'hétéroscédasticité (variance minimale dans l'ensemble de la classe des estimateurs linéaires non biaisés).

- $\boxed{Y_i = \beta_1 + \beta_2 X_i + u_i}$ (1)

$\Rightarrow \quad Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i$ (2)

où $X_{0i} = 1, \forall i$

Supposons que σ_i^2 soient connues.

$$\Rightarrow \quad \boxed{\frac{Y_i}{\sigma_i} = \beta_1 \left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2 \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right)} \equiv \text{FRP}$$

$$\Rightarrow \quad Y^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^*$$

- <u>Pq transformer le modèle de cette façon</u> ?

$$\begin{aligned}
\text{var}(u_i^*) &= E\left(u_i^* - E(u_i^*)\right)^2 \\
&= E(u_i^*)^2 && \text{car } E(u_i^*) = 0 \\
&= E\left(\frac{u_i}{\sigma_i}\right)^2 \\
&= \frac{1}{\sigma_i^2} \cdot E(u_i^2) && \text{où } \sigma_i^2 \text{ est connue} \\
&= \frac{1}{\sigma_i^2} \cdot \sigma_i^2 && \text{car } E(u_i^2) = \sigma_i^2 \\
&= 1
\end{aligned}$$

(u_i^* est homoscédastique !)

- <u>En résumé</u> : méthode des MCG consiste à appliquer les MCO aux variables transformées, càd. aux variables qui satisfont les hypothèses du modèle de régr. linéaire classique.

Estimateurs de MCG sont BLUE en présence d'hétéroscédast.

- <u>Comment obtenir estimateurs des MCG $\hat{\beta}_1^*$ et $\hat{\beta}_2^*$?</u>

$$FRE \equiv \frac{Y_i}{\sigma_i} = \hat{\beta}_1^* \left(\frac{X_{0i}}{\sigma_i}\right) + \hat{\beta}_2^* \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{\hat{u}_i}{\sigma_i}\right)$$

$$\Rightarrow \quad Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^*$$

$$\min \sum \hat{u}_i^{2*} \equiv \min \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2$$

$$\Rightarrow \min \sum \left(\frac{\hat{u}_i}{\sigma_i}\right)^2$$

$$\Rightarrow \min \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_1^* \left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_2^* \left(\frac{X_i}{\sigma_i}\right)\right]^2$$

soit $\boxed{w_i = \frac{1}{\sigma_i^2}}$

$$\Rightarrow \min \sum w_i \hat{u}_i^2$$

$$\Rightarrow \boxed{\min \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2}$$

MCG = minimise somme des carrés des résidus pondérée où poids w_i sont inversmt proport. à la variance conditionnelle des résidus $\hat{u}_i$ ($var(\hat{u}_i / X_i) = var(Y_i / X_i)$)

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \hat{\beta}_1^*} = 2 \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-1)$$

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \hat{\beta}_2^*} = 2 \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-X_i)$$

$$\sum w_i Y_i = \hat{\beta}_1^* \sum w_i + \hat{\beta}_2^* \sum w_i X_i$$
$$\sum w_i X_i Y_i = \hat{\beta}_1^* \sum w_i X_i + \hat{\beta}_2^* \sum w_i X_i^2$$

$\Rightarrow$

$$\hat{\beta}_1^* = \overline{Y}^* - \hat{\beta}_2^* \, \overline{X}^*$$

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}$$

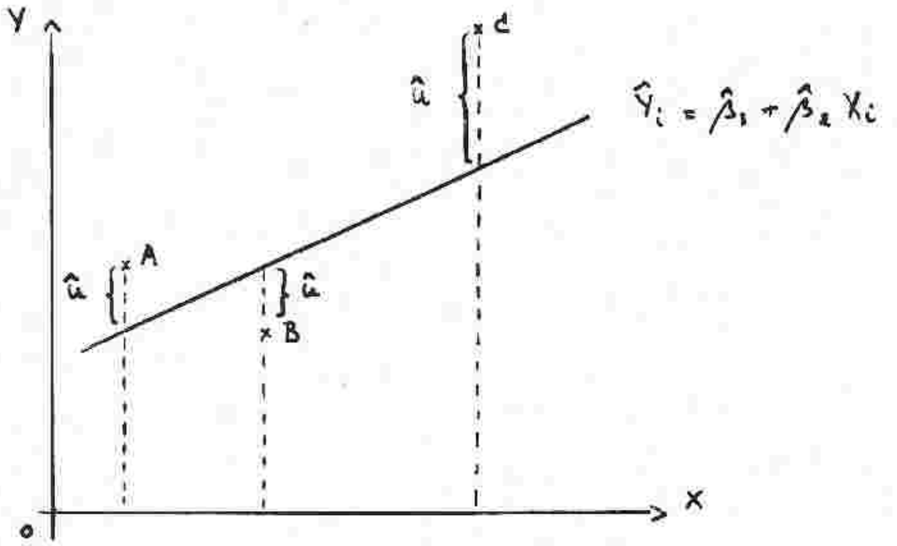
$$\text{var}(\hat{\beta}_2^*) = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}$$

- $\neq$ les MCO et MCG

MCO : $\min \sum \hat{u}_i^2 \equiv \min \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$

MCG : $\min \sum w_i \hat{u}_i^2 \equiv \min \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2$

où $w_i = \frac{1}{\sigma_i^2}$



Utilisation des MCG en présence d'hétéroscédasticité attribue un poids + élevé aux observations proches de leur moyenne com ce qui permet estimation + précise de la FRP.

2.2.4. Csqueuces de l'utilisation des MCO en présence d'hétéroscédasticité

Estimateurs $\hat{\beta}_2^*$ (MCG) et $\hat{\beta}_2$ (MCO) sont linéaires et non biaisés.
$\hat{\beta}_2^*$ est efficient (pas $\hat{\beta}_2$).

i) Estimation par MCO en tenant compte de l'hétérosc.

$$\hat{\beta}_2 = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

$$\text{var}(\hat{\beta}_2) = \frac{\Sigma x_i^2 \sigma_i^2}{(\Sigma x_i^2)^2}$$ (prend en compte hétérosc.)

Généralmt : $\text{var}(\hat{\beta}_2^*) < \text{var}(\hat{\beta}_2)$

(ou encore : $\text{var}(\hat{\beta}_2^*) \leq \text{var}(\hat{\beta}_2)$)

ii) Estimation par MCO en ignorant l'hétéroscédasticité

$$\hat{\beta}_2 = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_i^2}$$

$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_i^2}$ est un estimateur biaisé de la vraie variance de l'estimateur $\hat{\beta}_2$ en présence d'hétérosc. qui est donné par expression suivante : $\text{var}(\hat{\beta}_2) = \frac{\Sigma x_i^2 \sigma_i^2}{(\Sigma x_i^2)^2}$

$\Rightarrow$ En moyenne, on sur ou sous-estime la vraie valeur de la variance de $\hat{\beta}_2$. Généralmt, on ne sait pas si sur ou sous-estimation car dépdt relat° entre σ_i^2 et X.

Biais découle du fait que estimateur de la variance c^dtionnelle des erreurs : $\hat{\sigma}^2 = \frac{\Sigma \hat{u}_i^2}{(n-2)}$ est un estimateur biaisé de σ_i^2.

2.2.5. Comment détecter l'hétéroscédasticité ?

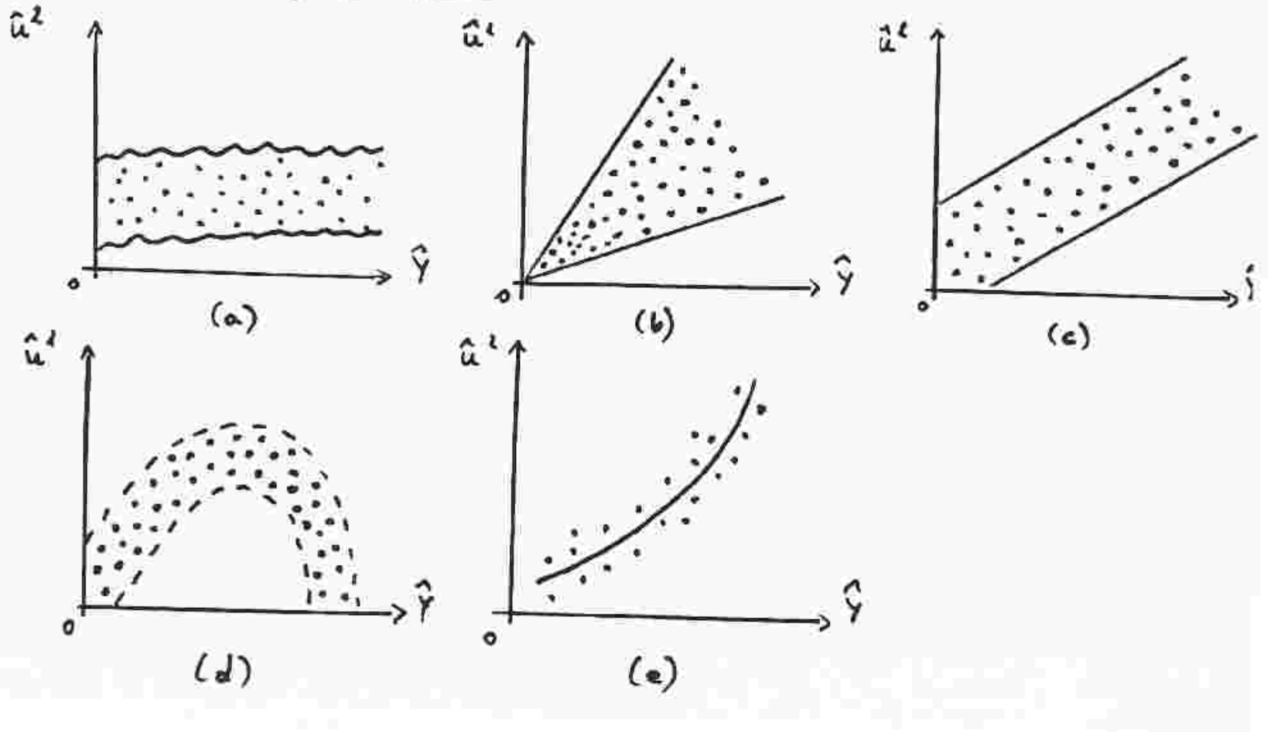
Repose en gde partie sur intuition et expérience.
Plupart des méthodes basées sur analyse de résidus des MCO $\hat{u}_i$ (pas sur terme d'erreur u_i). On espère que les résidus $\hat{u}_i$ fournissent bonne estimation des erreurs u_i (espérance générlmt satisfaite pour de gds échantillons).

i) Méthodes informelles

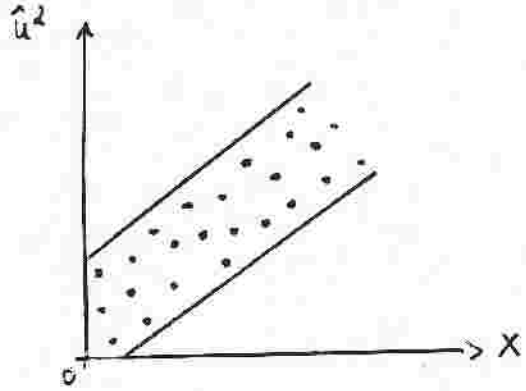
1) Méthode basée sur la nature du pblme.

Lorsqu'on travaille avec des données en « cross-section » faisant intervenir des unités hétérogènes, l'hétérosc. est davantage la règle que l'exception.

2) Méthode graphique



__Exemple :__ Epargne $\sim$ Revenu



Variance hétéroscédastique des résidus est proport. au revenu.

ii) __Méthodes formelles__

1) __Test de Park__

Hyp. : σ_i^2 est une fction de X_i

__Suggestion :__ $\sigma_i^2 = \sigma^2 X_i^{\beta} e^{v_i}$

$$\Rightarrow \boxed{\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i} \qquad (1)$$

où v_i = terme d'erreur stochastique

On utilise $\hat{u}_i^2$ comme proxy de σ_i^2 :

$$\begin{aligned} \ln \hat{u}_i^2 &= \ln \sigma^2 + \beta \ln X_i + v_i \\ &= \alpha \quad + \beta \ln X_i + v_i \end{aligned} \qquad (2)$$

__Règle de décision :__ si β est significatif $\longrightarrow$ prés. d'hétérosc.

__Inconvénient :__ v_i ne satisfait pas tjs hyp. modèle de régr. linéaire classique (peut m̂ s'avérer hétérosc.)

cf. Goldfeld et Quandt (1971)

2) Test de Goldfeld - Quandt

Hyp. : σ_i^2 dépend „positivement" d'une seule variable explicative du mod. de régression.

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\sigma_i^2 = \sigma^2 . X_i^2$$

Procédure :

- Etape 1 : ordonner observations en fct° des valeurs de X_i (en commençant par + petite valeur de X)
- Etape 2 : omettre „c" observations centrales et diviser les $(n-c)$ observations restantes en 2 groupes de $\frac{(n-c)}{2}$ observations.
- Etape 3 : Régression par MCO pour 2 sous-échantillons pris séparément, obtenir SCR de chaque régression

  $SCR_1 \equiv$ SCR pour groupe d'obs. dont variance est supposée faible.

  $SCR_2 \equiv$ SCR pour groupe d'obs. dont variance est supposée imp.

  Chaque SCR caractérisée par :

  $$\frac{(n-c)}{2} - k \text{ ou } \frac{(n-c-2k)}{2} \text{ df}$$

- Etape 4 : Calculer valeur de

$$\boxed{\lambda = \frac{SCR_2 / df}{SCR_1 / df}}$$

Si u_i suit loi normale et homoscéd. :

λ suit dist. Fisher avec $\left(\frac{n-c-2k}{2}\right)$ df

Règle de décision :

si λ calculée $> F_{\alpha}, \left(\frac{n-c-2k}{2}\right) df \Rightarrow$ rejette hyp. nulle d'homoscédasticité

Remarques :

- on omet „ c" observations centrale pour accentuer ≠ lel SCR_1 & SCR_2
- puissance du test (= 1 - risque 2nd espèce) dépend du nombre d'obs. „ c" omises.

Rappel : $\begin{cases} \text{risque 2d espèce} = \text{prob. d'accepter une hyp. fausse} \\ \text{puissance test} = \text{prob. rejeter l'hyp. nulle lorsqu'elle est fausse.} \end{cases}$

- d'après Judge et al. (1982), si $k = 2$
  $\Rightarrow$ „ c" = 4 lorsque $n = 30$
  „ c" = 10 lorsque $n = 60$

3) Test de Breusch - Pagan - Godfrey

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

Soit : $\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi})$

où $\begin{cases} Z = \text{variables non stochastiques} \\ \text{Certaines ou ttes les variables } X \text{ peuvent} \\ \text{servir comme variable } Z \end{cases}$

Plus spécifiquement :

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi}$$

Si $\alpha_2 = \alpha_3 = \dots = \alpha_m = 0 \Rightarrow \sigma_i^2 = \alpha_1$ (= constante)

1 2m

Procédure :

- Etape 1 : Estimer le modèle de régression linéaire à k variables par MCO, obtenir n résidus : $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n$
- Etape 2 : Calculer $\tilde{\sigma}^2 = \dfrac{\Sigma \hat{u}_i^2}{n}$ (estim. MV de σ^2) $\neq$ est. MCO $\hat{\sigma}^2 = \dfrac{\Sigma \hat{u}_i^2}{(n-k)}$
- Etape 3 : Construire variables p_i, où

$$p_i = \frac{\hat{u}_i^2}{\tilde{\sigma}^2}$$

- Etape 4 : Régresser les p_i sur variables Z :

$$p_i = \alpha_1 + \alpha_2 Z_{2i} + \ldots + \alpha_m Z_{mi} + v_i$$

où v_i = terme d'erreur

- Etape 5 : SCE de régress. des p_i sur cst et Z et calculer :

$$\Omega = \frac{1}{2} (SCE)$$

Sous hyp. normalité du terme d'erreur u_i, en présence d'homoscédasticité :

$$\Omega \underset{asy}{\sim} \chi^2_{m-1}$$

Règle de décision :

si Ω calculé $> \chi^2_{\alpha, (m-1)\, df}$ $\Rightarrow$ rejette hyp. nulle d'homoscédasticité.

- Exemple :

Echantillon de 30 obs. concernant dép. de cons. et revenu des ménages US en 1988.

Test d'homoscédasticité de Breusch - Pagan - Godfrey :

Etape 1 : $\hat{Y}_i = 9,2903 + 0,6378\ X_i$

$se = (5,2314) \quad (0,0286)$

$SCR = 2361,153 \quad ; \quad R^2 = 0,9466$

Etape 2 : $\tilde{\sigma}^2 = \dfrac{\Sigma \hat{u}_i^2}{30} = \dfrac{2361,153}{30} = 78,7051$

Etape 3 : On divise chaque carré de résidu $\hat{u}_i^2$ par $\tilde{\sigma}^2 = 78,7051$, pour obtenir p_i

Etape 4 : Hyp. : p_i dépend linéairement de X_i ($= Z_i$)

$\hat{p}_i = -0,7426 + 0,0101\ X_i$

$se = (0,7529) \quad (0,0041)$

$SCE = 10,4280 \quad ; \quad R^2 = 0,18$

Etape 5 : Calcule de la stat. de test :

$\Omega = \frac{1}{2}(SCE) = 5,2140$

$(\Omega \underset{asy}{\sim} \chi^2_{1df})$

$\chi^2_{0,05\,;\,1df} = 3,841$

$\Rightarrow$ rejette hyp. nulle d'homoscéd. à 5%.

⚠ Test de Breusch - Pagan - Godfrey est asympt.
+ vérifier hyp. de normalité de u_i.

4) Test de White

Ne nécessite pas hyp. de normalité de u_i (cf. B-P-G)
Ne requiert pas d'ordonner obs. d'après X supposée à la base du pblme d'hétérosc. (cf. G-Q)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (1)$$

Procédure :

- Etape 1 : Estimer modèle de rég. (1) et obtenir $\hat{u}_i$

- Etape 2 : Estimer régression auxiliaire :

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_4 X_{2i} X_{3i} + v_i \qquad (2)$$

Obtenir R^2 de cette rég.

- Etape 3 : $n . R^2 \underset{asy}{\sim} \chi^2_{df}$

df = nbre de var. expl. dans rég. auxiliaire (sans l'intercepte)

- Etape 4 : si $n . R^2 > \chi^2_{\alpha, df}$ $\Rightarrow$ rejette hyp. nulle d'homoscédasticité
$\Rightarrow$ rejette hyp nulle que $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$

Remarques :

- consomme svt bcp de df
- si nR^2 significatif, il est possible qu'il y ait une erreur de spécification plutôt qu'un pblme d'hétéroscédasticité.

Si produits croisés des variables pas inclus dans régression auxiliaire : test d'hétéroscédasticité pur.
Sinon, test de White teste simultanément pour hétéroscéd. et biais de spécification.

2.2.6. Remèdes à l'hétéroscédasticité

i) Lorsque σ_i^2 est connue : moindres carrés pondérés (MCP)

Moindres carrés pondérés (MCP) (Weighted least squares, WLS)
MCP = cas particulier de MCG, MCP fournit estim. BLUE

TABLE 11.1
COMPENSATION PER EMPLOYEE ($) IN NONDURABLE MANUFACTURING INDUSTRIES ACCORDING TO EMPLOYMENT SIZE OF ESTABLISHMENT, 1958

	Employment size (average number of employees)								
Industry	1–4	5–9	10–19	20–49	50–99	100–249	250–499	500–999	1000–2499
Food and kindred products	2994	3295	3565	3907	4189	4486	4676	4968	5342
Tobacco products	1721	2057	3336	3320	2980	2843	3072	2969	3822
Textile mill products	3600	3657	3674	3437	3340	3334	3225	3163	3168
Apparel and related products	3494	3787	3533	3215	3030	2834	2750	2967	3453
Paper and allied products	3498	3847	3913	4135	4445	4885	5132	5342	5326
Printing and publishing	3611	4206	4695	5083	5301	5269	5182	5395	5552
Chemicals and allied products	3875	4660	4930	5005	5114	5248	5630	5870	5876
Petroleum and coal products	4616	5181	5317	5337	5421	5710	6316	6455	6347
Rubber and plastic products	3538	3984	4014	4287	4221	4539	4721	4905	5481
Leather and leather products	3016	3196	3149	3317	3414	3254	3177	3346	4067
Average compensation	3396	3787	4013	4104	4146	4241	4388	4538	4843
Standard deviation	742.2	851.4	727.8	805.06	929.9	1080.6	1241.2	1307.7	1110.5
Average productivity	9355	8584	7962	8275	8389	9418	9795	10,281	11,750

Source: The Census of Manufacturers, U.S. Department of Commerce, 1958 (computed by author).

Soit $\begin{cases} Y & \text{le salaire moyen d'un travailleur en \$} \\ X & \text{la taille de l'entreprise} \end{cases}$

$$\boxed{Y_i/\sigma_i = \hat{\beta}_1^* (1/\sigma_i) + \hat{\beta}_2^* (X_i/\sigma_i) + (\hat{u}_i/\sigma_i)}$$

où σ_i = écart-type des erreurs pour chaque valeur de X

$$\widehat{(Y_i/\sigma_i)} = 3406,639 \ (1/\sigma_i) + 154,153 \ (X_i/\sigma_i)$$
$$(80,983) \qquad\qquad (16,959)$$
$$t = (42,066) \qquad\qquad (9,090)$$

$$\hat{Y}_i = 3417,833 + 148,767 \ X_i \qquad \Rightarrow \quad var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$
$$(81,136) \qquad\qquad (14,418)$$
$$t = (42,125) \qquad\qquad (10,318)$$

Remarque :

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \quad \text{et} \quad \hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{2*}}$$

où $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$

$$x_i^* = \left(\frac{X_i}{\sigma_i}\right) - \overline{\left(\frac{X_i}{\sigma_i}\right)}, \quad y_i^* = \left(\frac{Y_i}{\sigma_i}\right) - \overline{\left(\frac{Y_i}{\sigma_i}\right)}$$

(ii) Lorsque σ_i^2 est inconnue

Comment obtenir une estimation consistante de la variance des estimateurs des MCO en présence d'hétéroscédasticité ?

- White heteroscedasticity - consistent variances and stand. errors

Utiliser des erreurs stand. „robustes".

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (1)$$

où $var(u_i) = \sigma_i^2$

$$var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \quad (2) \Rightarrow \boxed{var(\hat{\beta}_2) = \frac{\sum x_i^2 \hat{u}_i^2}{(\sum x_i^2)^2}} \quad (3)$$

(3) est un estimateur consistant de (2)

Généralisation pour k variables :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \qquad (4)$$

$$\boxed{var(\hat{\beta}_j) = \frac{\sum \hat{\omega}_{ji}^2 \hat{u}_i^2}{(\sum \hat{\omega}_{ji}^2)^2}} \qquad (5)$$

où :

$\hat{u}_i$ = résidu de la régression (4)

$\hat{\omega}_{ji}$ = résidus d'une régression auxiliaire où on régresse X_j sur les autres régresseurs (var. expl. & intercepte) contenus dr modèle de régression (4).

<u>Exemple</u> : $\frac{\text{dépenses d'enseignmt}}{\text{tête}} \sim \frac{\text{revenu}}{\text{tête}}$, $\left(\frac{\text{revenu}}{\text{tête}}\right)^2$ dans 50 Etats américains en 1979.

$$\hat{Y}_i = 832,91 - 1834,2 \left(\frac{\text{Revenu}}{\text{tête}}\right) + 1587,04 \left(\frac{\text{Revenu}}{\text{tête}}\right)^2$$

se (MCO) =	(327,3)	(829,0)	(519,1)
t =	(2,54)	(2,21)	(3,06)
se (White) =	(460,9)	(1243,0)	(830,0)
t =	(1,81)	(-1,48)	(1,91)

Procédure uniq. valable pour de gds échantillons

Estimateurs obtenus par procédure de White généralmt moins efficients que ceux que l'on obtient lorsqu'on utilise une methode de transf. qui corrige pour hétérosc.

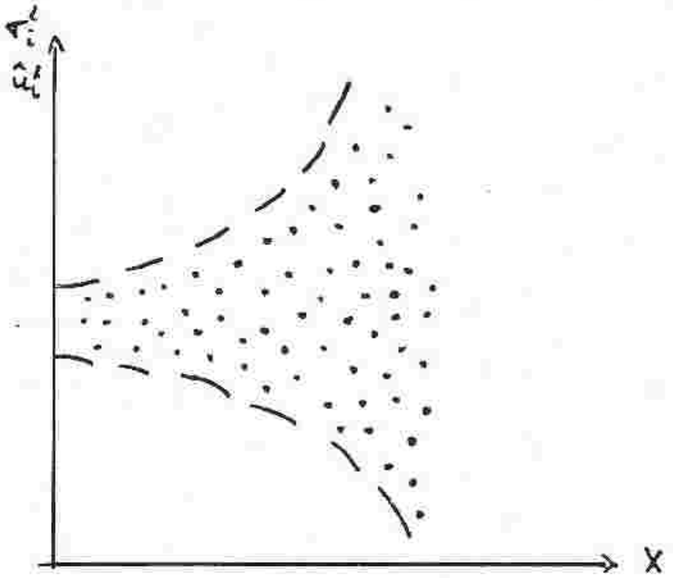
Rappel : $var(\hat{\beta}_2^*) \leq var(\hat{\beta}_2)$ où $\hat{\beta}_2^*$ estimateur MCP et $\hat{\beta}_2$ estimateur des MCO en prés. d'hétéros. $\left(var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}\right.$

...

- Hypothèses relatives à la forme structurelle de l'hétéroscédasticité

$Y_i = \beta_1 + \beta_2 X_i + u_i$

$\boxed{\begin{array}{l}\text{Hypothèse 1 : La variance de l'erreur est proportionnelle à } X_i^2 \\ \Longrightarrow \quad E(u_i^2) = \sigma^2 X_i^2\end{array}}$



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

$$= \beta_1 \left(\frac{1}{X_i}\right) + \beta_2 + v_i$$

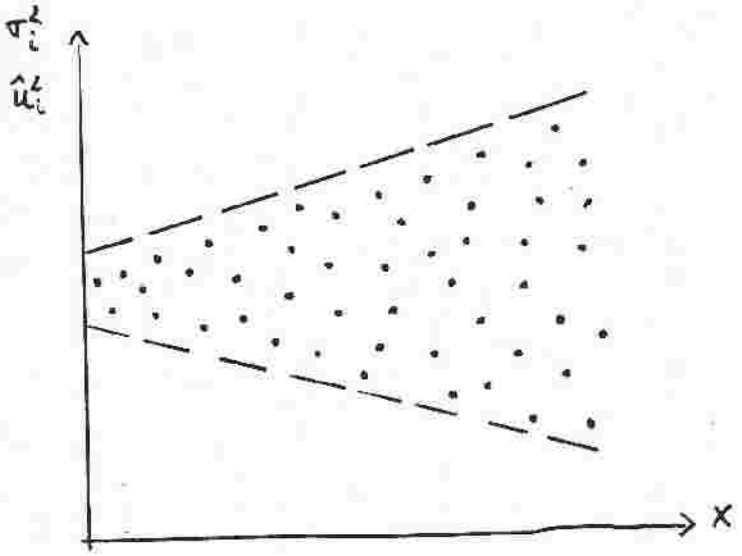
où $v_i = \dfrac{u_i}{X_i}$

$$E(v_i^2) = E\left(\frac{u_i}{X_i}\right)^2 = \frac{1}{X_i^2} E(u_i^2) = \sigma^2$$

car $E(u_i^2) = \sigma^2 X_i^2$

v_i est homoscédastique, on peut appliquer MCO pour estimer modèle de rég. transformé.

$\boxed{\text{Hypothèse 2 : La variance de l'erreur est proportionnelle à } X_i \quad \Rightarrow \quad E(u_i^2) = \sigma^2 X_i}$



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

$$= \beta_1 \frac{1}{\sqrt{X_i}} + \beta_2 \sqrt{X_i} + v_i$$

où $\quad v_i = \dfrac{u_i}{\sqrt{X_i}} \qquad$ avec $X_i > 0$

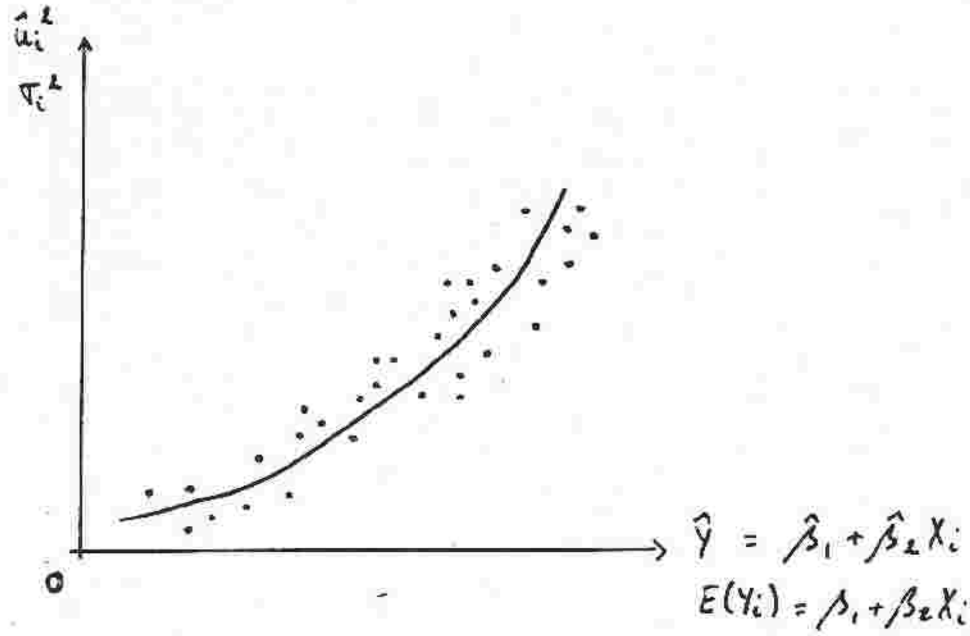
$$E(v_i^2) = E\left(\frac{u_i}{\sqrt{X_i}}\right)^2 = \frac{1}{X_i} E(u_i^2) = \sigma^2$$

car $E(u_i^2) = \sigma^2 . X_i$

v_i est homoscédastique, MCO peuvent être utilisés pour estimer modèle de rég. transformé.

Hypothèse 3 : La variance de l'erreur est proportionnelle au carré de la valeur moyenne (conditionnelle) de Y

$$\Rightarrow E(u_i^2) = \sigma^2 \left(E(Y_i)\right)^2$$



$$\frac{Y_i}{E(Y_i)} = \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)}$$

$$= \beta_1 \left[\frac{1}{E(Y_i)}\right] + \beta_2 \frac{X_i}{E(Y_i)} + v_i$$

où $v_i = \dfrac{u_i}{E(Y_i)}$

$$E(v_i^2) = E\left(\frac{u_i}{[E(Y_i)]}\right)^2 = \frac{1}{[E(Y_i)]^2} \cdot E(u_i^2) = \sigma^2$$

car $E(u_i^2) = \sigma^2 [E(Y_i)]^2$

v_i est homoscédastique

Ne peut être appliqué car $E(Y_i) = \beta_1 + \beta_2 X_i$ et (β_1, β_2) sont inconnus.

$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \equiv$ estimateur de $E(Y_i)$

- Trois étapes :

i) Estimer modèle de régression initial $\Rightarrow \hat{Y}_i$

ii) Transformation :

$$\frac{Y_i}{\hat{Y}_i} = \beta_1 \left(\frac{1}{\hat{Y}_i}\right) + \beta_2 \left(\frac{X_i}{\hat{Y}_i}\right) + v_i \qquad (1)$$

où $v_i = \left(\frac{u_i}{\hat{Y}_i}\right)$

iii) Estimation du modèle transformé (1)

Procédure asymptotique.

Hypothèse 4 : Lorsqu'on a un modèle de rég. lin. $Y_i = \beta_1 + \beta_2 X_i + u_i$ très souvent on réduit l'hétéroscédasticité en y appliquant une transf. log., càd. en estimant $\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$

Exemple : $80 = 10 \times$ + gd que 8

$\ln(80) = \pm\, 2 \times$ + gd que $\ln(8)$

Remarques :

Nature de la variance cond. des erreurs σ_i^2 est svt incertaine
$\longrightarrow$ svt difficile de déterminer quelle transf. est adéquate
$\longrightarrow$ si on dispose d'un gd échantillon, svt + facile d'utiliser procédure de White m̂ si estimateurs pas tjs efficients.

Transformations posent certains pblmes :

1) Quelle var. expl. utiliser pour transformer les données ?
2) Transf. log. ne peut pas être utilisée lorsque var. prennent valeurs nulles ou négatives.
3) Corrélation fallacieuse („ Spurious correlation")
Variables initiales non corrélées mais ratio de ces var. corrélés
Exemple : $Y_i = \beta_1 + \beta_2 X_i + u_i \Rightarrow Y_i$ et X_i non corrélés
$Y_i/X_i = \beta_1 (1/X_i) + \beta_2 + (u_i/X_i) \Rightarrow \left(\frac{Y_i}{X_i}\right)$ et $\left(\frac{1}{X_i}\right)$ corrélés
4) Inférence uniq. permise asympt.

- Exemple :

TABLE 11.5
INNOVATION IN AMERICA: RESEARCH AND DEVELOPMENT (R&D) EXPENDITURE IN THE UNITED STATES, 1988 (All Figures in Millions of Dollars)

Industry grouping	Sales	R&D expenses	Profits
1. Containers and packaging	6,375.3	62.5	185.1
2. Nonbank financial	11,626.4	92.9	1,569.5
3. Service industries	14,655.1	178.3	276.8
4. Metals and mining	21,869.2	258.4	2,828.1
5. Housing and construction	26,408.3	494.7	225.9
6. General manufacturing	32,405.6	1,083.0	3,751.9
7. Leisure time industries	35,107.7	1,620.6	2,884.1
8. Paper and forest products	40,295.4	421.7	4,645.7
9. Food	70,761.6	509.2	5,036.4
10. Health care	80,552.8	6,620.1	13,869.9
11. Aerospace	95,294.0	3,918.6	4,487.8
12. Consumer products	101,314.1	1,595.3	10,278.9
13. Electrical and electronics	116,141.3	6,107.5	8,787.3
14. Chemicals	122,315.7	4,454.1	16,438.8
15. Conglomerates	141,649.9	3,163.8	9,761.4
16. Office equipment and computers	175,025.8	13,210.7	19,774.5
17. Fuel	230,614.5	1,703.8	22,626.6
18. Automotive	293,543.0	9,528.2	18,415.4

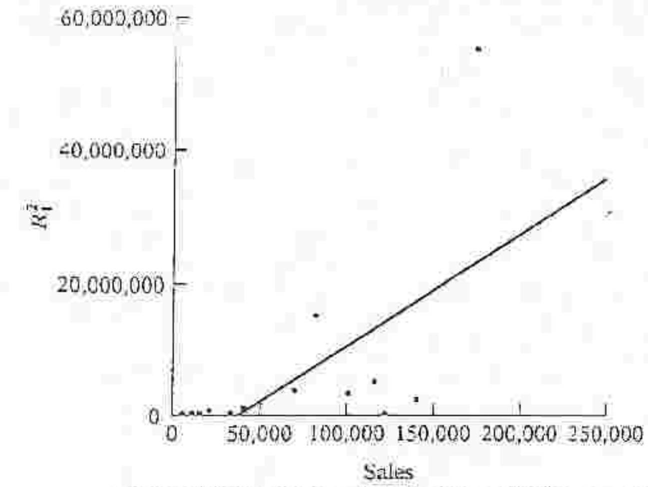
Source: Business Week, Special 1989 Bonus Issue, R&D Scorecard, pp. 180–224.
Note: The industries are listed in increasing order of sales volume.

Estimation par MCO :

$$\widehat{(R\&D)}_i = 192,9931 + 0,0319 \text{ Sales}_i \quad (1) \quad \left(\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_i^2}\right)$$

$$se = (533,9317) \quad (0,0083)$$

2.41. $$t = (0,3614) \quad (3,8433) \qquad R^2 = 0,4783$$



Squared residuals from the regression of R&D on sales

FIGURE 11.13

Test de Park

$$\hat{u}_i^2 = -974469,1 + 86,2321 \, Sales_i$$
$$se = (4802,343) \qquad (40,3625)$$
$$t = (-0,2029) \qquad (2,1364)$$

$$R^2 = 0,2219$$

Test de White

$$\hat{u}_i^2 = -6219665 + 229,3508 \, Sales_i - 0,000537 \, Sales_i^2$$
$$se = (6459809) \qquad (126,2197) \qquad (0,0004)$$
$$t = (0,9628) \qquad (1,8170) \qquad (-1,3425)$$
$$R^2 = 0,2895$$

$$n = 18 \implies n.R^2 = 5,2124 \sim \chi^2_{2df} \text{ (sous } H_0)$$

$$p\text{-value} = 0,074$$

$\Rightarrow$ Analyse graphique, test de Park et test de White suggèrent presence d'hétéroscédasticité.

σ_i^2 n'est pas connue $\longrightarrow$ impossible d'utiliser les MCP

- Analyse graphique suggère que : $E(u_i^2) = \sigma^2 . Sales_i$

$\longrightarrow$ <u>diviser ensemble des variables</u> <u>par $\sqrt{Sales_i}$ et estimation</u> <u>par MCO</u> :

$$\widehat{\left(\frac{R\&D_i}{\sqrt{Sales_i}}\right)} = -246,6769 \frac{1}{\sqrt{Sales_i}} + 0,0367 \sqrt{Sales_i} \qquad (2)$$

$$se = (381,1285) \qquad (0,0071)$$

$$t = (-0,6472) \qquad (5,1690)$$

$R^2 = 0,3648$

Rappel : si application des MCO aux données <u>non</u> transf. :

$$\widehat{(R\&D_i)} = 192,9931 + 0,0319 \; Sales_i$$

$$se = (533,9317) \quad (0,0083)$$

$$t = (0,3614) \quad (3,8433)$$

- <u>White heteroscedasticity - consistent standard errors</u> :

$$\widehat{(R\&D_i)} = 192,9931 + 0,0319 \; Sales_i$$

$$se = (533,9931) \quad (0,0101)$$

$$t = (0,3614) \quad (3,1584) \qquad R^2 = 0,4783$$

- En résumé :

1) En présence d'hétéroscédasticité, estimateurs des MCO sont non biaisés et consistants mais leur variance n'est plus minimale → ils ne sont pas efficients. (ne sont pas BLUE).

2) Lorsque σ_i^2 est connue, MCP fournit des estimateurs BLUE.

3) En présence d'hétéroscédasticité, les variances des estimateurs des MCO pas fournies par formules habit.
→ si on persiste à utiliser formules habit., tests en t, F, etc... peuvent conduire à ccl. erronées.

4) Plus facile de décrire csquences de l'hét. que de la détecter. ∃ ≠ test mais aucun n'est infaillible.

5) Remèdes lorsque σ_i^2 inconnue :
- si n est grand : erreurs standard de White
→ permet inférence (m̂ si estimateurs pas tjs efficients).
- si n petit : utiliser résidus des MCO, diagn. forme fctionnelle terme d'erreur hétérosc.
→ transformat° des données, estimat° par MCO.

2.3. L'autocorrélation

- <u>Cross-section</u> porte sur un échantillon aléatoire d'unités (par exemple ménages, firmes).

  A priori, pas de raison que termes d'erreur soient corrélés. Le cas échéant, on parle d'autocorrélation spatiale.

  Généralmt pas d'ordre logique dans les unités d'un cross-section ⟶ il suffit svt de changer ordre des unités pour que pblme d'autocorrélat° soit résolu.

- <u>Time series</u> : observations ordonnées de façon logique dans le tmps.

  Forte probabilité que les observ. successives soient corrélées surtt lorsque l'intervalle de tmps est court.

⟹ Quelle est la nature de l'autocorrélation ?
Quelles sont les csquences théoriques et pratiques ?
Comment détecter l'autocorrélation ?
Comment remédier à l'autocorrélation ?

2.3.1. La nature du problème

Autocorrélation = corrélation entre une série d'observations qui sont ordonnées dans le temps (time series) ou dans l'espace (cross-section).

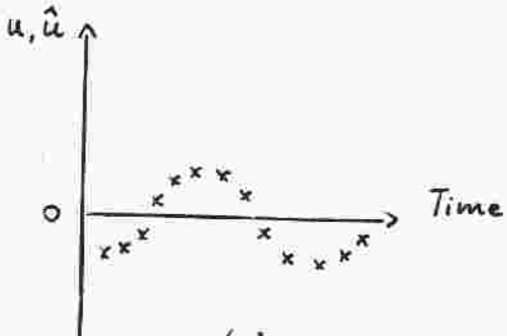
Absence d'autocorrélation des erreurs :

$\boxed{E(u_i u_j) = 0} \qquad i \neq j$

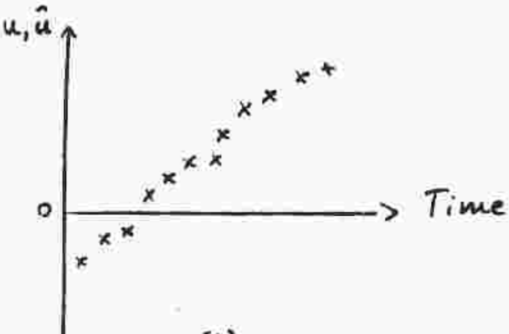
$\Rightarrow$ terme d'erreur d'une obs. n'est pas influencé par terme d'erreur d'une autre observation.

Causes de l'autocorrélation ?

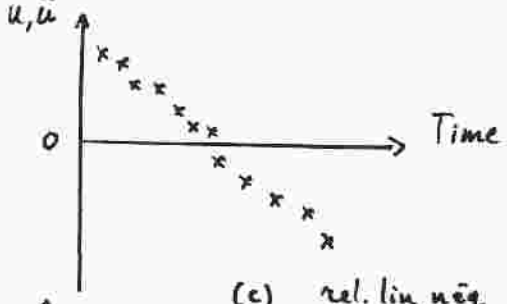
1) L'inertie
2) $\exists$ d'un biais de spécification suite à l'exclusion de variables explicatives importantes.



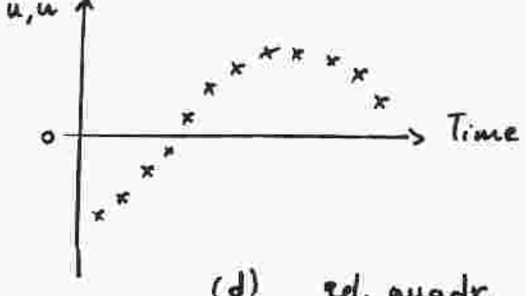
(a) rel. cub.



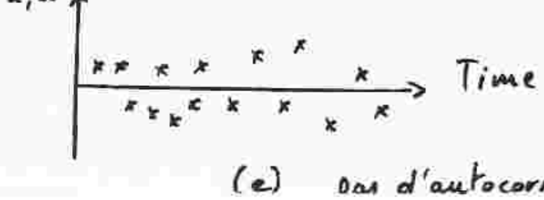
(b) rel. lin pos.



(c) rel. lin nég.



(d) rel. quadr.



(e) pas d'autocorr.

Exemple :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad (1)$$

où

$$\begin{cases} Y = \text{qté de viande de boeuf demandée} \\ X_2 = \text{prix du boeuf} \\ X_3 = \text{revenu des consommateurs} \\ X_4 = \text{prix du porc} \\ t = \text{temps} \\ u = \text{terme d'erreur} \end{cases}$$

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \qquad (2)$$

Si modèle (1) est correct $\Rightarrow v_t = \beta_4 X_{4t} + u_t$

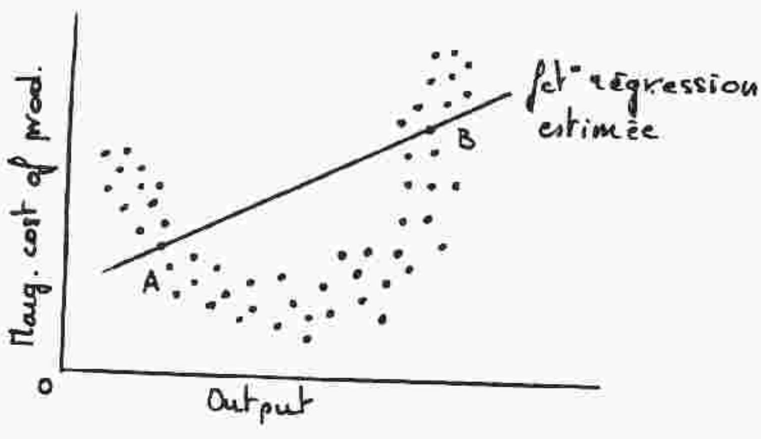
v_t seront autocorrélés.

3) ∃ d'un biais de spécification suite à l'utilisation d'une forme fonctionnelle inadéquate.

Exemple :

$$\text{Marginal cost}_i = \beta_1 + \beta_2 \text{ Output}_i + \beta_3 \text{ Output}_i^2 + u_i \qquad (1)$$

$$\text{Marginal cost}_i = \alpha_1 + \alpha_2 \text{ Output}_i + v_i \qquad (2)$$



$\begin{cases} (1) \equiv \text{vrai modèle} \\ (2) \equiv \text{modèle estimé} \end{cases}$

$v_i = \text{output}_i^2 + u_i \Rightarrow v_t$ est autocorrélé.

4) Omission d'une variable dépendante retardée (laguée).

Exemple :

$$\text{Consommation}_t = \beta_1 + \beta_2 \,\text{Income}_t + \beta_3 \,\text{Cons}_{t-1} + u_t$$

$\Rightarrow$ modèle autorégressif.

5) „Manipulation" des données.

(lissage, extra & intrapolations)

6) „Transformation" des données

Exemple :

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (1)$$

où $\begin{cases} Y = \text{dépenses de cons.} \\ X = \text{revenu} \end{cases}$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad (2)$$

$(1)-(2)$: $\Delta Y_t = \beta_2 \Delta X_t + v_t$ $\equiv$ mod. en $\neq$ 1ères.

où $\begin{cases} \Delta = \text{opérateur différence première} \\ \Delta Y_t = Y_t - Y_{t-1} \\ \Delta X_t = X_t - X_{t-1} \\ v_t = \Delta u_t = u_t - u_{t-1} \end{cases}$

Même si u_t n'est pas autocorrélé, v_t le sera.

Démonstration :

a) $v_t = u_t - u_{t-1}$

$\Rightarrow E(v_t) = E(u_t - u_{t-1}) = E(u_t) - E(u_{t-1}) = 0$

car par hyp. $E(u_t) = 0, \forall t$. $\Rightarrow$ esp. cond. v_t est nulle

b) $var(v_t) = var(u_t - u_{t-1}) = var(u_t) + var(u_{t-1}) = 2\sigma^2$

Pq ?

- par hyp. : $\begin{cases} var(u_t) = \sigma^2, & \forall t \\ E(u_t u_s) = 0, & t \neq s \end{cases} \Rightarrow cov(u_t, u_s) = 0$

- par déf. : $var(a-b) = var(a) + var(b) - 2cov(a,b)$

$\Rightarrow v_t$ est homoscédastique

c) $cov(v_t, v_{t-1}) = E(v_t v_{t-1}) = E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})]$

$= E(u_t u_{t-1}) - E(u_t u_{t-2}) - E(u_{t-1}^2) + E(u_{t-1} u_{t-2})$

$= 0 - 0 - \sigma^2 - 0$

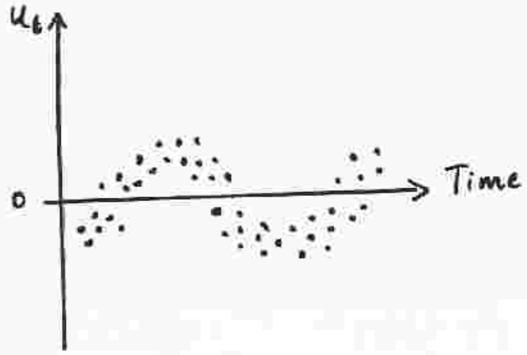
$= -\sigma^2$

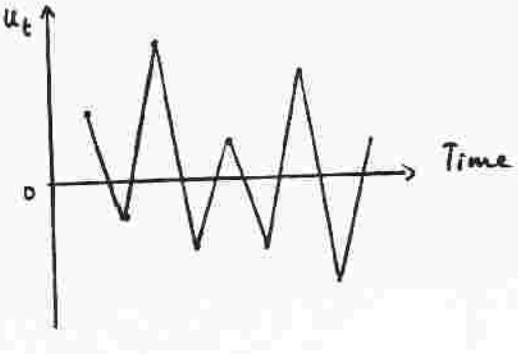
Pq ? Par hyp. u_t ne sont pas autocorrélés $\rightarrow E(u_t u_s) = 0$ $\forall t \neq s$

$\Rightarrow$ m̂ si u_t ne sont pas autocorrélés, leurs ≠ 1ères v_t le seront.

7) La non stationnarité des variables

- Remarque : Autocorrélation peut être ⊕ comme ⊖ (mais + svt ⊕)





2.3.2. Estimation par MCO en présence d'autocorrélation

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (1)$$

$$\boxed{u_t = \rho u_{t-1} + \varepsilon_t} \qquad (2)$$

où $\begin{cases} \rho \equiv \text{coefficient d'autocovariance} \\ \varepsilon_t \equiv \text{terme d'erreur stochastique tq :} \\ \quad \begin{cases} E(\varepsilon_t) = 0 \\ var(\varepsilon_t) = \sigma_\varepsilon^2 \\ cov(\varepsilon_t, \varepsilon_{t+s}) = 0 \quad , s \neq 0 \end{cases} \\ \quad \Rightarrow \text{„ bruit blanc "} \\ \quad \text{(white noise error term)} \end{cases}$

Forme structurelle du terme d'erreur u_t est un processus autorégressif d'ordre 1 $\equiv AR(1)$

__Rem.__ : $AR(2) \equiv u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$

__Propriétés__ : ρ s'interprète comme coeff. de corrélation d'ordre 1

$$\Rightarrow \rho = \frac{E\{[u_t - E(u_t)][u_{t-1} - E(u_{t-1})]\}}{\sqrt{var(u_t)}\ \sqrt{var(u_{t-1})}}$$

$$= \frac{E(u_t u_{t-1})}{var(u_{t-1})} \qquad \text{car} \begin{cases} E(u_t) = 0 \quad \forall t \\ var(u_t) = var(u_{t-1}) \text{ car} \\ \text{hyp. d'homosc. maintenue} \end{cases}$$

ρ = coeff. de pente de régression de u_t sur u_{t-1}

- Quid de la variance, de la covariance et de la corrélation de u_t ?

$$\boxed{\begin{aligned} var(u_t) &= E(u_t^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2} && (1) \\ cov(u_t, u_{t+s}) &= E(u_t u_{t+s}) = \rho^s \frac{\sigma_\varepsilon^2}{1-\rho^2} && (2) \\ cor(u_t, u_{t+s}) &= \rho^s && (3) \end{aligned}}$$

$\Rightarrow$ i) $\rho \in [-1, 1]$, variance u_t tjs homoscédastique.

ii) u_t est corrélé avec sa valeur à +ieurs lags d'écart.

iii) $|\rho| = 1 \longrightarrow var(u_t)$ & $cov(u_t, u_{t+s})$ indéterminés

$|\rho| < 1 \longrightarrow$ processus AR (1) est stationnaire (moyenne, var. et cov. sont définies et leurs valeurs ne chgent pas ds le temps).

- soit : $Y_t = \beta_1 + \beta_2 X_t + u_t$

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2}$$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2} \quad (1), \quad var(\hat{\beta}_2)_{AR(1)} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho \frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2 \frac{\sum x_t x_{t-2}}{\sum x_t^2} + \dots + 2\rho^{n-1} \frac{x_1 x_n}{\sum x_t^2}\right] \quad (2)$$

Difficile de prédire si $var(\hat{\beta}_2)$ est $>, <$ ou $=$ $var(\hat{\beta}_2)_{AR(1)}$

Si $\rho = 0 \Rightarrow var(\hat{\beta}_2) = var(\hat{\beta}_2)_{AR(1)}$

Si corrélations ld valeurs de X faible $\Rightarrow \left\{\begin{array}{l}\text{variance} \\ \text{estimateurs MCO} \\ \text{peu biaisée}\end{array}\right.$

- <u>Hypothèse</u> : X suit un processus $AR(1)$

$$\Rightarrow \text{cor}(X_t, X_{t+s}) = r^s$$

$$\text{où} \quad r = \frac{\sum x_t x_{t+1}}{\sum x_t^2}$$

On peut montrer que :

$$\text{var}(\hat{\beta}_2)_{AR(1)} = \frac{\sigma^2}{\sum x_t^2}\left(\frac{1+r\rho}{1-r\rho}\right) = \text{var}(\hat{\beta}_2)_{OLS}\left(\frac{1+r\rho}{1-r\rho}\right)$$

Exemple : $r = 0{,}6$ et $\rho = 0{,}8$

$$\Rightarrow \text{var}(\hat{\beta}_2)_{AR(1)} = 2{,}8461 \ \text{var}(\hat{\beta}_2)_{OLS}$$

$$\Rightarrow \text{var}(\hat{\beta}_2)_{OLS} = 0{,}3513 \ \text{var}(\hat{\beta}_2)_{OLS}$$

Formule usuelle de la variance de $\hat{\beta}_2$ sous-estime de +/- 65% la variance de $\hat{\beta}_2$ lorsque terme d'erreur suit un processus $AR(1)$.

- Quid si on utilise estimateur des MCO $\hat{\beta}_2$ et qu'on ajuste formule usuelle de la variance pour tenir compte de la structure $AR(1)$ du terme d'erreur ?

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2}$$

$$\text{var}(\hat{\beta}_2)_{AR(1)} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \dots + 2\rho^{n-1}\frac{x_1 x_n}{\sum x_t^2}\right]$$

$\hat{\beta}_2$ est linéaire et non biaisé mais variance n'est pas minimale $\Rightarrow$ estimateur n'est pas efficient.

2.3.3. Estimateur BLUE en présence d'autocorrélation

Technique des MCG fournit des estimateurs BLUE.
Pour modèle de rég. à 2 variables, sous hyp. que terme d'erreur suit processus AR(1), estimateur BLUE de β_2 est donné par :

$$\hat{\beta}_2^{GLS} = \frac{\sum_{t=2}^{n} (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n} (x_t - \rho x_{t-1})^2} + C$$

$$\text{var}(\hat{\beta}_2^{GLS}) = \frac{\sigma^2}{\sum_{t=2}^{n} (x_t - \rho x_{t-1})^2} + D$$

où C et D sont des termes correcteurs

Estimateur des MCG est BLUE parce qu'il utilise au mieux l'information disponible.

2.3.4. Conséquences de l'utilisation des MCO en présence d'autocorrélation

Estimateurs des MCO sont tjs linéaires, non biaisés, consistants et asympt. normaux mais pas efficients.

Quid des procédures habituelles d'inférence statistique ?

(i) Lorsqu'on tient compte de l'autocorrélation.
(ii) Lorsqu'on ignore la présence d'autocorrélation.

(i) <u>Estimation par MCO en tenant compte de l'auto-corrélation</u>

Estimateur des MCO $\hat{\beta}_2$ n'est pas BLUE. Même si on utilise $var(\hat{\beta}_2)_{AR(1)}$ intervalles de confiance + larges que si on utilise $var(\hat{\beta}_2^{GLS})$

$\Rightarrow$ $\hat{\beta}_2$ n'est pas efficient même asymptotiqumt.

$\Rightarrow$ pour faire inférence utiliser MCG.

(ii) <u>Estimation par MCO en ignorant autocorrélation</u>

Quid si on utilise (en présence d'autocorrélation) :

$$\hat{\beta}_2 = \frac{\Sigma x_t y_t}{\Sigma x_t^2} \quad , \quad var(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_t^2} \quad ?$$

- variance résiduelle $\hat{\sigma}^2 = \Sigma \hat{u}_t^2 / (n-2)$ sous-estime généralmt la vraie valeur de σ^2 $\Rightarrow$ on sur-estime génerlmt R^2

- même si valeur de σ^2 n'est pas sous-estimée, $var(\hat{\beta}_2)$ peut sous-estimer $var(\hat{\beta}_2)_{AR(1)}$

  (! $var(\hat{\beta}_2)_{AR(1)}$ inefficiente // $var(\hat{\beta}_2^{GLS})$)

$\Rightarrow$ tests usuels de significativité ne sont plus valables.

- Sous hyp. modèle rég. lin. classique :

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{(n-2)}$$ est un estimateur non biaisé de σ^2

$$\Rightarrow E(\hat{\sigma}^2) = \sigma^2$$

Si terme d'erreur suit processus AR(1) :

$$E(\hat{\sigma}^2) = \sigma^2 \frac{\{n - [2/(1-\rho)] - 2\rho r\}}{n-2}$$

où $\begin{cases} n = \# \text{ d'observations} \\ \rho = \text{coeff. autocorr. terme d'erreur} \\ r = \text{coeff. de corrélation les valeurs successives de } X \longrightarrow r = \sum_{t=1}^{n-1} x_t x_{t-1} \Big/ \sum_{t=1}^{n} x_t^2 \end{cases}$

Si ρ et r positifs (hyp. rais. pour plupart séries Eco.)
$\Rightarrow E(\hat{\sigma}^2) < \sigma^2$

Formule habituelle de la variance des résidus sous-estime en moyenne la vraie valeur de σ^2.

Biais dans $\hat{\sigma}^2$ se transmet dans var $(\hat{\beta}_2)$
car var $(\hat{\beta}_2) = \hat{\sigma}^2 / \sum x_t^2$

- Même si σ^2 n'est pas sous-estimé, on peut montrer que var $(\hat{\beta}_2)$ est un estimateur biaisé de var $(\hat{\beta}_2)_{AR(1)}$.

Si ρ et r positifs : var $(\hat{\beta}_2) <$ var $(\hat{\beta}_2)_{AR(1)}$

$\Rightarrow$ si on utilise var $(\hat{\beta}_2)$, on sur-estime précision de $\hat{\beta}_2$ (on sur-estime t), on rejette trop vite H_0.

2.3.5. Relation entre les salaires et la productivité dans secteur privé aux Etats-Unis entre 1959 et 1988

- Estimation par MCO :

$$\hat{Y}_t = 29,5192 + 0,7136 \, X_t$$
$$se = (1,9423) \qquad (0,0241)$$
$$t = (15,1977) \qquad (29,6066)$$

$R^2 = 0,9584$

$d = 0,1229$

$\hat{\sigma} = 2,6755$

où

$$\begin{cases} Y = \text{indice du salaire horaire} \\ X = \text{indice de l'output par heure} \end{cases}$$

Fiabilité des résultats ?
En présence d'autocorrélation, se de estimateurs sont biaisés (t stat pas fiables)

2.3.6. Détecter l'autocorrélation

(i) La méthode graphique

Résidus des MCO fournissent une approximation des erreur

⟶ représentation graphique des résidus // temps.

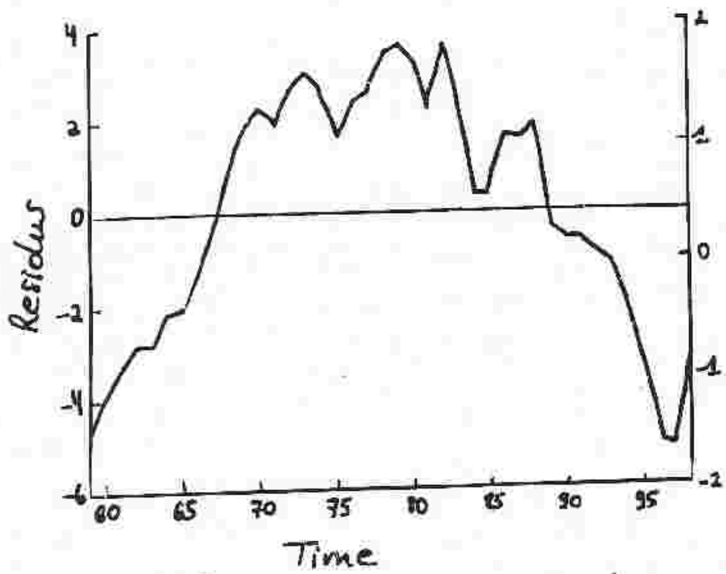


Fig. Residuals from wages – productivity regression

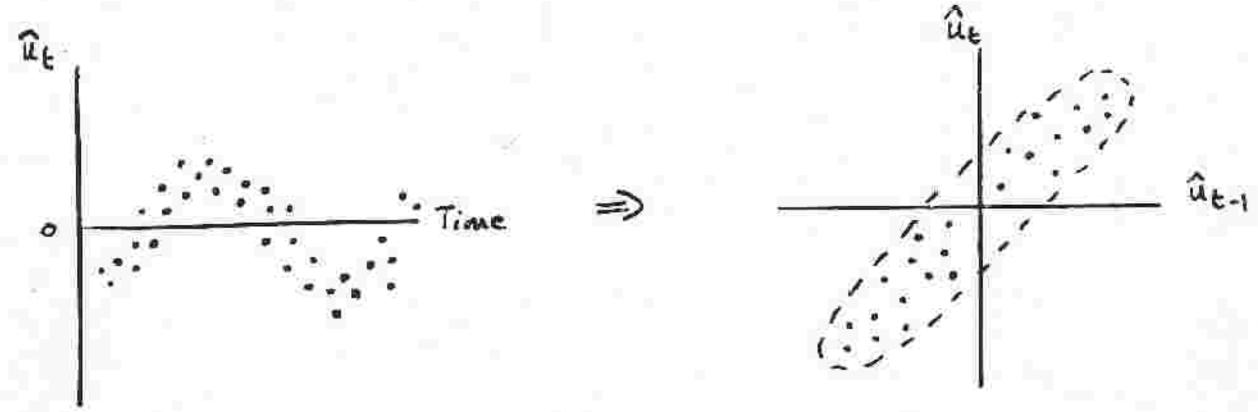


Fig. Positive autocorrelation

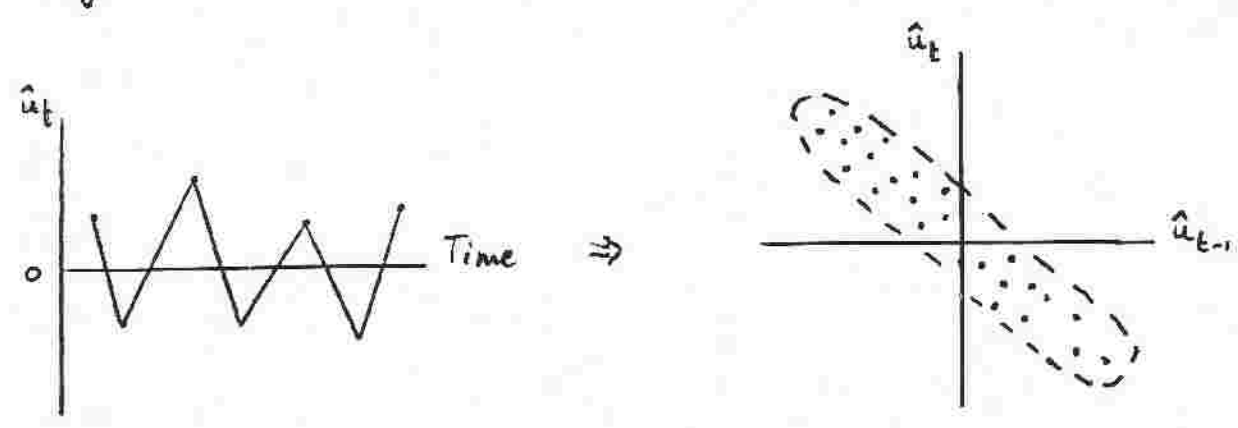


Fig. Negative autocorrelation

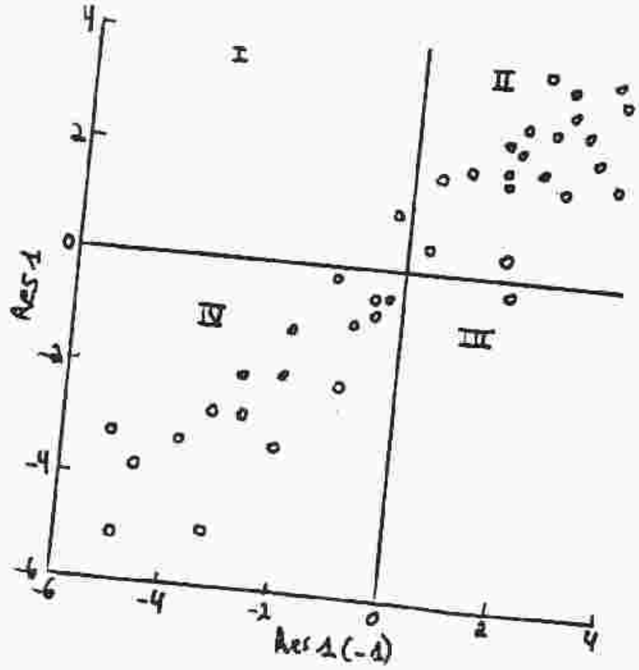


Fig. Current residuals versus lagged residuals

Analyse graphique est subjective et qualitative.

(ii) <u>Le test des séquences</u> („runs test")

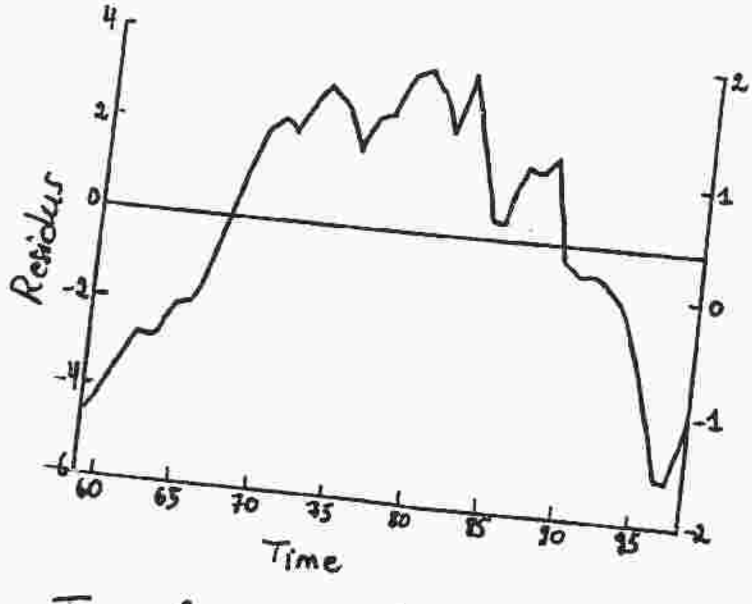


Fig. Residuals from wage - productivity regression.

Test des séquences ou „Geary - test" est non paramétrique

- Série de signes :

(-----------)(++++++++++++++++++++)(-----------)

Séquence = série ininterrompue de signes positifs ou négatifs
Longueur d'une séquence = nbre de symboles qu'elle contie

- Autocorrélation ? Obtenir 3 séquences pour 40 observat°, est-ce trop élevé / faible au regard de ce qu'on aurait obtenu pour 40 observat° pftmt aléatoires ?

Intuition :
- si # de séq. trop élevé ⇒ résidus chgt trop fréq. de signe ⇒ autocorrélation négative.
- si # de séq. trop faible ⇒ résidus ne chgt pas suffismt de signe ⇒ autocorrélation positive.

- Soit : N = nbre total d'observations $= N_1 + N_2$
  N_1 = nbre de symboles „+"
  N_2 = nbre de symboles „−"
  R = nbre de séquences

⇒ Sous hyp. nulle que résidus sont indépendants et lorsque $N_1 > 10$ et $N_2 > 10$, nbre de séquences suit asymph. une distrib. normale tq. Moyenne : $E(R) = \dfrac{2N_1N_2}{N} + 1$

Variance : $\sigma_R^2 = \dfrac{2N_1N_2\,(2N_1N_2 - N)}{(N)^2\,(N-1)}$

⇒ Sous hyp. nulle d'indépendance, d'après prop. distr. normale : $Prob\left[-1{,}96 \leq \dfrac{R-E(R)}{\sigma_R} \leq +1{,}96\right] = 0{,}9$

$$\Rightarrow \boxed{\text{Prob}\left[E(R) - 1{,}96 . \sigma_R \leq R \leq E(R) + 1{,}96\ \sigma_R\right] = 0{,}95}$$

<u>Règle de décision</u> :

Rejette hyp. nulle d'absence d'autocorrélation au niveau de prob. 95% lorsque nbre de séquences R n'est pas contenu ds l'intervalle de confiance. Sinon on ne rejette pas H_0.

<u>Exemple</u> :

$$\left.\begin{array}{l} N_1 = 19 \\ N_2 = 21 \\ R = 3 \end{array}\right\} \Rightarrow \quad \begin{array}{l} E(R) = \dfrac{2 N_1 N_2}{N} + 1 = 10{,}975 \\[2ex] \sigma_R^2 = \dfrac{2 N_1 N_2 \, (2 N_1 N_2 - N)}{N^2 \ (N-1)} = 9{,}6936 \end{array}$$

$$\Rightarrow \sigma_R = 3{,}1134$$

$$IC : [10{,}975 +/- 1{,}96\ (3{,}1134)] = [4{,}8728\ ;\ 17{,}0722]$$

On rejette H_0 au niveau de prob. 95%.
Nbre de séq. trop faible $\Rightarrow$ autocorrélation positive.

<u>Quid si $N < 20$?</u>

Tables de Swed & Eisenhart

(iii) <u>Le test de Durbin-Watson</u> (d test)

$$\boxed{d = \dfrac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n} \hat{u}_t^2}}$$

Hypothèses à satisfaire pour utiliser la stat. "d"

1. Le modèle de régression contient un intercepte
   Sinon, réestimer modèle avec constante ou utiliser les tables stat. de Farebrother (1980).
2. Les variables explicatives X doivent être non stochastiques.
3. Les erreurs u_t doivent être générées par processus AR(1) : $u_t = \rho u_{t-1} + \varepsilon_t$
   DW pas en mesure de détecter processus autorégressif dont ordre > 1.
4. Terme d'erreur suit une distribution normale.
5. Le Modèle de régression ne contient pas de variable dépendante laguée (parmi var. expl.)
   DW ne peut pas être utilisé si on estime modèle autoregressif du type :
   $$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \lambda Y_{t-1} + u_t$$
6. Pas de valeurs manquantes parmi les observations.

Distribution d'échantillonnage exacte de la stat. "d" ?

Difficile à dériver car elle dépend de manière complexe des valeurs prises par les variables expl. X
⇒ pas de valeur critique unique permettant de rejeter ou de ne pas rejeter H₀ que perturbations u_i suivent processus AR(1)
⇒ limites supérieure (d_U) et inférieure (d_L) dont valeur dépend uniq. du nbre d'observations (n) et du nbre de variables explicatives (et pas de la valeur des variables explicatives).

- Valeur de la stat. « d" comprise lel 0 et 4

$$d = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n} \hat{u}_t^2} \implies d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{n} \hat{u}_t^2}$$

$\sum \hat{u}_t^2$ & $\sum \hat{u}_{t-1}^2$ diffèrent d'une seule observation $\Rightarrow$ deux sommes sont approx. égales

$\Rightarrow$

$$d \simeq 2\left[1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}\right]$$

Comme, on sait que :

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$ (coeff. d'autocorr. d'ordre 1 des résidus)

$$\Rightarrow \quad \boxed{d \simeq 2(1 - \hat{\rho})}$$

Comme, on sait que :

$$-1 \leq \rho \leq 1$$

$$\Rightarrow \quad \boxed{0 \leq d \leq 4}$$

Lorsque :

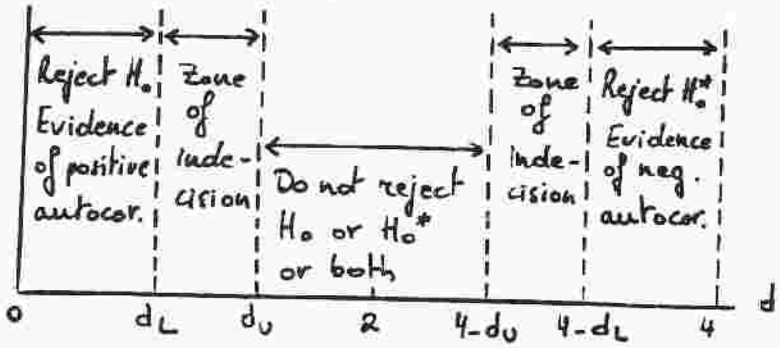
$\hat{\rho} = 0 \implies d \approx 2$ (absence d'autoc.

$\hat{\rho} = 1 \implies d \approx 0$ (autoc. positive)

$\hat{\rho} = -1 \implies d \simeq 4$ (autoc. négative)

En pratique :

1. On estime le modèle par les MCO et on obtient les résidus
2. On calcule la statistique „d".
3. On trouve les valeurs critiques d_L et d_U étant donné la taille de notre échantillon et le nbre de variables expl. incluses de le modèle.
4. On applique, règle de décision suivante :



Legend : H_0 : No positive autocorrelation
H_0^* : No negative autocorrelation

Fig. Durbin-Watson d statistic

a) $0 \leq d < d_L$, rejette H_0 d'absence d'autocor. positive d'ordre 1.
b) $d_L \leq d < d_U$, zone d'indétermination (pas de ccl.)
c) $d_U \leq d < 4-d_U$, on ne rejette pas H_0 d'absence d'aut. d'ord.
d) $4-d_U \leq d \leq 4-d_L$, zone d'indétermination (pas de ccl.)
e) $4-d_L < d \leq 4$, rejette H_0 d'absence d'autocor. nég. d'ordre 1

Exemple : $d = 0,1229$, 40 obs, 1 var. explicative
$d_L = 1,44$ et $d_U = 1,54$ $(\alpha = 0,05)$

$d < d_L \Rightarrow$ rejette hyp. d'absence d'autocor. positive les résidus

- „Principal" inconvénient du test de DW : $\exists$ une zone d'indétermination.

→ propositions pour contourner ce pblme : idée générale = dans de nbrx cas, d_U = valeur critique.

→ lorsque d comprise ds zone d'indétermination, suivre règle de décision suivante :

1) si $H_0 : \rho = 0$ vs. $H_1 : \rho > 1 \Rightarrow RH_0(\alpha)$ si $d < d_U$
2) si $H_0 : \rho = 0$ vs. $H_1 : \rho < 0 \Rightarrow RH_0(\alpha)$ si $(4-d) < d_U$
3) si $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0 \Rightarrow RH_0(2\alpha)$ si $\begin{cases} d < d_U \\ \text{ou} \\ (4-d) < d_U \end{cases}$

- Quid si on utilise test de DW pour modèle autorégressif (variable dpte laguée parmi var. explic.) ?

d sera généralmt proche de 2 m̂ en présence d'autocorrélation d'ordre 1. Solution : „h test" de DW mais moins puissant que test de Breusch - Godfrey → mieux vaut utiliser BG.

- Quid si terme d'erreur ne suit pas dist. normale ?

Si échantillon petit ⇒ stat d ne peut pas être utilisée.
Si échantillon gd ⇒ stat d peut être utilisée
Pq ?

$$\sqrt{n}\left(1 - \tfrac{1}{2} d\right) \underset{asy}{\approx} N(0,1)$$

Exemple : $d = 0{,}1229$, $n = 40$

$\sqrt{n}(1 - 1/2\, d) = \sqrt{40}\left(1 - \frac{0{,}1229}{2}\right) = 5{,}94$ → RH_0 d'absence d'autocor

Valeurs critiques à 5 et 1 % : 1,96 et 2,58

- Quid si variables explicatives stochastiques ?

Test de DW n'est pas valable m̂ lorsque l'échantillon est grand.

Avec séries chronologiques sut l'une des variables explicatives est la variable dpdt laguée → 1 régresseur stochastique, DW non utilisable !

(iv) Le test de Breusch - Godfrey

Aussi appelé LM - test (basé sur le principe du multiplicateur de Lagrange)

Test „ général " d'autocorrélation car permet :

* que les régresseurs soient stochastiques.
* de testes pour structures autorégressives d'ordre supérieur à 1.
* de vérifier si terme d'erreur peut être représenté comme une moyenne mobile d'un bruit blanc ε_t

↳ Exemple : $Y_t = \beta_1 + \beta_2 X_t + u_t$

? où $u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2}$

(terme d'erreur u_t = moyenne mobile d'ordre 3 du bruit blanc ε_t)

Fonctionnement du test de Breusch - Godfrey ?

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

où $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$

(processus $AR(p)$)

$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$

1) On estime le modèle de régression : $Y_t = \beta_1 + \beta_2 X_t + u_t$ et on obtient les résidus $\hat{u}_t$.

2) On régresse $\hat{u}_t$ sur la/les variables explicatives X_t et sur les valeurs laguées des résidus obtenus en 1ère étape (càd. $\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-p}$).

Si $p = 4$ ($\rightarrow n-4$ observations) :

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \rho_3 \hat{u}_{t-3} + \rho_4 \hat{u}_{t-4} + \varepsilon_t$$

On calcule le R^2 de cette régression auxiliaire.

(on inclut X_t ds rég. aux. pour tenir compte du fait qu'elle n'est p-ê pas strictmt non stochast.)

3) BG ont montré que lorsque la taille de l'échantillon est grande :

$$(n-p) R^2 \underset{asy}{\sim} \chi^2_p$$

(asympt. $(n-p) . R^2$ suit loi de χ^2 à p df)

Règle de décision :

si $(n-p) R^2$ excède valeur critique de la chi-carré $\Rightarrow$ RHo (au moins un des paramètres „rho" est signif. $\neq$ de zéro).

Remarques :

1) Si on teste présence processus AR(1), test de BG est appelé le test „m" de Durbin.
2) Test de BG ne spécifie pas la valeur de p
Dans certains cas, on peut utiliser critères d'information d'Akaike et Schwartz.

2.3.7. Remèdes à l'autocorrélation

4 options :

1) Déterminer si le problème d'autocorrélation est „pure", càd. s'il ne découle pas d'un biais de spécification (modèle mal spécifié, variables imp. omises, forme fctionnelle pas adaptée, ...)

2) Si autocorrélation pure, on peut essayer de transformer modèle initial pour faire disparaître le pblme → MCG (méthode de Moindres Carrés Généralisés).

3) Si échantillon est grand, ô option = utiliser méthode de Newey - West (fournit des se corrigés ntmt pour autocorrélation).

4) Dans certains cas, on peut continuer à utiliser MCO.

2.3.8. Mauvaise spécification versus autocorrélation „pure"

Exemple : Relation lel salaires & productivité

$d = 0,1229$ → suggère autocorrélation positive

→ modèle mal spécifié ?

2 séries chronologiques → possible qu'elles aient une tendance commune → inclure trend ds modèle pour analyser relation lel salaires & productivité nette de la tendance commune.

$$\hat{Y}_t = 1,4752 + 1,3057\, X_t - 0,9032\, t$$
$$se = (13,18) \quad (0,2765) \quad (0,4203)$$
$$t = (0,1119) \quad (4,7230) \quad (-2,1490)$$

$R^2 = 0,9631$ $\qquad$ $d = 0,2046$

d reste faible, suggère autocorrélation „pure".

Relation lel salaires & productivité quadratique ?

$$\hat{Y}_t = -16,2181 + 1,9488\, X_t - 0,0079\, X_t^2$$
$$t = (-5,4891) \quad (24,9868) \quad (-15,9363)$$

$R^2 = 0,9947$ $\qquad$ $d = 1,02$

$d_L = 1,3191$, $d_U = 1,60$ ⇒ ccl : RH0 d'absence d'autocorrélat° positiv

Rem. : pour tester normalité utiliser p.ex. test de JB.

2.3.3. Corriger pour l'autocorrélation (pure) : la méthode des moindres carrés généralisés (MCG)

Manque d'efficience des estimateurs des MCO en présence d'autocorrélation → remède dépend de notre connaissance qt à la nature de l'interdépendance ie/ les erreurs.

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (1)$$

où $u_t = \rho u_{t-1} + \varepsilon_t \qquad (2)$

$(-1 < \rho < 1)$

Deux possibilités :

(i) Valeur de ρ est connue

(ii) Valeur de ρ est inconnue (mais elle peut être estimée)

(i) Lorsque valeur de ρ est connue

Si ρ est connue pblme d'autocorrélation facilmt résolu.

Si éq. (1) valable en t, elle l'est églmt en $(t-1)$ ⟹ $Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad (3)$

$$\Rightarrow \rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \qquad (4)$$

(1) − (4) :

$$(Y_t - \rho Y_{t-1}) = \beta_1 (1-\rho) + \beta_2 (X_t - \rho X_{t-1}) + \varepsilon_t$$

où $\varepsilon_t = u_t - \rho u_{t-1}$

$$\Rightarrow \quad Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t$$

où
$$\begin{cases} \beta_1^* = \beta_1 (1-\rho) \\ Y_t^* = (Y_t - \rho Y_{t-1}) \\ X_t^* = (X_t - \rho X_{t-1}) \\ \beta_2^* = \beta_2 \\ \varepsilon_t = u_t - \rho u_{t-1} \end{cases}$$

Comme ε_t satisfait aux hyp. habituelles des MCO, on peut appliquer MCO aux données transformées Y^* et X^*
$\Rightarrow$ estimateurs BLUE.

Rappel : MCG = appliquer MCO aux données modifiées qui satisfont hyp. classiques.

Lorsque ρ est connue $\longrightarrow$ appliquer MCG : on soustrait de chaque variable une proportion ρ de sa valeur à la période précédente et on estime le M par MCO.

On appelle cette régression l'équation en différence généralisée („generalized difference equation").

Opération engendre perte de la première obs.
Pour éviter cela, on transforme première obs de Y et X comme suite :

$$Y_1^* = Y_1 \sqrt{1-\rho^2}$$
$$X_1^* = X_1 \sqrt{1-\rho^2}$$

$\Rightarrow$ transformation de Prais - Winsten.

Rem. : si on laisse 1ère obs. sans la transf. $\Rightarrow$ hétérosc.

(ii) Lorsque valeur de ρ est inconnue

En pratique, ρ rarement connue → MCG difficile à utiliser → trouver méthode pour estimer valeur de ρ.

- Méthode en différence première

Soit $\rho = +1$: éq en ≠ généralisée ≡ éq. en ≠ 1ère

$$\Rightarrow Y_t - Y_{t-1} = \beta_2 (X_t - X_{t-1}) + (u_t - u_{t-1})$$

$$\Rightarrow \Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \quad (\text{où } \varepsilon_t \text{ white noise})$$

Qd utiliser cette méthode ?

Lorsque ρ est très élevé (p.ex. > 0,8) ou d° très faib

Lorsque $d < R^2$ (cf. Maddala)

Exemple : $d\ (= 0{,}1229) < R^2\ (= 0{,}9584)$

$$\widehat{\Delta Y_t} = 0{,}7199\ \Delta X_t$$
$$t = (9{,}2073)$$
$$R^2 = 0{,}3610$$
$$d = 1{,}5096$$

→ autocorrélat° ne pose plus problème

$$\hat{Y}_t = 29{,}5192 + 0{,}7136\ X_t$$
$$se = (1{,}9423) \quad (0{,}0241)$$
$$t = (15{,}1977) \quad (29{,}6066)$$
$$R^2 = 0{,}9584$$
$$d = 0{,}1229 \quad , \quad \hat{\sigma} = 2{,}6755$$

Autre avantage : (dans certains cas) transf. en $\neq$ 1ère permet de rendre des séries chronologiques stationnaires

Si terme d'erreur suit processus AR(1) : $u_t = \rho u_{t-1} + \varepsilon_t$, on a que :

$$Var(u_t) = E(u_t^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

$$cov(u_t, u_{t+s}) = E(u_t u_{t+s}) = \rho^s \cdot \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

Si $\rho = 1$, $var(u_t)$ et $cov(u_t, u_{t+s})$ sont infinies
$\Rightarrow u_t$ est non stationnaire.

Si $\rho = 1 \Rightarrow u_t = u_{t-1} + \varepsilon_t$
$\Rightarrow u_t - u_{t-1} = \Delta u_t = \varepsilon_t$

Δu_t est stationnaire

Test de Berenblutt - Webb $(H_0 : \rho = 1)$

$$\boxed{g = \frac{\sum_{t=2}^{n} \hat{e}_t^2}{\sum_{t=1}^{n} \hat{u}_t^2}}$$

où $\hat{u}_t$: résidus des MCO de la régression en niveau.
$\hat{e}_t$: résidus des MCO de la régression en $\neq$ 1ère.

$d \approx 2(1-\hat{\rho}) \longrightarrow$ sous $H_0 : \rho = 1$, on s'attend à ce que d soit proche de zéro.

Exemple : $\sum_{t=1}^{n} \hat{u}_t^2 = 272,022$ et $\sum_{t=2}^{n} \hat{e}_t^2 = 0,33427$

$$\Rightarrow g = \frac{0,33427}{272,022} = 0,0012$$

Pour 39 obsv. et 1 var. expl. : $d_L = 1,435$ et $d_U = 1,540$ à 5%

g calculée $< d_L$ $\Rightarrow$ on ne rejette pas $H_0 : \rho = 1$.

$\Rightarrow$ résultats régression en $\neq$ 1ère paraissent acceptables.

- <u>Valeur de ρ basée sur „d" de Durbin-Watson</u>

$d \approx 2(1-\hat{\rho}) \quad \Rightarrow \quad \boxed{\hat{\rho} \approx 1 - \frac{d}{2}} \qquad (1)$

Si gd échantillon, utiliser valeur de d pour estimer ρ, transformer nos données conformément à l'éq. en diff. généralisée (à p. de $\hat{\rho}$).

Si petit échantillon, utiliser relation (modifiée) de Theil & Nagar (cf. Gujarati)

<u>Exemple</u> : $d = 0,1229 \Rightarrow \hat{\rho} \approx 0,9386$

$(Y_t - \hat{\rho} Y_{t-1}) = \beta_1 (1-\hat{\rho}) + \beta_2 (X_t - \hat{\rho} X_{t-1}) + \varepsilon_t$

- <u>Valeur de ρ estimée à partir des résidus</u>

si terme d'erreur u_t suit processus AR(1) : $u_t = \rho u_{t-1} + \varepsilon_t$
$\Rightarrow$ estimer ρ en régressant $\hat{u}_t$ sur $\hat{u}_{t-1}$
car $\hat{u}_t \equiv$ estimateur consistent du terme d'erreur u_t
$\Rightarrow$ estimer : $\hat{u}_t = \hat{\rho}\, \hat{u}_{t-1} + v_t$
(où $\hat{u}_t$ sont les résidus de la régression en niveau)

Rem. : pas d'intercepte
<u>Exemple</u> : $\hat{u}_t = 0,9142\ \hat{u}_{t-1}$
$t = (16,2281) \qquad R^2 = 0,8736$
$\Rightarrow \hat{\rho} = 0,9142$ ... à utiliser pour transformer les variables & estimer l'éq. en $\neq$ généralisée.

• <u>Méthodes itératives pour estimer ρ</u>

Exemples : méthode itérative de Cochrane - Orcutt
méthode en 2 étapes de Durbin
méthode de scanning de Hildreth - Lu

<u>Remarques</u> :

1) Estimateur MCO sont consistants (indpd de l'hyp d'autocorrélat°. Pour gds échantillons, qlq soit méthode d'estimation de ρ (d de DW, $résidus_t \sim résidus_{t-1}$, Cochrane - Orcutt), on obtiendra estimation consistante de la vraie valeur de ρ.

2) Méthodes d'estimation en 2 étapes. Comme on utilise $\hat{\rho}$ plutôt que ρ → moindres carrés généralisés "estimés" ("estimated" GLS ou EGLS ; feasible GLS ou FGLS)

3) Pour de petits échantillons EGLS fournit estimateurs qui ne sont pas nécess. BLUE.

4) Pour de petits échantillons, important d'inclure la 1ère observation à la Prais - Winsten.

→ MCG estimés avec inclusion de la première observation à la Prais - Winsten
≡ Full estimated generalized least squares, FEGLS ou full GLS.

2.3.10. Correction des erreurs standard des MCO par la méthode de Newey - West

Procédure de Newey - West fournit des erreurs standard qui sont HAC (heteroscedasticity and autocorrelation consistent).

≡ méthode asymptotique
(si gd échantillon pas lieu d'utiliser MCG estimés).

Exemple :

Procédure de Newey - West	Sans correction pour autocorr. / hétérosc.
$\hat{Y}_t = 29,5192 + 0,7136\ X_t$	$\hat{Y}_t = 29,5192 + 0,7136\ X_t$
se = (4,1180) (0,0512)	se = (1,9423) (0,0241)
	t = (15,1977) (29,6066)
$R^2 = 0,9584$ d = 0,1229	$R^2 = 0,9584$
	d = 0,1229 $\hat{\sigma} = 2,675$

⇒ Méthode des MCO habituelle (sans correction) sous-estime la vraie valeur des erreurs standard.

! stat. „d" est la m̄ ds les 2 cas mais ds procédure de Newey - West les erreurs standard ont été corrigées pour autocorrélat° & hétéroscédasticité.

- En présence d'autocorrélation, estimateurs des MCO sont non biaisés, consistants et asympt. normaux. Mais, ils ne sont pas efficients → inférence basée sur t, F et χ^2 n'est pas appropriée.

- MCG estimés et procédure de Newey - West fournissent estimateurs asympt. efficients.

  Mais propriétés stat. de ces estimateurs mal connues pour petits échantillons → si échantillon petit, procédure de Newey - West ou MCG estimés peuvent être moins appropriés que MCO.

  Griliches et Rao (1969) ont montré (à l'aide de simulations Monte Carlo) que lorsque l'éch. est relativmt petit et que $\rho < 0{,}3$, la méthode MCO aussi bonne voire meilleure que celle des MCG estimés

2.3.11. <u>Modèles ARCH et GARCH</u>

Si <u>carré</u> des erreurs au tmps t est corrélé au <u>carré</u> des erreurs au tmps $(t-1)$: „ autoregressive conditional heterosceda sticity " ≡ ARCH

Si carré des erreurs est corrélé au carré des erreurs à +ieurs périodes d'intervalle : „ generalized autoregressive conditional heteroscedasticity " ≡ GARCH
⇒ autocorrélat° pas confinée à la relat° (el erreurs présentes et passées, elle peut aussi concerner la variance présente et passée des erreurs.

2.4. Modélisation économétrique : spécification du modèle et tests de diagnostique

Questions :

1) Quels sont les critères pour choisir la spécification d'un modèle ?
2) Quels types d'erreurs de spécification rencontre-t-on en pratique ?
3) Quelles sont les csquences des erreurs de spécificat° ?
4) Comment détecte-t-on les erreurs de spécificat° ?
5) Quels sont les remèdes et bénéfices ?
6) Comment évalue-t-on les performances relatives de modèles concurrents ?

2.4.1. Critères de sélection d'un modèle

Critères qu'un bon modèle devrait satisfaire (d'après Hendry et Richard) :

* Il faut que les données soient admissibles
* Il faut que le modèle soit compatible avec la théorie.
* Il faut que les régresseurs soient (au moins) faiblement exogènes. Au mieux non stochastiques.
* La valeur des paramètres doit être stable dans le temps.
* Il faut que les données soient cohérentes.

2.4.2. Types d'erreurs de spécification

$$\boxed{Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i}} \qquad (1)$$

où Y = coût total de production

X = output

- $$\boxed{Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i}} \qquad (2)$$

Si on estime (2) et que (1) est le bon modèle, erreur de spécification = omettre variable explicative d'intérêt (X_i^3).

$$\Rightarrow \quad u_{2i} = u_{1i} + \beta_4 X_i^3$$

- $$\boxed{Y_i = \lambda_1 + \lambda_2 X_i + \lambda_3 X_i^2 + \lambda_4 X_i^3 + \lambda_4 X_i^4 + u_{3i}} \qquad (3)$$

Si on estime (3) et que (1) est correct, erreur de spécification = inclure une variable explicative non nécessaire (superflue) car λ_5 ds (1) est nul.

$$\Rightarrow \quad u_{3i} = u_{1i} - \lambda_5 X_i^4$$

$$= u_{1i} \qquad (\text{car } \lambda_5 = 0)$$

- $$\boxed{\ln Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 X_i^2 + \gamma_4 X_i^3 + u_{4i}} \qquad (4)$$

Erreur de spécification = utilisation d'une forme fctionnelle inadéquate.

- $$\boxed{Y_i^* = \beta_1^* + \beta_2^* X_i^* + \beta_3^* X_i^{*2} + \beta_4 X_i^{*3} + u_i^*} \qquad (5)$$

où $Y_i^* = Y_i + \varepsilon_i$
$X_i^* = X_i + \omega_i$

et ε_i, ω_i sont des erreurs de mesure.

Erreurs de mesure : plutôt que d'utiliser les vraies variables, on utilise les proxy Y_i^* et X_i^* (mesurées avec erreurs).

- Finalmt, erreur de spécification concernant modélisation du terme d'erreur.

<u>Exemple</u> : $\boxed{Y_i = \beta X_i u_i} \qquad (7)$

où $(\ln u_i)$ satisfait hyp. classiques.
Plutôt que d'estimer (7), on estime (8) :

$$\boxed{Y_i = \alpha X_i + u_i} \qquad (8)$$

Si (7) est bon modèle, $\hat{\alpha}$ est estimateur biaisé de $\beta \Rightarrow E(\hat{\alpha}) \neq \beta$

<u>En résumé</u> :

1. Omettre variable importante.
2. Inclure variable superflue.
3. Choisir forme fctionnelle inadaptée.
4. Utiliser des variables mesurées avec erreur.
5. Ne pas spécifier correctmt le terme d'erreur stochastique.

2.4.3. Conséquences des erreurs de spécification

(i) Omission d'une variable d'intérêt

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \text{(vrai modèle)}$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i \qquad \text{(modèle estimé)}$$

Conséquences de l'omission de X_3 :

1. Si les variables X_2 et X_3 sont corrélées, $\hat{\alpha}_1$ et $\hat{\alpha}_2$ sont des estimateurs biaisés et inconsistants $\rightarrow$
$E(\hat{\alpha}_1) \neq \beta_1$, $E(\hat{\alpha}_2) \neq \beta_2$

$$\underset{n \to \infty}{\text{plim}}\ \hat{\alpha}_1 \neq \beta_1 \quad , \quad \underset{n \to \infty}{\text{plim}}\ \hat{\alpha}_2 \neq \beta_2$$

Biais persiste asymptotiquement.

2. Si les variables X_2 et X_3 sont non corrélées, $\hat{\alpha}_2$ est non biaisé mais $\hat{\alpha}_1$ est biaisé.
3. La variance homoscédastique du terme d'erreur n'est pas estimée correctement.
4. La mesure habituelle de la variance de l'estimateur $\hat{\alpha}_2$ $\left(= \sigma^2 / \Sigma x_{2i}^2 \right)$ est un estimateur biaisé de la variance du vrai estimateur $\hat{\beta}_2$.
5. Dès lors, intervalles de confiance et tests d'hyp. habituels sont susceptibles de conduire à des ccl. erronées.
6. Prévisions à partir du modèle et intervalles de confiance des prévision ne sont pas corrects.

Intuition :

* On peut montrer que : $\boxed{E(\hat{\alpha}_2) = \beta_2 + \beta_3 \, b_{23}}$ (1)

où b_{23} = coefficient de pente dans la régression de la variable X_3 sur la variable X_2 →

$$b_{23} = \frac{\sum x_{3i} x_{2i}}{\sum x_{2i}^2}$$

$E(\hat{\alpha}_2) \neq \beta_2$ sauf si β_3 et/ou $b_{23} = 0$

Par hyp., on sait que $\beta_3 \neq 0$ (sinon pas de biais de spéc.

$b_{23} = 0$, si X_2 et X_3 non correlées (hyp. rarement satisfaite avec séries éco.)

⇒ ampleur du biais dépend du produit de β_3 et b_{23}

si β_3 et $b_{23} > 0$ ⇒ en moyenne $\hat{\alpha}_2$ surestime β_2

* Nous savons que :

$$var(\hat{\alpha}_2) = \sigma^2 / \sum x_{2i}^2$$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} = \frac{\sigma^2}{\sum x_{2i}^2} \; VIF$$

Comme $var(\hat{\beta}_2)$ est non biaisée, $var(\hat{\alpha}_2)$ est généralement biaisée : $0 < r_{23}^2 < 1$ ⇒ $var(\hat{\alpha}_2) < var(\hat{\beta}_2)$

Rem : $r_{23} \neq 0$ car variables éco. très st corrélées.

$r_{23} \neq 1$ car par hyp. pas de colinéarité pfte.

⇒ dilemme : $\hat{\alpha}_2$ est biaisé mais variance + faibl que l'estimateur non biaisé $\hat{\beta}_2$.

Estimation de σ^2 sera ≠ selon qu'on estime 1er ou 2nd modèle.

$$\hat{\sigma}^2 = SCR \,/\, df$$

où SCR = somme des carrés résiduelle
df = degrees of freedom

Si nbre de variables explicatives ↑ ⇒ { SCR ↓ généralmt ; df ↓ }

Si on ajoute variable supplém. au pouvoir explicatif imp. ⇒ baisse de SCR sera + imp. que baisse du nbre de degrés de liberté ⇒ inclusion variable expl. supplém. ↓ le biais et variance de l'estimateur.

Si variable incluse a un impact limité sur variable dpdte et que var. expl. sont fortmt corrélées (VIF est gd), ajout de cette var. réduira biais de l'estimateur mais variance ↑ ⇒ trade off

lel précision & précision de l'estimateur. Solution ?

Par exemple, choisir estimateur dont erreur moyenne au carré (mean squared error ≡ MSE) est la + faible.

<u>Rappel</u> :

$$\begin{aligned}
MSE(\hat{\theta}) &= E(\hat{\theta}-\theta)^2 \\
&= E\left[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta\right]^2 \\
&= E\left[\hat{\theta} - E(\hat{\theta})\right]^2 + E\left[E(\hat{\theta}) - \theta\right]^2 \\
&\quad + \underbrace{2\,E\{[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]\}}_{=0} \\
&= E\left[\hat{\theta} - E(\hat{\theta})\right]^2 + E\left[E(\hat{\theta}) - \theta\right]^2 \\
&= var(\hat{\theta}) + bias(\hat{\theta})^2 = \text{variance de } \hat{\theta} \text{ plus}
\end{aligned}$$

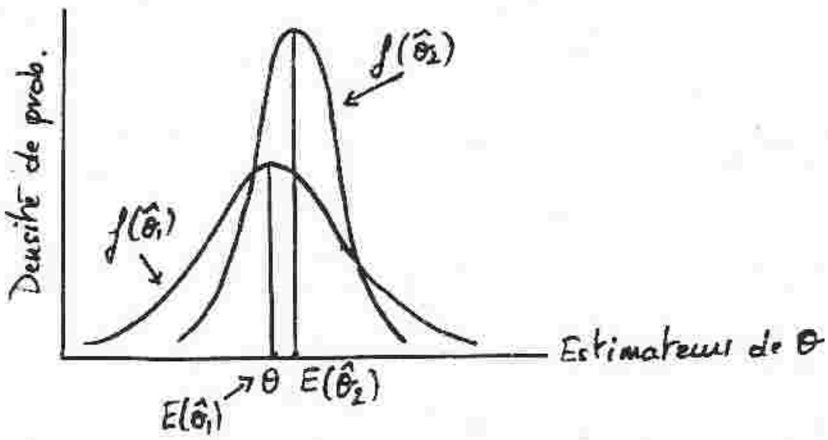
Remarque :

Pq $2\, E\left\{ [\hat{\theta} - E(\hat{\theta})]\,[E(\hat{\theta}) - \theta] \right\} = 0$?

$$\Rightarrow 2 \left\{ [E(\hat{\theta})]^2 - [E(\hat{\theta})]^2 - \theta E(\hat{\theta}) + \theta E(\hat{\theta}) \right\} = 0$$

$E[E(\hat{\theta}) - \theta]^2 = [E(\hat{\theta}) - \theta]^2$ (car espérance d'une cste = cste)

Graphiquement :



* Supposons que X_2 et X_3 non corrélés $\Rightarrow r_{23} = 0 \Rightarrow b_{23} =$

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 b_{23} \quad \Rightarrow \quad E(\hat{\alpha}_2) = \beta_2$$

Estimateur $\hat{\alpha}_2$ est non biaisé.

On suit que :

$var(\hat{\alpha}_2) = \sigma^2 / \Sigma x_{2i}^2$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\Sigma x_{2i}^2 (1 - r_{23}^2)} = \frac{\sigma^2}{\Sigma x_{2i}^2}\, VIF$$

$\Rightarrow$ variance de $\hat{\alpha}_2$ et $\hat{\beta}_2$ semble identique.

En réalité, $var(\hat{\alpha}_2)$ biaisée car estimation de σ^2 n'est pas la m̂ lorsqu'on omet une var.

* Quid de $\hat{\alpha}_1$ lorsque $r_{23} = 0$?
$\hat{\alpha}_1$ est tjs biaisé.
Pq ?
$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$
$\hat{\alpha}_1 = \bar{Y} - \hat{\alpha}_2 \bar{X}_2$

(ii) <u>Inclusion d'une variable superflue</u>

$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$ (= bon modèle)
$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i$ (= modèle estimé)

<u>Conséquences de l'inclusion d'une variable superflue</u> :

1. Estimateurs des MCO des paramètres du „mauvais" modèle sont non biaisés et consistants $\rightarrow$
$E(\hat{\alpha}_1) = \beta_1$, $E(\hat{\alpha}_2) = \beta_2$, $E(\hat{\alpha}_3) = \beta_3 = 0$

$\underset{n \to \infty}{\text{plim}}\, \hat{\alpha}_1 = \beta_1$, $\underset{n \to \infty}{\text{plim}}\, \hat{\alpha}_2 = \beta_2$

2. Variance du terme d'erreur σ^2 correctmt estimée.
3. Intervalles de confiance et tests d'hypothèse habituels restent valables.
4. Cpdt, estimateurs $\hat{\alpha}$ générlmt inefficients (variance des $\hat{\alpha}$ générlmt + gd que celle des $\hat{\beta}$)

Pq ? $$\left.\begin{array}{l} Var(\hat{\beta}_2) = \sigma^2 / \Sigma x_{2i}^2 \\ Var(\hat{\alpha}_2) = \dfrac{\sigma^2}{\Sigma x_{2i}^2 (1 - r_{23}^2)} \end{array}\right\} \Rightarrow \frac{Var(\hat{\alpha}_2)}{Var(\hat{\beta}_2)} = \frac{1}{1 - r_{23}^2}$$

Comme $0 \leq r_{23}^2 \leq 1 \Rightarrow Var(\hat{\alpha}_2) \geq Var(\hat{\beta}_2)$

2.4.4. Les tests d'erreur de spécification

(i) Détecter la présence de variables superflues

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

Faut-il inclure X_k ? Test en „t" habituel → si $t = \hat{\beta}_k / se(\hat{\beta}_k)$ significative, on garde X_k

Faut-il inclure X_2, X_3 et X_4 ? F test habituel → $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$

Il est déconseillé de construire un modèle de façon „itérative" → „data mining", „regression fishing", „data cruching", ...

(ii) Tester pour des variables omises ou une forme fonctionnelle inadaptée

Examiner les résidus

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i} \quad \text{(vrai modèle)}$$

où $\begin{cases} Y = \text{coût total de production} \\ X = \text{niveau de production} \end{cases}$

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \quad \text{(modèle alternatif (a))}$$

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \quad \text{(modèle alternatif (b))}$$

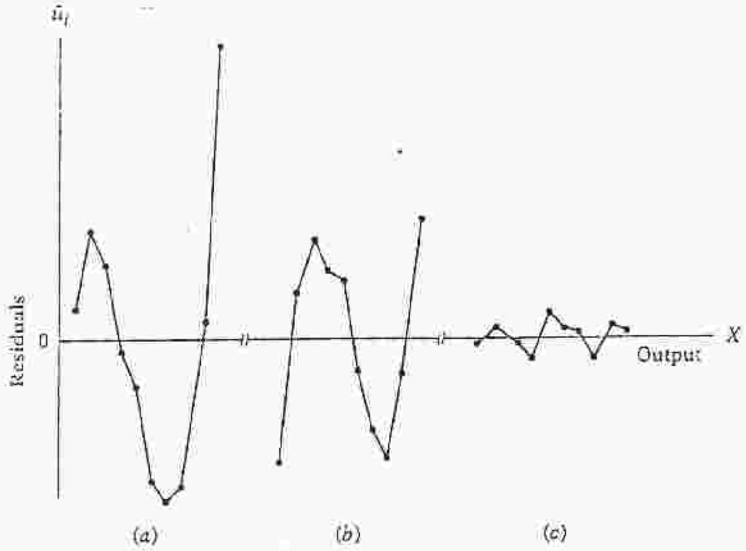


FIGURE 13.1 Residuals $\hat{u}_i$ from (a) linear, (b) quadratic, and (c) cubic total cost functions.

La statistique d de Durbin-Watson

TABLE 13.1 ESTIMATED RESIDUALS FROM THE LINEAR, QUADRATIC, AND CUBIC TOTAL COST FUNCTIONS

Observation number	$\hat{u}_i$, linear model*	$\hat{u}_i$, quadratic model†	$\hat{u}_i$, cubic model**
1	6.600	−23.900	−0.222
2	19.667	9.500	1.607
3	13.733	18.817	−0.915
4	−2.200	13.050	−4.426
5	−9.133	11.200	4.435
6	−26.067	−5.733	1.032
7	−32.000	−16.750	0.726
8	−28.933	−23.850	−4.119
9	4.133	−6.033	1.859
10	54.200	23.700	0.022

$$^{*}\hat{Y}_i = \underset{\substack{(19.021)\\(8.752)}}{166.467} + \underset{\substack{(3.066)\\(6.502)}}{19.933X_i} \qquad R^2 = 0.8409 \quad \bar{R}^2 = 0.8210 \quad d = 0.716$$

pour $n = 10$, $k = 1$, $\alpha = 0,0$:
$d_L = 0,879$; $d_U = 1,320$

$$^{\dagger}\hat{Y}_i = \underset{\substack{(23.488)\\(9.468)}}{222.383} - \underset{\substack{(9.809)\\(-0.818)}}{8.0250X_i} + \underset{\substack{(0.869)\\(2.925)}}{2.542X_i^2} \qquad R^2 = 0.9284 \quad \bar{R}^2 = 0.9079 \quad d = 1.038$$

$d_L = 0,697$; $d_U = 1,641$

$$^{**}\hat{Y}_i = \underset{\substack{(6.375)\\(22.238)}}{141.767} + \underset{\substack{(4.778)\\(13.285)}}{63.478X_i} - \underset{\substack{(0.9856)\\(-13.151)}}{12.962X_i^2} + \underset{\substack{(0.0592)\\(15.861)}}{0.939X_i^3} \qquad R^2 = 0.9983 \quad \bar{R}^2 = 0.9975 \quad d = 2.70$$

$d_L = 0,525$; $d_U = 2,016$

Corrélat° positive dans résidus des 2 premières spécifications ne découle pas d'un pblme d'autocorrélation « pure » mais bien d'un biais de spécification.

Si on exclut X_i^3 de la fction de coût $\Rightarrow u_{2i} = (u_{1i} + \beta_4 X_i^3)$

Test de Durbin - Watson en pratique :

1. On estime le modèle de régression et on obtient les résidus des MCO.
2. Imaginons qu'on veuille test si modèle mal spécifié en raison de l'exclusion de la variable Z.
   Dans ce cas, on ordonne résidus (obtenus en 1ère étape) de manière croissante // à Z.
3. Calcule statistique d à partir des résidus ordonnés

$$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2}$$

(peut être utilisé pour time series & cross-section).

4. Règle de décision : si statistique d est significative, on rejette hypothèse nulle que modèle est correctmt spécifié. Remède (si d signif.) : inclure variable Z !

Exemple : la variable Z (= X), le coût de production, était ordonnée de façon croissante $\longrightarrow$ il en va de m̂ pour X^2 et X^3 $\longrightarrow$ stat. d suggèrent la présence d'une erreur de spécif. ds modèle linéaire & quadrat. $\longrightarrow$ estimer modèle cubique.

Le test de Ramsey (RESET)

RESET ≡ regression specification error

Intuition :

$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i}$

où $\begin{cases} Y = \text{coût total} \\ X = \text{niveau de production} \end{cases}$

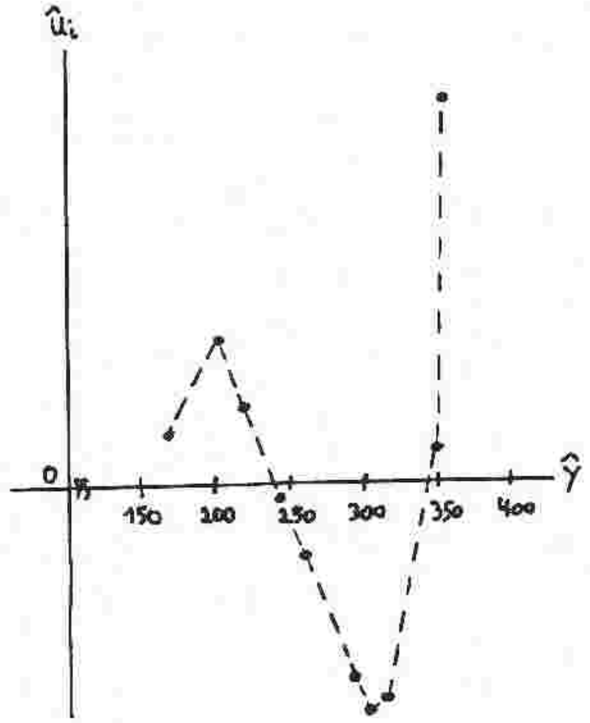


Fig. Residuals $\hat{u}_i$ and estimated Y from the linear cost function : $Y_i = \lambda_1 + \lambda_2 X_i + u_i$

Résidus présentent une structure systématique et leur moyenne dépend de la valeur de $\hat{Y}_i$ ⇒ introduct° de $\hat{Y}_i$ (sous une forme particulière) devrait pouvoir ↑ R^2.

Règle de décision : si ↑ du R^2 est statist. significative sur base d'un F-test, on rejette hyp. que fct° de

 coût linéaire est une bonne spécification.

En pratique :

1. On estime modèle, p.ex. $\boxed{Y_i = \lambda_1 + \lambda_2 X_i + u_{3i}}$ (1) et on obtient estimations des observations Y_i ($\hat{Y}_i$)
2. On réestime modèle en incluant $\hat{Y}_i$ sous forme additive.
3. Comme on observe relation curviligne entre $\hat{u}_i$ et $\hat{Y}_i$ (cf. Fig. slide 2.88.), on pourrait décider d'inclure $\hat{Y}_i$ au carré et au cube:

$$\boxed{Y_i = \beta_1 + \beta_2 X_i + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + u_i} \quad (2)$$

4. R^2_{new} ≡ coeff. dét. modèle (2)
   R^2_{old} ≡ coeff. dét. modèle (1)

$$\boxed{F = \frac{(R^2_{new} - R^2_{old}) \,/\, \text{nber of new regressors}}{(1 - R^2_{new}) \,/\, (n - \text{nber of parameters in new model})}}$$

5. Si statistique F est significative, on rejette hypothèse que modèle (1) est correctmt spécifié.

Exemple :

$$\hat{Y}_i = 166,467 + 19,933\, X_i$$
$$se = (19,021) \quad (3,066)$$
$$R^2 = 0,8409$$

$$\hat{Y}_i = 2140,7223 + 476,6557\, X_i - 0,09187\, \hat{Y}_i^2 + 0,000119\, \hat{Y}_i^3$$
$$se = (132,0044) \quad (33,3951) \quad (0,0062) \quad (0,000074)$$
$$R^2 = 0,9983$$

$$F = \frac{(0,9983 - 0,8409) \,/\, 2}{(1 - 0,9983) \,/\, (10-4)} = 284,4035 \quad \text{(htmt sign.)}$$

⇒ modèle (1) mal spécifié.

Avantage et inconvénient du test RESET :
ne requiert pas de spécification alternative ...
si modèle s'avère mal spécifié on ne sait pas quelle alternative choisir.

Le test du multiplicateur de Lagrange (LM test)

Fonction de coût linéaire = version restreinte de la fct° de coût cubique. (suppose que coeff. relatifs à X_i^2 et X_i^3 sont égaux à zéro).

$\Rightarrow$ LM test :

1. Estimation du modèle restreint : $\boxed{Y_i = \lambda_1 + \lambda_2 X_i + u}$
   $\longrightarrow$ on obtient résidus des MCO ($\hat{u}_i$)
2. Si modèle non restreint : $\boxed{Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u}$ est correct, résidus du modèle restreint devraient dépendre de X_i^2 et X_i^3.
3. $\boxed{\hat{u}_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + \alpha_4 X_i^3 + v_i}$ (*)
   où v_i terme d'erreur qui satisfait aux hyp. class.
4. Asympt. n (taille de l'éch.) fois R^2 (coeff. de déterminat° de rég. auxiliaire (*)) suit une loi de chi-carré où df = nbre de restrictions imposées dans modèle restreint.

$$\Rightarrow \boxed{n . R^2 \underset{asy}{\sim} \chi^2_{df = \text{nbre de restrictions}}}$$

5. Règle de décision : si χ^2 calculée excède valeur critique au niveau prob. α, on rejette reg. restreint

Exemple :

$$\hat{Y}_i = 166,467 + 19,333 \; X_i$$

où $\begin{cases} Y = \text{coût total} \\ X = \text{niveau d'output} \end{cases}$

Régression auxiliaire :

$$\hat{u}_i = -24,7 + 43,5443 \; X_i + 12,9615 \; X_i^2 + 0,9396 \; X_i^3$$
$$se = (6,375) \quad (4,779) \quad (0,986) \quad (0,059)$$
$$R^2 = 0,9896$$

$$n * R^2 = 10 * 0,9896 = 9,896$$

Valeur critique χ^2 à 2 df $= 9,21$ $(\alpha = 0,01)$
$\chi^2_{calculée}$ est significative à 1% $\Rightarrow$ on rejette modèle restreint.

2.4.5. Modèles imbriqués et non imbriqués

Modèle A : $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$
Modèle B : $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

→ Modèle B est imbriqué ds modèle A car si sur base d'un F-test, on ne rejette pas $H_0 : \beta_4 = \beta_5 = 0$, modèle A se résume au modèle B.

Modèle C : $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$
Modèle D : $Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$

où $X \neq Z$

→ Modèles C et D ne sont pas imbriqués car aucun n'est un cas particulier de l'autre..

Modèles peuvent être non imbriqués suite à l'utilisation d'une forme fctionnelle ≠, par exemple :

Modèle D : $Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$
Modèle E : $Y_i = \beta_1 + \beta_2 \ln Z_{2i} + \beta_3 \ln Z_{3i} + w_i$

Tests relatifs aux modèles non imbriqués

Approche „discriminante" : sélectionner un modèle sur base d'un critère relatif à la qualité de l'ajstmt des modèle.

Approche „de discernement" : choisir modèle qui est le plus adéquat en utilisant information contenue dans l'ô modèle

(i) Approche discriminante

Modèle C : $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$
Modèle D : $Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$

Le coefficient de détermination

$$\boxed{R = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}}$$

Pblmes ?
R^2 mesure qualité de l'ajustmt au sein de l'échantillon.
Il faut que variable dpdt soit identique.
R^2 ne diminue jamais lorsque # var. expl. ↑

Le coefficient de détermination ajusté

$$\boxed{\bar{R}^2 = 1 - \frac{SCR / (n-k)}{SCT / (n-1)} = 1 - (1-R^2)\,\frac{n-1}{n-k}}$$

$\Rightarrow \quad \bar{R}^2 \leq R^2$

(! pour que comparaison ait un sens, il faut que variables dépendantes soient identiques).

Le critère d'information de Akaike (AIC)

$$AIC = e^{2k/n}\,\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\,\frac{SCR}{n}$$

où $\begin{cases} k = \text{nbre de régresseurs (intercepte compris)} \\ n = \text{nbre d'observations} \end{cases}$

$$\Rightarrow \quad \boxed{\ln AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{SCR}{n}\right)}$$

où $\frac{2k}{n}$ = facteur de pénalité

Choisir modèle dont valeur critère AIC est la plus faible

Permet d'évaluer qualité de la prévision au sein de l'échantillon et en dehors.

Souvent utilisé pour déterminer le nbre de lags dans un modèle AR(p).

Le critère d'information de Schwartz (SIC)

$$SIC = n^{k/n} \frac{\sum \hat{u}^2}{n} = n^{k/n} \frac{SCR}{n}$$

$$\Rightarrow \boxed{\ln SIC = \frac{k}{n} \ln(n) + \ln\left(\frac{SCR}{n}\right)}$$

où $\left[\frac{k}{n} \ln(n)\right]$ = facteur de pénalité

Même avantages que critère AIC mais impose pénalité + imp. pour inclusion de var. expl. suppl.

On choisit modèle pour lequel critère SIC prend valeur la + faible.

(ii) Approche de discernement

Le F test non imbriqué (Encompassing F test)

Modèle C : $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$
Modèle D : $Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$

Faire un choix ICI C & D ?

Modèle F : $Y_i = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \lambda_4 Z_{2i} + \lambda_5 Z_{3i} + u_i$

$\Rightarrow$ Modèle F imbrique ou englobe modèles C et D.

Modèles C & D non imbriqués.

$$\left.\begin{array}{l}\text{Si modèle C correct : } \lambda_4 = \lambda_5 = 0 \\ \text{Si modèle D correct : } \lambda_2 = \lambda_3 = 0\end{array}\right\} \rightarrow \text{F test tradit.}$$

Inconvénients :

1. Le choix du modèle de référence peut avoir une influence déterminante sur le modèle sélectionné (surtt lorsqu'il y a une forte multicolinéarité le/ régresseurs des ≠ modèles)

2. Modèle F ne donne pas tjs lieu à une interprétat° économique cohérente.

Le J test de Davidson - MacKinnon

Supposons qu'on compare les modèles suivants :

Modèle C : $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$

Modèle D : $Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$

Différentes étapes :

1. On estime modèle D $\longrightarrow \hat{Y}_i^D$
2. On ajoute $\hat{Y}_i^D$ comme régresseurs du modèle C :
   $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 \hat{Y}_i^D + u_i$
3. $H_0 : \alpha_4 = 0$ (utiliser t test)
4. Si $\not{R}H_0 \longrightarrow$ on ne rejette pas l'hyp. que modèle C est le bon.

   Intuition : si $\not{R}H_0$ que $\alpha_4 = 0$, variables incluses dans D et pas dans C ne permettent pas d'améliorer perf. du modèle C $\Rightarrow$ modèle C englobe (encompasses) modèle D.
5. On renverse rôle des hyp : on estime C, on obtient $\hat{Y}_i^C$, on estime D en incluant $\hat{Y}_i^C$ ($Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 \hat{Y}_i^C + v_i$), on
2.95. teste $H_0 : \beta_4 = 0$. Si $\not{R}H_0$, on choisit D, sinon C.

Inconvénients :

1. 4 types de résultats :

Hyp. $\beta_4 = 0$ (H_0 : D vs. H_1 : C) \ Hyp. $\alpha_4 = 0$ (H_0 : C vs. H_1 : D)	Ne pas rejeter	Rejeter
Ne pas rejeter	Accepter C & D	Accepter D Rejeter C
Rejeter	Accepter C Rejeter D	Rejeter C & D

→ si 2 modèles rejetés : aucun modèle ne reflète correctmt le cmpt de la variable Y.

si 2 modèles acceptés : (apparemment) les données ne contiennent pas assez d'info pour les départager

2. t stat. suit uniq distrib. normale centrée réduite asympt. (pblme parfois pour tester signif. de $\hat{y}_i^{C \text{ ou } D}$)
→ pas très puissant si petit échantillon.
(rejette trop svt H_0 lorsque petit éch.)