Econométrie 1

Master Financial Economics

Michel BEINE

michel.beine@uni.lu

Universite du Luxembourg

Manuel

- La partie économétrique du cours d'économétrie I et Analyse des données est basé sur le livre Introductory Econometrics, a Modern Approach (second edition) de *Jeffrey M. Wooldridge*, édition Thomson South-Western.
- Sur le site web : http://www.swcollege.com/bef/wooldridge/ ... wooldridge2e/wooldridge2e.html
  il y a des données nécessaires pour les exercices et plusieurs liens intéressants.

Plan du cours (Partie Économétrie)

- Chapitre 1. Nature de l'économétrie et des données économique.

Partie I: Analyse de régression avec des données "cross-section".

- Chapitre 2. Le modèle de régression simple.
- Chapitre 3. Le modèle de régression multiple: estimation.
- Chapitre 4. Le modèle de régression multiple: inférence.
- Chapitre 5. Le modèle de régression multiple: MCO asymptotiques.

Plan du cours (Partie Économétrie)

- Chapitre 1. Nature de l'économétrie et des données économique.

Partie I: Analyse de régression avec des données “cross-section”.

- Chapitre 6. Le modèle de régression multiple: questions avancées.
- Chapitre 7. Le modèle de régression multiple avec information qualitative.
- Chapitre 8. Hétéroscédasticité.
- Chapitre 9. Spécification et problèmes liés aux données.

Chapitre 1: Nature de l'économétrie et données économiques

Nature de l'économétrie

- L'économétrie est une technique de plus en plus répandue.
- Elle permet de répondre à des questions économiques (ou non) telles que :
  - Quels sont les déterminants du salaire ?
  - Quelle est la meilleur prévision (selon moi) du PNB belge de l'année prochaine ?
  - Est-ce que le plan de résorption du chômage a été efficace ?
- L'économétrie est basée sur les développements des méthodes statistiques pour estimer des relations économiques, tester des théories économiques et évaluer des politiques (économiques).
- Une caractéristique importante de l'économétrie est qu'elle se base sur des données non-expérimentales.

Étapes d'une analyse économique

- Une analyse empirique utilise des données pour tester une théorie ou estimer une relation (économique) $\longrightarrow$ 5 étapes.
- 1. Formuler la question.
- 2. Dans certains cas un modèle économique est développé : équations mathématiques qui décrivent des relations. Exemple: maximization de fonction d'utilité $\longmapsto$ les individus agissent de manière à maximiser leur bien-être sous des contraintes de ressources.
- Extension de ce concept par le prix Nobel Gary Becker: modèle économique expliquant la participation d'un individu à un crime.
- Certains crimes ont des récompenses économiques mais la plupart ont des coûts. Selon Becker il y a un arbitrage coût-bénéfice.

Modèle économique du crime

- $y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$, où
- $y =$ nombre d'heures passées dans des activités criminelles.
- $x_1 =$ salaire pour une heure passées dans une activité criminelle.
- $x_2 =$ salaire pour une heure passées dans une activité légale.
- $x_3 =$ autres revenus (que dans des activités criminelles et légales).
- $x_4 =$ probabilité d'être attrapé.
- $x_5 =$ probabilité d'être reconnu coupable si attrapé.
- $x_6 =$ sentence attendue si attrapé et reconnu coupable.
- $x_7 =$ âge.

Étapes d'une analyse économique

- 3. Dans beaucoup de cas il faut se baser également sur l'intuition plutôt que sur un modèle économique établi. Exemple: $wage = f(educ, exper, training)$.
- 4. Ensuite il faut spécifier un modèle économétrique.
- a) $f(.)$?
- b) Quelles sont les variables observées. Que faire avec les variables non-observées (ou non-observables) ?
- Exemple du crime: $crime = \beta_0 + \beta_1 wage_m + \beta_2 otheinc + \beta_3 freqarr + \beta_4 freqconv + \beta_5 avgsen + \beta_6 age + u$ où entre autre $wage_m$ est le salaire horaire dans une activité légale, $freqconv$ est la fréquence de condamnation et $avgsen$ est la sentence moyenne.
- u capture tous les facteurs non observés tels que x_1, origine sociale mais aussi les erreurs de mesure.

Étapes d'une analyse économique

- Une fois le modèle spécifié, on peut poser des questions du style *est-ce que* $wage_m$ *influence le comportement criminel ?* $\rightarrow \beta_1 = 0$ contre $\beta_1 >=< 0$?
- 5. Finalement, une fois que nous disposons de données, nous pouvons estimer les paramètres $\beta_0, \ldots, \beta_6$ et procéder aux tests.

Structure des données

Il existe 3 types de données. Chaque type de données peut nécessiter des techniques économétriques particulières.

- Données “cross-section” ou transversales.
- Séries temporelles.
- Données de panel ou longitudinales.

Données “Cross-section”

- Échantillon d’individus, ménages, firmes, ..., pris à un point du temps donné.
- Important : on peut souvent supposer que les obs. = échantillon aléatoire $\longrightarrow$ simplifie l’analyse.
- Données très utilisées en économie et sciences sociales $\longrightarrow$ micro appliquée: marché du travail, finances publiques, organisation industrielle, économie spatiale, démographie, économie de la santé, etc.
- Exemple: Wage1.
- On se focalisera sur ce type de données dans ce cours.

Séries temporelles

- Séries chronologiques. Ex: PNB, importations, indices de prix, etc.
- Important: Rarement indépendantes au court du temps $\longrightarrow$ complexifie l'analyse.
- Différentes fréquences: annuel, trimestriel, mensuel, hebdomadaire, journalier, intra-journalier.
- Données très utilisées en macro-économie et en finance.
- Exemple: PRMINWGE.

Données de panel

- Série temporelle pour chaque unité/individu.
- Important : la même unité est observée plusieurs fois au court du temps.
- Exemple: WAGEPAN.

Causalité et notion *Ceteris Paribus*

- La plupart du temps, l'objectif d'un économiste est de montrer qu'une variable à un effet causal sur une autre. Exemple : l'éducation a un effet causal sur la productivité d'un travailleur.
- La notion *Ceteris Paribus* joue un rôle important dans cette représentation causale.
- Dans le modèle $wage = f(educ, exper, training)$ on peut être intéressé pas savoir l'effet d'une semaine de formation supplémentaire sur le salaire attendu, toute chose restant égale par ailleurs (ici $educ$ et $exper$).
- $\rightarrow$ le but est d'isoler l'effet de $training$ sur $wage$.
- Nous allons imposer certaines contraintes permettant ce type d'interprétation.

Chapitre 2: Modèle de régression simple

Modèle de régression simple

- La plupart des analyses économétriques commencent comme ceci: y *et* x *sont deux variables représentant une population et nous voulons expliquer* y *en fonction de* x, *c-à-d comment varie* y *lorsque* x *varie ?*
- En écrivant un modèle qui “explique y en fonction de x” on est face à 3 problèmes.
- 1. Comme il n’existe pas de relation exacte entre deux variables, comment tenir compte que d’autres facteurs (non observés) peuvent expliquer y ?
- 2. Quelle est la relation fonctionnelle entre y et x ?
- 3. Comment s’assurer que l’on capture une relation *Ceteris Paribus* entre y et x ?

Modèle de régression simple

- Objectif: estimer un modèle du type

$$wage = \beta_0 + \beta_1 educ + u.$$

- De manière générale:

$$y = \beta_0 + \beta_1 x + u.$$

- $\rightarrow$ cette relation est supposée tenir sur la population d'intérêt $=$ modèle linéaire de régression simple.

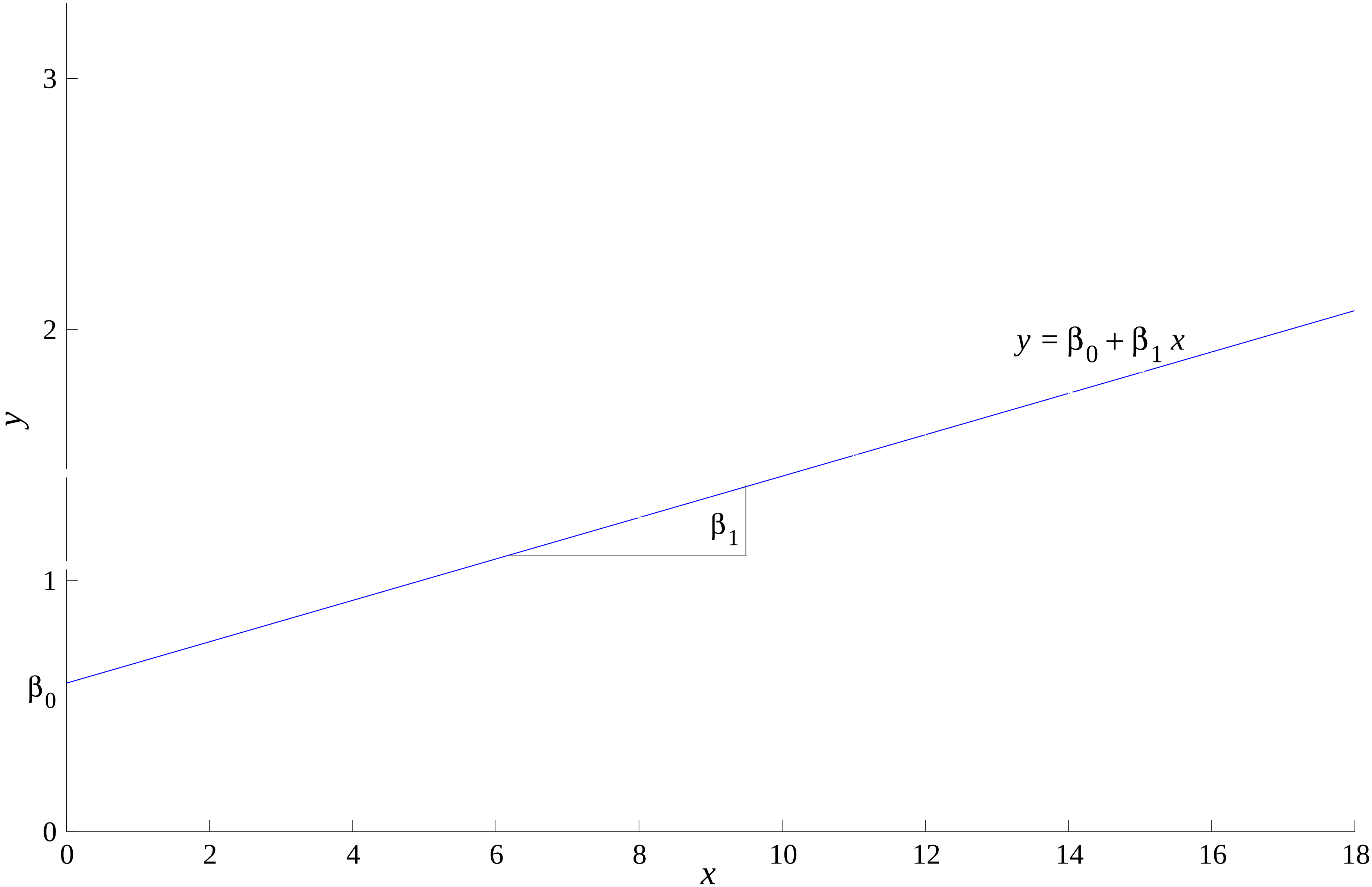
-

y	x
Variable dépendante	Variable indépendante
Variable à expliquer	Variable explicative
Variable de réaction	Variable de contrôle
Régressant	Régresseur

Modèle de régression simple

- u est le terme d'erreur (aléa) = facteurs *non-observés* autres que x qui affectent y.
- u et x sont des variables aléatoires.
- Relation fonctionnelle entre y et x ?
- $y = \beta_0 + \beta_1 x + u$.
  $\rightarrow$ Si les autres facteurs dans u restent inchangés, c-à-d $\Delta u = 0$, alors x a un effet linéaire sur y:
  $\Delta y = \beta_1 \Delta x$ si $\Delta u = 0$.

Modèle linéaire



Exemple

- $wage = \beta_0 + \beta_1 educ + u.$
- Si $wage$ est le salaire horaire en dollars et $educ$ le nombre d'années d'éducation.
- $\rightarrow \beta_1$ mesure le changement de salaire horaire dû à une année supplémentaire d'éducation, *Ceteris Paribus*.
- La linéarité implique que l'effet est le même pour les bas que pour les hauts niveaux d'éducation.
- Cette restriction est discutable $\rightarrow$ on étendra ce modèle plus tard (voir Section 2.4).

Modèle de régression simple

- Pour estimer β_0 et β_1 et garder cette interprétation *Ceteris Paribus* il faut faire certaines hypothèses.
- 1) $E(u) = 0$: normalisation $\rightarrow$ on ne perd rien.
- 2) $E(u|x) = E(u)$: pour toute valeur de x, la moyenne des u correspondantes est la même.
  $\rightarrow$ implique la non-corrélation (linéaire). Ex:
  $wage = \beta_0 + \beta_1 educ + u \rightarrow E(abil|8) = E(abil|16)$.
- En combinant ces 2 hypothèses:
  $\rightarrow E(u|x) = E(u) = 0$: hypothèse de moyenne conditionnelle nulle.
  $\rightarrow E(y|x) = \beta_0 + \beta_1 x$: fonction de régression de la population.

Estimation de β_0 et β_1

$(x_i, y_i) : i = 1, \ldots, n =$ échantillon aléatoire de taille n tiré de la population, c-à-d du modèle $y = \beta_0 + \beta_1 x + u$.

- $\Rightarrow y_i = \beta_0 + \beta_1 x_i + u_i$, $\forall i$.
- $E(u) = 0 \rightarrow E(y - \beta_0 - \beta_1 x) = 0$.
- $E(u|x) = 0 \rightarrow$ Covariance nulle entre u et x
  $\rightarrow E[(y - \beta_0 - \beta_1 x)x] = 0$.
- $\Rightarrow$ on a deux restrictions sur la distribution jointe de (x, y) dans la population.
- Pour obtenir $\hat{\beta}_0$ et $\hat{\beta}_1$ on va résoudre ce système en remplaçant $E(.)$ par son équivalant empirique $1/n \sum_{i=1}^{n}(.)$.

Estimation de β_0 et β_1

$$n^{-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\hookrightarrow \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$n^{-1}\sum_{i=1}^{n}[(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i] = 0$$

$$\Leftrightarrow \quad \sum_{i=1}^{n}[x_i(y_i - \overline{y} + \hat{\beta}_1 \overline{x} - \hat{\beta}_1 x_i)] = 0$$

$$\Leftrightarrow \quad \underbrace{\sum_{i=1}^{n} x_i(y_i - \overline{y})}_{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})} = \hat{\beta}_1 \underbrace{\sum_{i=1}^{n} x_i(x_i - \overline{x})}_{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

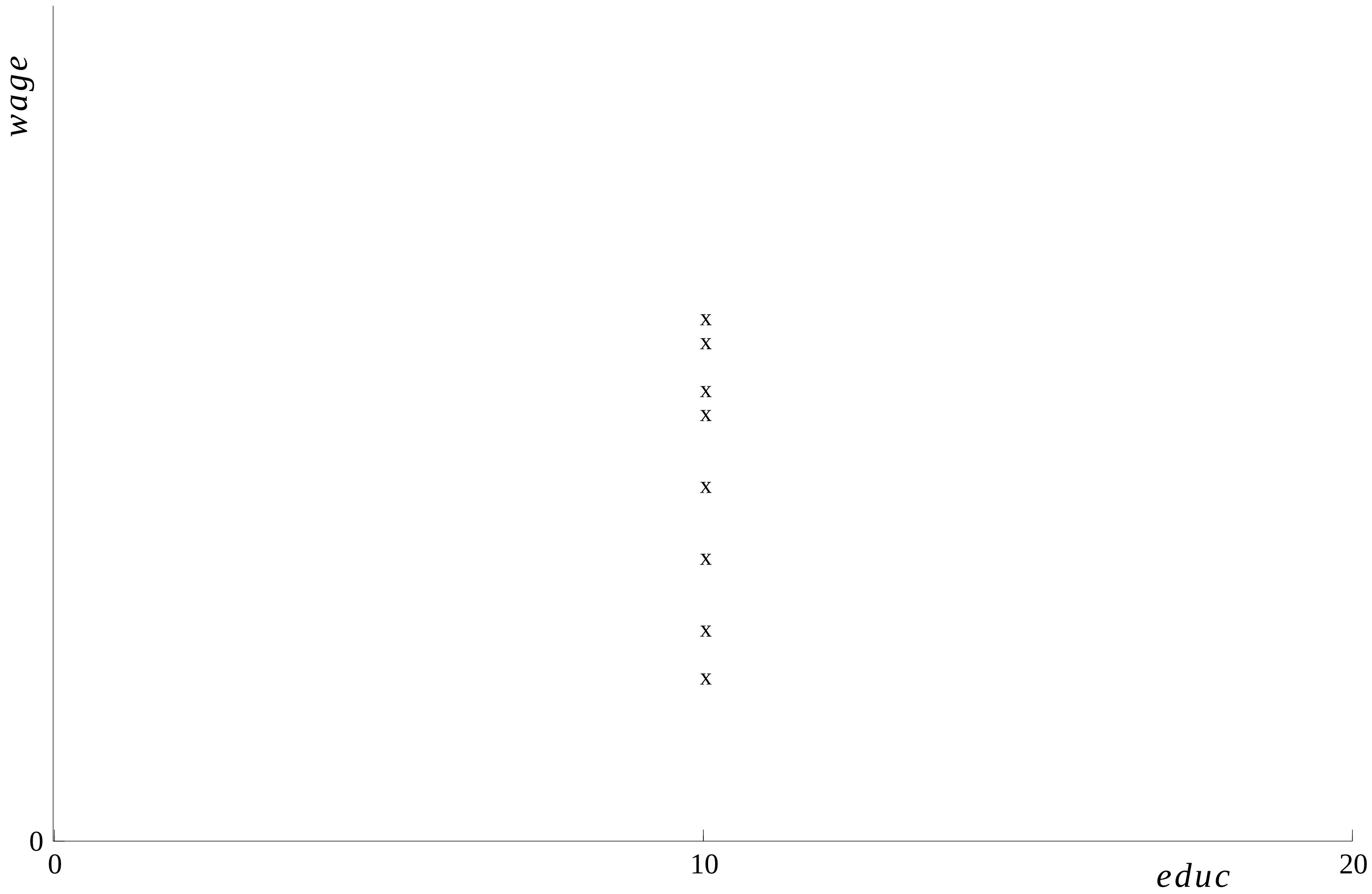
Estimation de β_0 et β_1

- Donc pour autant que $\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0$,

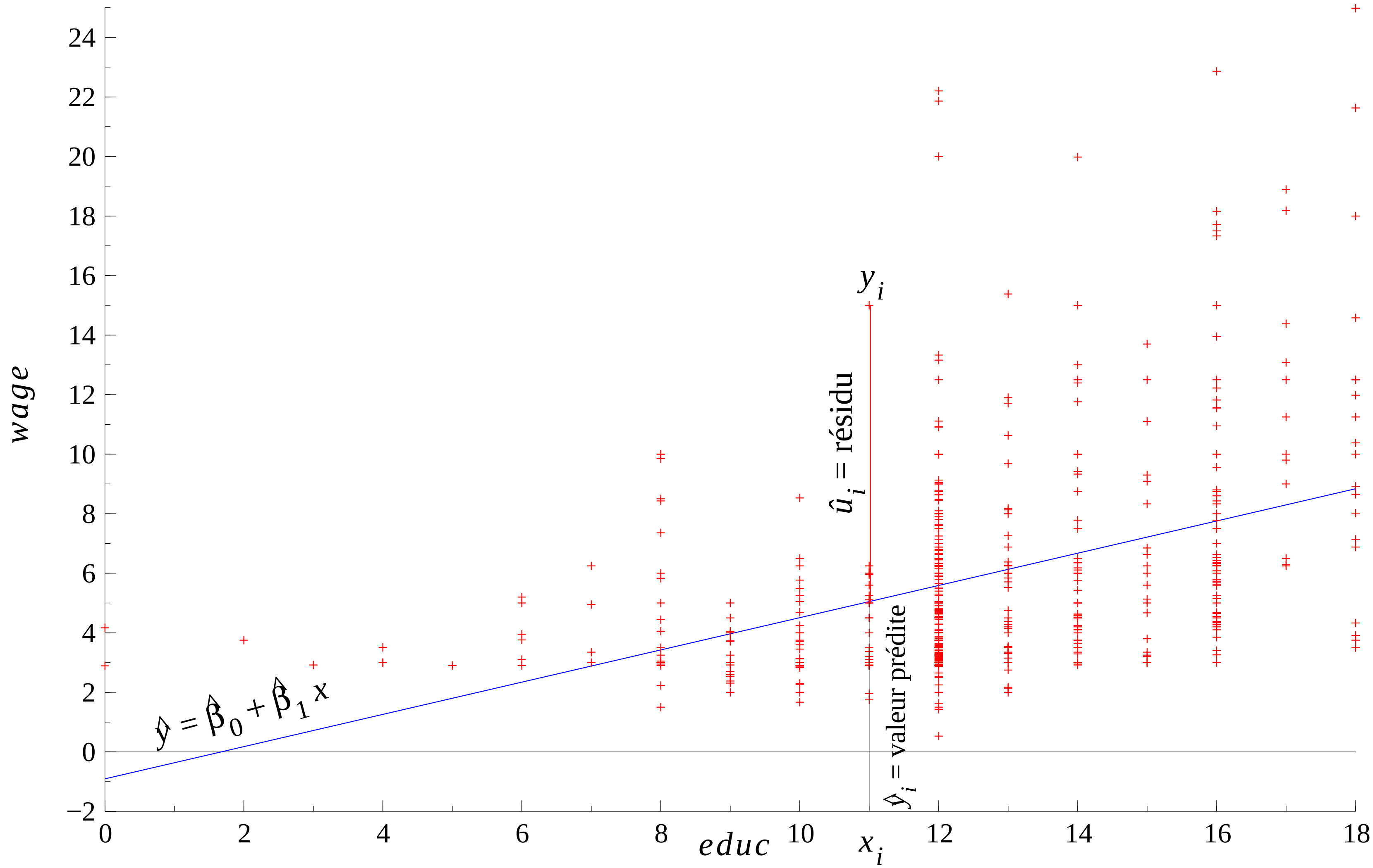
$$\hookrightarrow \hat{\beta}_1 \quad = \quad \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

- $\hat{\beta}_1 = \dfrac{\text{Covariance empirique entre} x \text{ et } y}{\text{Variance empirique de} x}.$
- $\rightarrow \hat{\beta}_1 > 0 (< 0)$ quand x et y sont positivement (négativement) corrélés dans l'échantillon.
- Cette méthode d'estimation de β_0 et β_1 s'appelle la méthode des moments.

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0$$



Moindres carrés ordinaires (MCO)



Moindres carrés ordinaires (MCO)

Trouver les valeurs de $\hat{\beta}_0$ et $\hat{\beta}_1$ qui minimisent le problème suivant:

$$\begin{aligned} & \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2 \\ \Leftrightarrow \quad & \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{aligned}$$

$\rightarrow$ 2 équations “de premier ordre”

Moindres carrés ordinaires (MCO)

$$
\begin{aligned}
-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 \quad &- \quad \hat{\beta}_1 x_i) = 0 \\
-2\sum_{i=1}^{n}[(y_i - \hat{\beta}_0 \quad &- \quad \hat{\beta}_1 x_i)x_i] = 0 \\
\hookrightarrow \hat{\beta}_0 \quad &= \quad \overline{y} - \hat{\beta}_1 \overline{x} \\
\hookrightarrow \hat{\beta}_1 \quad &= \quad \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2 \text{ si } > 0}.
\end{aligned}
$$

Estimation de β_0 et β_1 par la méthode des MCO donne les même $\hat{\beta}_0$ et $\hat{\beta}_1$.

Moindres carrés ordinaires (MCO)

- On a $min \sum_{i=1}^{n} \hat{u}_i^2$. On pourrait également $min \sum_{i=1}^{n} |\hat{u}_i|$.
- $\longrightarrow$ Ca compliquerait beaucoup les calculs. Obtenir des formules pour $\hat{\beta}_0$ et $\hat{\beta}_1$ serait difficile ($\hat{\beta}_0$ et $\hat{\beta}_1$ peuvent être obtenus par des techniques d'optimisation numérique).
- Une fois les paramètres estimés, on peut tracer la droite de régression $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ (cf graphique précédant).
- $\hat{\beta}_0$ est la valeur prédite de y (c-à-d $\hat{y}$) pour une valeur de $x = 0 \longrightarrow$ pas toujours interprétable.
- Par contre si on le retire $\longrightarrow \hat{\beta}_0 = 0$. DISCUTABLE.
- $\hat{\beta}_1$ est bien souvent plus intéressant. $\hat{\beta}_1 = \frac{\Delta \hat{y}}{\Delta x} =$ par combien $\uparrow \hat{y}$ quand $x \uparrow$ d'une unité.

Exemples

- Modèle expliquant le salaire des PDGs:
  $salary = \beta_0 + \beta_1 roe + u$,
- où $salary$ est le salaire annuel (en 10.000 de $) des PDGs et roe le rendement moyen des actions des firmes relatives à ces PDGs (sur les 3 dernières années). 209 observations (année 1990) $\rightarrow$ données cross-section.
- Exemple: CEOSAL1.xls.
- Exemple: CEOSAL1.in7.

Statistiques descriptives

	Salary	roe
Observations	209	209
Mean	1281.1	17.184
Std.Devn.	1369.1	8.4981
Minimum	223.00	0.5000
Maximum	14822.	56.300

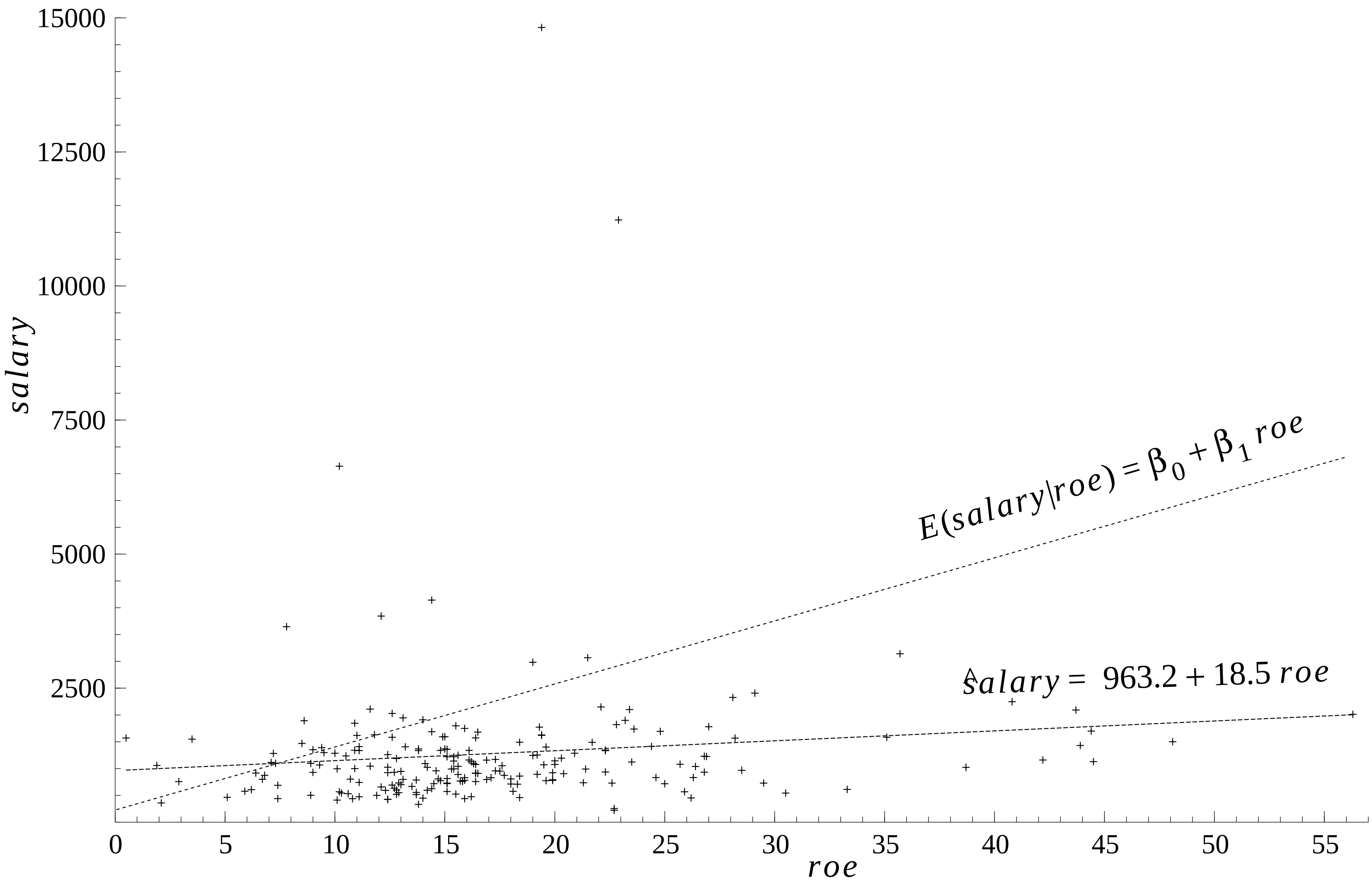
Correlation matrix:

	salary	roe
salary	1.0000	0.11484
roe	0.11484	1.0000

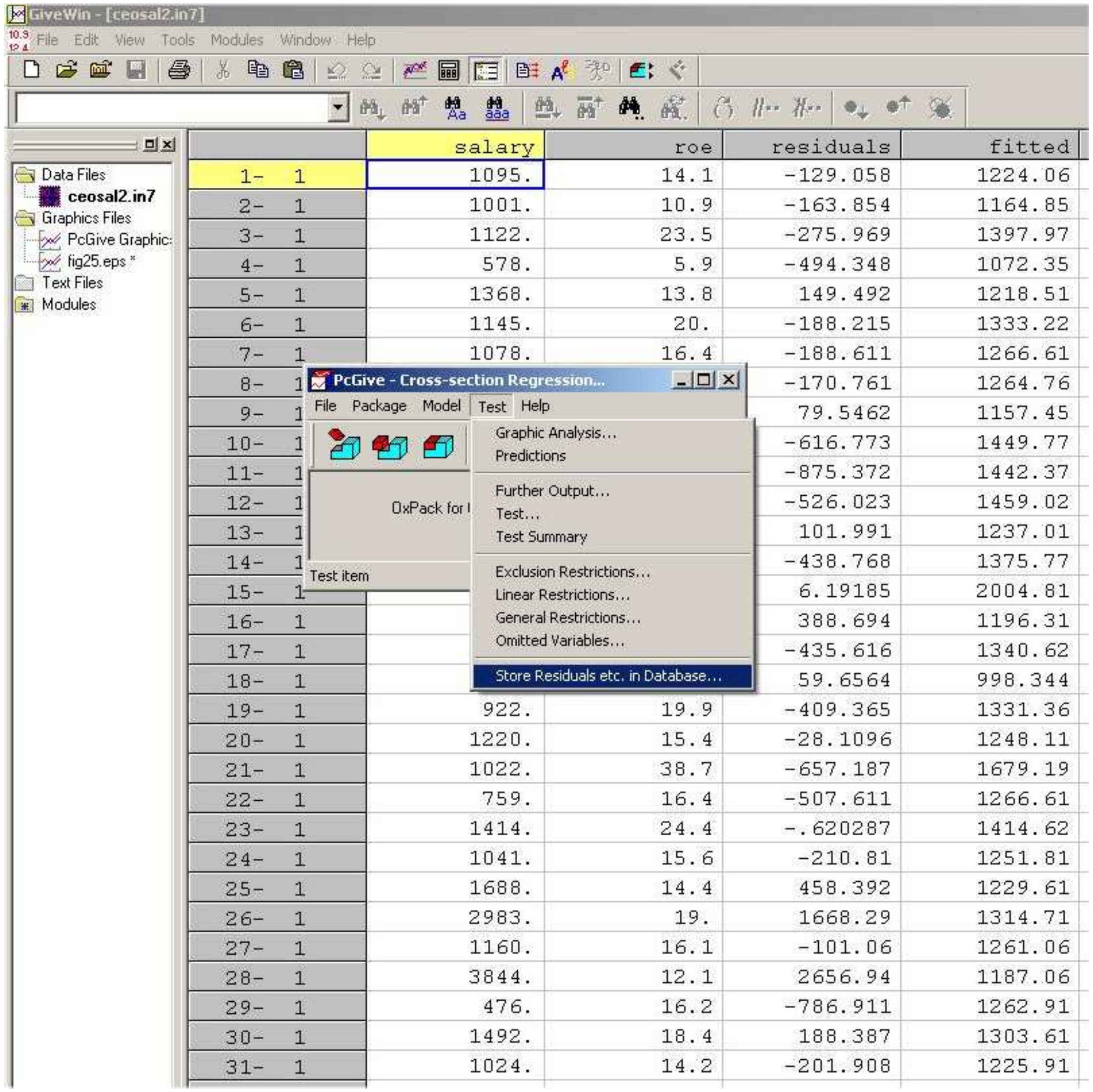
Régresser $salary$ sur roe

```
EQ( 1) Modelling salary by OLS-CS (using ceosal1.in7)
       The estimation sample is: 1 to 209
                 Coefficient  Std.Error  t-value  t-prob
Constant            963.191      213.2      4.52   0.000
roe                 18.5012      11.12      1.66   0.098

sigma                 1366.55  RSS                386566563
R^2                 0.0131886  F(1,207) =   2.767 [0.098]
log-likelihood       -1804.54  DW                       2.1
no. of observations       209  no. of parameters          2
mean(salary)          1281.12  var(salary)   1.87432e+006

    salary =  + 963.2 + 18.5*roe
    (SE)        (213)   (11.1)
```

$$\hat{salary} = 963.2 + 18.5roe$$



$\hat{y}_i$ et $\hat{u}_i$

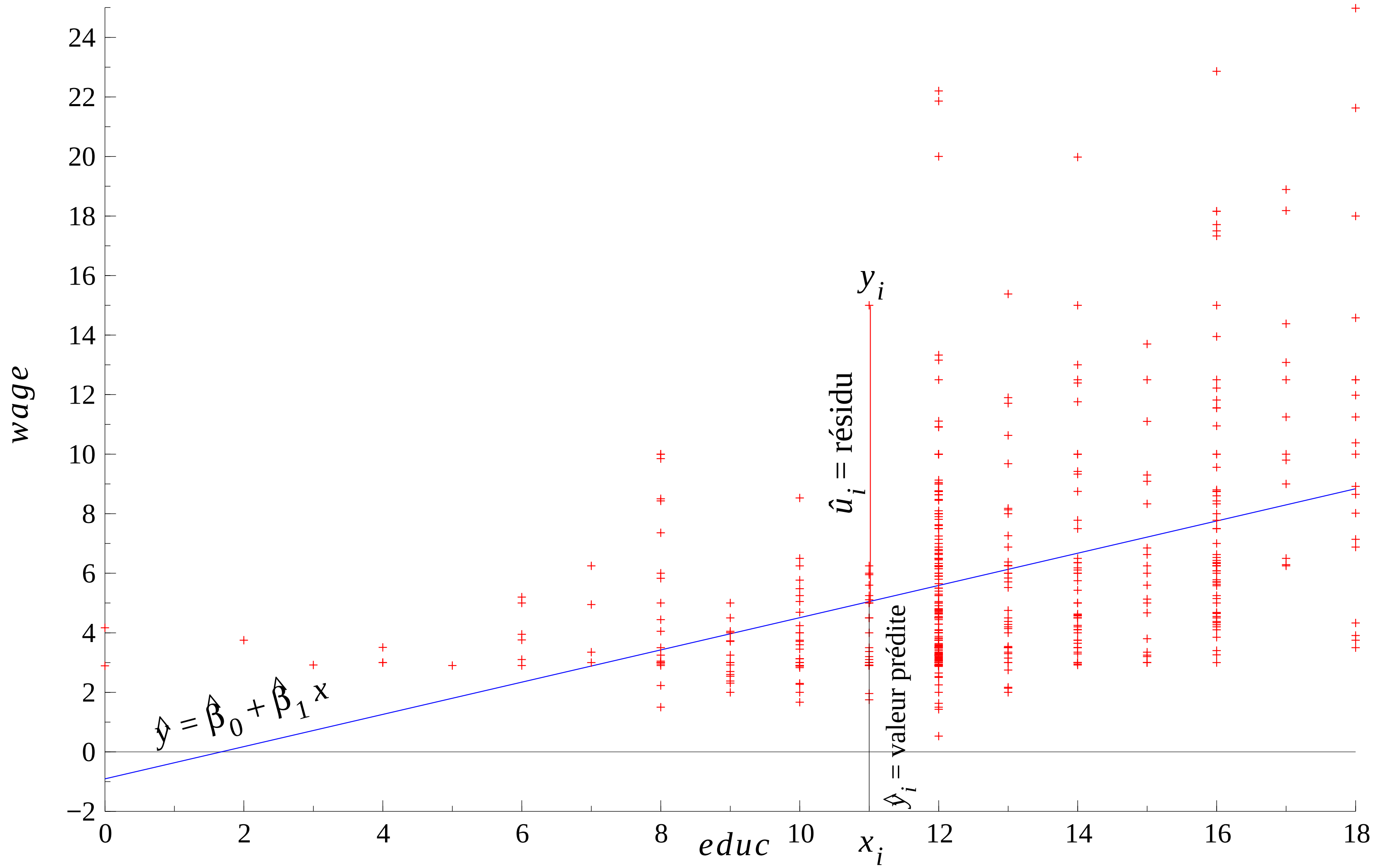


	salary	roe	residuals	fitted
1- 1	1095.	14.1	-129.058	1224.06
2- 1	1001.	10.9	-163.854	1164.85
3- 1	1122.	23.5	-275.969	1397.97
4- 1	578.	5.9	-494.348	1072.35
5- 1	1368.	13.8	149.492	1218.51
6- 1	1145.	20.	-188.215	1333.22
7- 1	1078.	16.4	-188.611	1266.61
8- 1			-170.761	1264.76
9- 1			79.5462	1157.45
10- 1			-616.773	1449.77
11- 1			-875.372	1442.37
12- 1			-526.023	1459.02
13- 1			101.991	1237.01
14- 1			-438.768	1375.77
15- 1			6.19185	2004.81
16- 1			388.694	1196.31
17- 1			-435.616	1340.62
18- 1			59.6564	998.344
19- 1	922.	19.9	-409.365	1331.36
20- 1	1220.	15.4	-28.1096	1248.11
21- 1	1022.	38.7	-657.187	1679.19
22- 1	759.	16.4	-507.611	1266.61
23- 1	1414.	24.4	-.620287	1414.62
24- 1	1041.	15.6	-210.81	1251.81
25- 1	1688.	14.4	458.392	1229.61
26- 1	2983.	19.	1668.29	1314.71
27- 1	1160.	16.1	-101.06	1261.06
28- 1	3844.	12.1	2656.94	1187.06
29- 1	476.	16.2	-786.911	1262.91
30- 1	1492.	18.4	188.387	1303.61
31- 1	1024.	14.2	-201.908	1225.91

Quelques propriétés alg. des MCO

- 1) $\sum_{i=1}^{n} \hat{u}_i = 0 \leftarrow$ première condition de premier ordre: $n^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$.
- 2) $\sum_{i=1}^{n} x_i \hat{u}_i = 0 \leftarrow$ deuxième condition de premier ordre: $n^{-1} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$.
- 3) $(\overline{x}, \overline{y})$ est toujours sur la droite de régression car on a vu que $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$.
- 4) $y_i = \hat{y}_i + \hat{u}_i \rightarrow$ les MCO décomposent chaque y_i en deux parties: une valeur prédite $+$ un résidu (valeur non-prédite).

Moindres carrés ordinaires (MCO)



Décomposition de la variance

- Comme $y_i = \hat{y}_i + \hat{u}_i$, on peut définir :

$$SST \equiv \sum_{i=1}^{n}(y_i - \overline{y})^2 = \text{ Total Sum of Squares.}$$

$$SSE \equiv \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = \text{ Explained Sum of Squares.}$$

$$SSR \equiv \sum_{i=1}^{n}\hat{u}_i^2 = \text{ Residual Sum of Squares.}$$

- SSE est parfois également appelé Regression Sum of Squares.
- $\Rightarrow SST = SSE + SSR$.

Décomposition de la variance

Preuve:

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\overline{y})^2 &= \sum_{i=1}^{n}\left[(y_i-\hat{y}_i)+(\hat{y}_i-\overline{y})\right]^2 \\
&= \sum_{i=1}^{n}\left[\hat{u}_i^2+(\hat{y}_i-\overline{y})\right]^2 \\
&= \sum_{i=1}^{n}\hat{u}_i^2+2\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i-\overline{y})+\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2 \\
&= SSR+2\underbrace{\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i-\overline{y})}_{(n-1)Cov(\hat{u}_i,\hat{y})=0}+SSE.\text{CQFD.}
\end{aligned}
$$

Goodness-of-fit (GoF)

- Pour autant :
  - qu'un intercept a été inclus dans la régression (β_0);
  - et qu'il y a de la variabilité dans les y_i,
  $\rightarrow$ on peut calculer le *R-carré* de la régression (ou coefficient de détermination).
- $R^2 \equiv SSE/SST = 1 - SSR/SST$.
- $\rightarrow 100 \times R^2$ est le pourcentage de la variance de y expliqué par x.
- $0 \leq R^2 \leq 1 \rightarrow 1$ si tous les points sont sur la droite de régression.
- $R^2 = \rho^2_{y_i,\hat{y}_i}$.
- Ne pas juger la qualité d'une régression uniquement sur le R^2.

R^2

```
        The estimation sample is: 1 to 209
                 Coefficient  Std.Error  t-value  t-prob
Constant           963.191      213.2      4.52    0.000
roe                18.5012      11.12      1.66    0.098

sigma                  1366.55  RSS                  386566563
R^2                  0.0131886  F(1,207) =   2.767 [0.098]
log-likelihood        -1804.54  DW                         2.1
no. of observations        209  no. of parameters            2
mean(salary)           1281.12  var(salary)    1.87432e+006
```

$$\rightarrow 1 - 386566563/(209 * 1874320) = 0.01318$$

Unités de mesure

- Dans l'exemple précédant $salary$ est exprimé en $10000\$$ et roe en %.
- $salary = 100 \longrightarrow 100 * 10000\$ = 1000000\$$ et $roe = 10 \longrightarrow 10\%$.
- Si on défini maintenant $salardol = 10000 \times salary$ on obtiendra:
  $salardol = 963191 + 18501 * roe$.
- Si $y \rightarrow c \times y \Rightarrow \hat{\beta}_0$ et $\hat{\beta}_0 \rightarrow c \times \hat{\beta}_0$ et $c \times \hat{\beta}_1$.
- Si $x \rightarrow c \times x \Rightarrow \hat{\beta}_1 \rightarrow c^{-1}\hat{\beta}_1$.
- Si $roedec = roe/100 \rightarrow salary = 963.191 + 1850.1 * roedec$.
- Le R^2 n'est pas affecté par ces changements d'unité de mesure.

Forme fonctionnelle

- Relation linéaire:

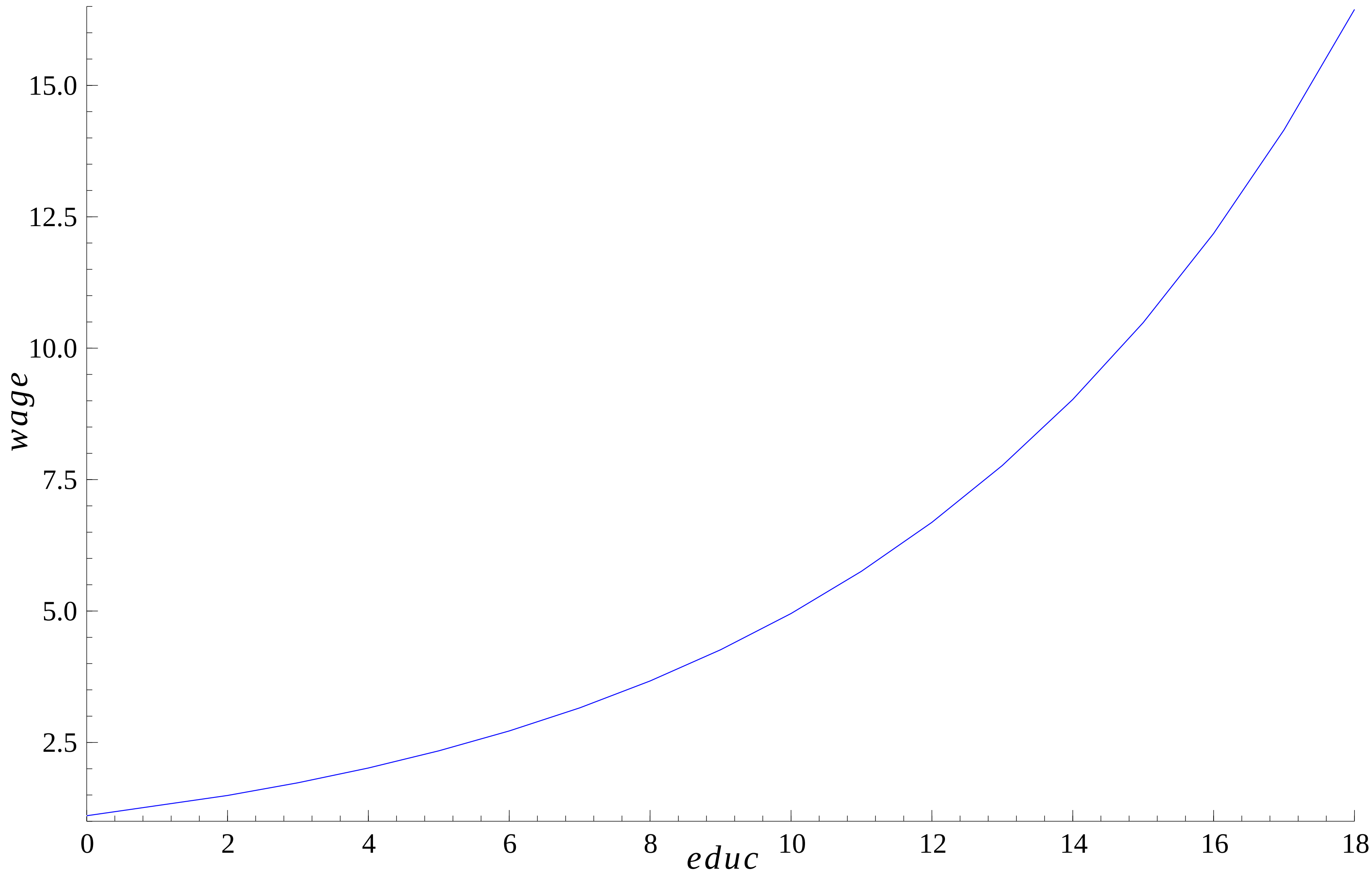
$$wage = \beta_0 + \beta_1 educ + u.$$

- $\hat{\beta}_1 = 0.54 \rightarrow$ chaque année supplémentaire devrait accroître le salaire horaire ($wage$) de 54 cents, quel que soit le niveau d'éducation.
- On suppose ici que l'effet absolu de l'éducation ne dépend pas du niveau d'éducation $\rightarrow$ irréaliste.
- En économie du travail, on voit souvent ce type de modèle:

$$log(wage) = \beta_0 + \beta_1 educ + u,$$

où $log(.)$ est le $ln(.)$ et non le log en base 10.
- $\rightarrow$ si $\Delta u = 0$, $\%\Delta wage \approx (100 \times \beta_1)\Delta educ$.

$log(y)$



$$wage = exp(\beta_0 + \beta_1 educ)$$, avec $\beta_1 > 0$.

Exemple

```
        The estimation sample is: 1 to 526
                               LWAGE        WAGE
Constant                     0.583773   -0.904852
EDUC                        0.0827444    0.541359

R^2                          0.185806    0.164758
```

$\rightarrow$ Le salaire ($wage$) augmente de 8.3 % pour chaque année d'éducation supplémentaire $(\Delta educ = 1)$.

Modèle à élasticité constante

- Dans certains cas, il est intéressant d'estimer un modèle $log - log$.
- $log(y) = \beta_0 + \beta_1 log(x) + u.$
- C'est toujours un modèle linéaire en les paramètres. Contrairement à $y = 1/(\beta_0 + \beta_1 x) + u.$
- Exemple: CEOSAL1.in7.
- $lsalary = 4.822 + 0.2567 lsales$
  $n = 209, R^2 = 0.211$.
- L'élasticité salaire-ventes est de 0.257 $\rightarrow$ une augmentation de 1% des ventes augmente le salaire d'environ 0.257%.

Changement d'unité et log

- $log(y_i) = \beta_0 + \beta_1 x_i + u_i.$
- $y_i \rightarrow cy_i \Rightarrow log(y_i) \rightarrow log(c) + log(y_i)$.
- $\rightarrow log(c) + log(y_i) = log(c) + \beta_0 + \beta_1 x_i + u_i.$
- $\rightarrow \beta_1$ est inchangé mais l'intercept est $log(c) + \beta_0.$
- Idem si on change l'unité de x.

Combinaisons de log

Modèle	Variable dépendante	Variable explicative	Interprétation de β_1
niveau-niveau	y	x	$\Delta y = \beta_1 \Delta x$
niveau-log	y	$log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
log-niveau	$log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
log-log	$log(y)$	$log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Propriétés des estimateurs MCO

SLR=Simple Linear Regression.

- Quelles sont les propriétés statistiques de l'estimateur des MCO ?
- $\hat{\beta}_0$ et $\hat{\beta}_1$ sont des estimateurs des paramètres de la population β_0 et β_1.
- Ce sont donc des variables aléatoires car on obtiendra des $\hat{\beta}_0$ et $\hat{\beta}_1$ différents si on utilise des échantillons différents (tirés de la même population).
- $\rightarrow$ on va devoir imposer certaines hypothèses pour étudier:
- 1. $E(\hat{\beta}_0)$ et $E(\hat{\beta}_1)$.
- 2. $Var(\hat{\beta}_0)$ et $Var(\hat{\beta}_1)$.

Caractère non biaisé des MCO

- **SLR.1** $y = \beta_0 + \beta_1 x + u \rightarrow$ linéaire en les paramètres.
- **SLR.2** $(x_i, y_i) : i = 1, \ldots, n \rightarrow$ échantillon aléatoire de taille n tiré de la population.
  $\rightarrow y_i = \beta_0 + \beta_1 x_i + u_i, i = 1, 2, \ldots, n.$
- **SLR.3** $E(u|x) = 0 \rightarrow$ moyenne conditionnelle nulle. Permet de dériver les propriétés des MCO *conditionnellement* aux valeurs de x_i dans notre échantillon. Techniquement identique à supposer x_i *fixes dans des échantillons répétés* (pas très réaliste).
- **SLR.4** $\sum_{i=1}^{n}(x_i - \overline{x})^2 \neq 0 \rightarrow$ variation dans les x.
- **Théorème 2.1**: Sous les hypothèses SLR.1 - SLR.4, $E(\hat{\beta}_0) = \beta_0$ et $E(\hat{\beta}_1) = \beta_1$.

Preuve Théorème 2.1

- Rappelons que si **SLR.4** est vérifié,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{SST_x}.$$

- Propriété de $\hat{\beta}_1$ à travers un ensemble infini d'échantillons différents ? $\rightarrow$ réécrire $\hat{\beta}_1$ en fonction de des paramètres de la population (et du terme d'erreur).
- $\rightarrow \hat{\beta}_1 == \frac{\sum_{i=1}^{n}(x_i - \overline{x})(\beta_0 + \beta_1 x_i + u_i)}{SST_x}.$

Preuve suite

- Le numérateur peut être réécrit comme suit:

$$\sum_{i=1}^{n}(x_i - \overline{x})\beta_0 + \sum_{i=1}^{n}(x_i - \overline{x})\beta_1 x_i + \sum_{i=1}^{n}(x_i - \overline{x})u_i$$

$$= \underbrace{\beta_0 \sum_{i=1}^{n}(x_i - \overline{x})}_{0} + \beta_1 \underbrace{\sum_{i=1}^{n}(x_i - \overline{x})x_i}_{\sum_{i=1}^{n}(x_i - \overline{x})^2 = SST_x} + \sum_{i=1}^{n}(x_i - \overline{x})u_i.$$

- Par conséquent,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{SST_x} = \beta_1 + \frac{\sum_{i=1}^{n} d_i u_i}{SST_x}.$$

- $E(\hat{\beta}_1) = \beta_1$?

Preuve suite

- Notons tout d'abord que SST_x et d_i ne dépendent que de $x_1, \ldots, x_n \rightarrow E(SST_x | x_1, \ldots, x_n) = SST_x$ (voir propriété CE.1 page 718).
- Donc,

$$\begin{aligned} E(\hat{\beta}_1 | x_1, \ldots, x_n) &= \beta_1 + \frac{1}{SST_x} \sum_{i=1}^{n} d_i \underbrace{E(u_i | x_1, \ldots, x_n)}_{0:\textbf{SLR.2-3}} \\ &= \beta_1. \end{aligned}$$

- Hors, si $E(\hat{\beta}_1 | x_1, \ldots, x_n) = \beta_1$
$\rightarrow \underbrace{E[E(\hat{\beta}_1 | x_1, \ldots, x_n)] = E(\hat{\beta}_1)}_{\text{voir propriété CE.4 page 719}} = \beta_1$.

Preuve suite et fin

- On a également que
  $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = \beta_0 + \beta_1\overline{x} + \overline{u} - \hat{\beta}_1\overline{x} = \beta_0 + (\beta_1 - \hat{\beta}_1)\overline{x} + \overline{u}$.
- $\rightarrow E(\hat{\beta}_0|x_1, \ldots, x_n) =$
  $\beta_0 + E[\underbrace{(\beta_1 - \hat{\beta}_1)}_{0}\overline{x}|x_1, \ldots, x_n] + \underbrace{E(\overline{u}|x_1, \ldots, x_n)}_{0:\textbf{SLR.2-3}}$.
- $\rightarrow E(\hat{\beta}_0) = E[E(\hat{\beta}_0|x_1, \ldots, x_n)] = \beta_0$.
- CQFD.
- Attention: le résultat ne dit rien sur l'estimateur obtenu sur un échantillon particulier $\rightarrow$ résultat “moyen”.
- Si **SLR.1-3** sont violées $\rightarrow$ estimateur biaisé. Si **SLR.4** violée $\rightarrow$ on ne peut pas calculer $\hat{\beta}_0$ et $\hat{\beta}_1$.

Importance de SLR.3

- Pour rappel: **SLR.3** $E(u|x) = 0 \rightarrow$ moyenne conditionnelle nulle.
- Exemple: $math10 = \beta_0 + \beta_1 lnchprg + u$, où $math10 = \%$age d'étudiants qui réussissent un examen de math standardisé et $lnchprg = \%$age d'étudiants qui sont éligibles à un programme fédéral de financement d'un repas de midi.
- $\hat{math}10 = 32.14 - 0.32 lnchprg \; (n = 408, R^2 = 0.171)$.
- $\rightarrow$ si $lnchprg$ augmente de 10%, $\hat{math}10$ diminue de 3.2%.
- $E(u|x) = 0$? $u \supset$ taux de pauvreté, qualité d'enseignement $\rightarrow$ corrélés avec $x : lnchprg$.

Variance de $\hat{\beta}_0$ et $\hat{\beta}_1$?

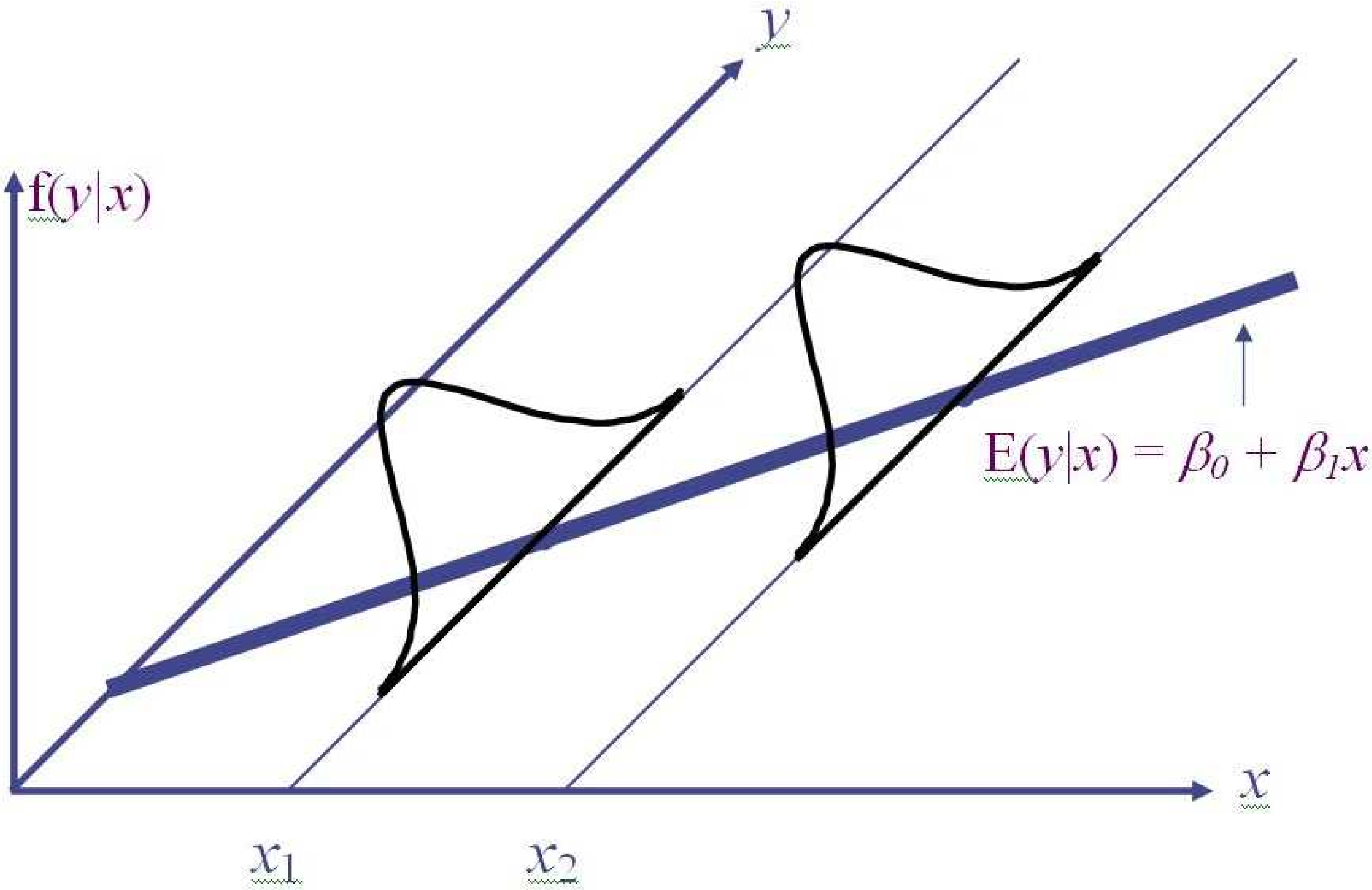
Mesure de dispersion de l'estimateur MCO ?

- **SLR.5** $Var(u|x) = \sigma^2 \rightarrow$ homoscédasticité.
- **Théorème 2.2**: Sous les hypothèses SLR.1 - SLR.5,

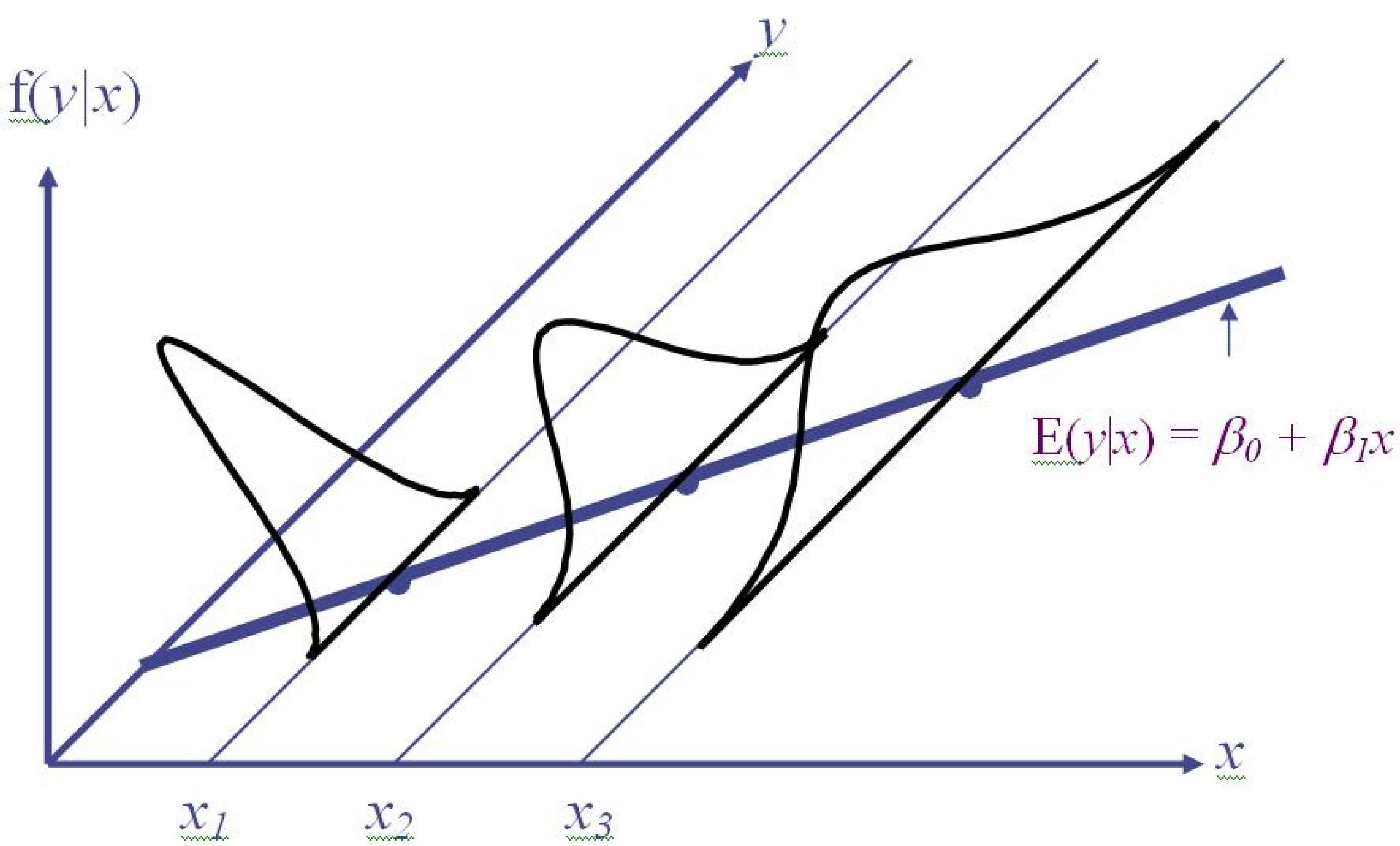
$$Var(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \overline{x})^2}$$
$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

- **Théorème 2.3**: Sous les hypothèses SLR.1 - SLR.5, $E(\hat{\sigma}^2) = \sigma^2$, où $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2$.
- $\rightarrow$ **SLR.5** ne joue aucun rôle pour prouver le caractère non-biaisé des MCO.

SLR avec homoscédasticité



SLR avec hétéroscédasticité



Preuve pour $\hat{\beta}_1$

- $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n d_i u_i}{SST_x}$.
- Comme β_1 est une cst et SST_x n'est fonction que de x,

$$
\begin{aligned}
Var(\hat{\beta}_1 | x_1, \ldots, x_n) &= \frac{Var\left(\sum_{i=1}^n d_i u_i | x_1, \ldots, x_n\right)}{SST_x^2} \\
&= \frac{\sum_{i=1}^n d_i^2 Var(u_i | x_1, \ldots, x_n)}{SST_x^2} \\
\textbf{(SLR.5)} &= \frac{\sum_{i=1}^n d_i^2 \sigma^2}{SST_x^2} = \frac{\sigma^2 \left(\sum_{i=1}^n d_i^2\right)}{SST_x^2} \\
&= \frac{\sigma^2 SST_x}{SST_x^2} = \frac{\sigma^2}{SST_x}.
\end{aligned}
$$

- $Var(\hat{\beta}_1 | x_1, \ldots, x_n) = Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$. CQFD.

$Var(\hat{\beta}_1) = \sigma^2/SST_x$?

- Plus il y a de la variation dans la partie non-prédite, c-à-d σ^2, plus β_1 est estimé de manière imprécise.
- Plus il y a de la variation dans la variable explicative, c-à-d SST_x, plus β_1 est estimé de manière précise. Comme $SST_x = \sum_{i=1}^{n}(x_i - \overline{x})^2$, $SST_x \uparrow$ avec n.
- On appelle $sd(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$ l'écart-type de $\hat{\beta}_1$.
- σ^2 est non observé $\rightarrow$ on doit l'estimer.

Estimation de σ^2

- σ^2 est la variance de u, c-à-d la variance du terme d'erreur de la population $u_i = y_i - \beta_0 - \beta_1 x_i \rightarrow$ jamais observé.
- Ce que l'on observe après estimation c'est $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$.
- Ou encore,

$$\begin{aligned}\hat{u}_i &= (\beta_0 + \beta_1 x_i + u_i) - \hat{\beta}_0 - \hat{\beta}_1 x_i \\ &= u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i.\end{aligned}$$

- Comme $\sigma^2 = E(u^2)$, un estimateur non-biaisé de $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2$. Mais u_i n'est pas observé.
- On serait tenté de replacer u_i par $\hat{u}_i \rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 = SSR/n$.

Estimation de σ^2

- Mais $E(SSR/n) \neq \sigma^2 \rightarrow SSR/n$ est un estimateur biaisé.
- Ceci vient du fait que pour calculer $\hat{u}_i$ par MCO (c-à-d $\hat{\beta}_0$ et $\hat{\beta}_1$) on a utilisé deux restrictions.
- $\sum_{i=1}^{n} \hat{u}_i = 0$ et $\sum_{i=1}^{n} x_i \hat{u}_i = 0 \rightarrow$ conditions de moments.
$\hat{u}_i$	x_i	$x_i\hat{u}_i$
-1	4	-4
2	6	12
-1	8	-8
0		0

  $\rightarrow$ 2 valeurs de $\hat{u}_i$ sont contraintes.
- On a donc $n-2$ degrés de libertés.

$\hat{\sigma}^2$

$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2 = SSR/(n-2)$ est un estimateur non-biaisé de $\sigma^2 \rightarrow E(\hat{\sigma}^2) = \sigma^2$.

Preuve:

$$\begin{aligned}
\hat{u}_i &= u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i \\
\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0 &= \overline{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)\overline{x} \\
\Rightarrow \hat{u}_i - 0 &= (\hat{u}_i - \overline{u}) - (\hat{\beta}_1 - \beta_1)(x_i - \overline{x}) \\
\Leftrightarrow \hat{u}_i^2 &= (\hat{u}_i - \overline{u})^2 + (\hat{\beta}_1 - \beta_1)^2(x_i - \overline{x})^2 \\
&- 2(\hat{u}_i - \overline{u})(\hat{\beta}_1 - \beta_1)(x_i - \overline{x})
\end{aligned}$$

◇

Suite de la preuve

$$
\begin{aligned}
\Leftrightarrow \sum_{i=1}^{n} \hat{u}_i^2 &= \sum_{i=1}^{n} (\hat{u}_i - \overline{u})^2 + (\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^{n} (x_i - \overline{x})^2 \\
&- 2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^{n} \hat{u}_i (x_i - \overline{x}) \\
\Rightarrow E\left(\sum_{i=1}^{n} \hat{u}_i^2\right) &= \underbrace{(n-1)\sigma^2}_{\text{C.5 p. 736}} + \underbrace{\sigma^2}_{E[(\hat{\beta}_1 - \beta_1)^2] = Var(\hat{\beta}_1) = \sigma^2 / SST_x} \\
&- \underbrace{2\sigma^2}_{2SST_x E[(\hat{\beta}_1 - \beta_1)^2]} = (n-2)\sigma^2 \\
\Rightarrow E\left(\frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2\right) &= E\left(\hat{\sigma}^2\right) = \sigma^2 \text{ CQFD.}
\end{aligned}
$$

Chapitre 3: Modèle de régression multiple

Modèle de régression multiple

- Objectif: estimer un modèle du type

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u.$$

- De manière générale:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u.$$

- β_0 est l'intercept et les $\beta_i (i = 1, \ldots, k)$ mesurent les changements de y respectivement à x_i, toutes choses restant égales par ailleurs.
- Notons que cette spécification permet les modèles du type:

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u.$$

$\rightarrow \frac{\Delta cons}{\Delta inc} \approx \beta_1 + 2\beta_2 inc$.

Mécanisme et interprétation des MCO

Population: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$.

- MCO: $min_{\hat\beta_0,\hat\beta_1,\ldots,\hat\beta_k} \sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 x_{i1} - \ldots - \hat\beta_k x_{ik})^2$.
- $\rightarrow k+1$ équations linéaires en $k+1$ paramètres inconnus $\rightarrow E(u)=0$ et $E(x_j u) = 0\ \forall i = 1,\ldots,k$:

$$\sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 x_{i1} - \ldots - \hat\beta_k x_{ik}) = 0$$

$$\sum_{i=1}^{n}[x_{i1}(y_i - \hat\beta_0 - \hat\beta_1 x_{i1} - \ldots - \hat\beta_k x_{ik})] = 0$$

$$\vdots$$

$$\sum_{i=1}^{n}[x_{ik}(y_i - \hat\beta_0 - \hat\beta_1 x_{i1} - \ldots - \hat\beta_k x_{ik})] = 0.$$

Interprétation

Population: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$.

- $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$.
- $\hat{\beta}_0$ est la valeur prédite de y quand $x_1 = x_2 = 0 \rightarrow$ pas toujours de sens bien que nécessaire.
- $\hat{\beta}_1$ a une interprétation d'effet partiel ou *Ceteris Paribus*.
- $\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2$.
- $\rightarrow$ quand $\Delta x_2 = 0, \Delta\hat{y} = \hat{\beta}_1 \Delta x_1$.
- $\rightarrow$ quand $\Delta x_1 = 0, \Delta\hat{y} = \hat{\beta}_2 \Delta x_2$.
- Exemple: Wage1.

Exemple: LWAGE ($n = 526$)

	Coefficient
Constant	0.284360
EDUC	0.0920290
EXPER	0.00412111
TENURE	0.0220672

$$\begin{aligned} \text{LWAGE} &= 0.2844 + 0.09203\text{EDUC} + 0.004121\text{EXPER} \\ &+ 0.02207\text{TENURE}. \end{aligned}$$

$\rightarrow$ *Ceteris Paribus*, une année supplémentaire d'éducation pourrait accroître $log(w\hat{a}ge)$ de 0.092 et donc environ le salaire 9.2% (0.092×100) ssi l'interprétation *Ceteris Paribus* tient.

$\rightarrow$ si $exper$ et $tenure \uparrow$ de 1 an et $educ$ reste inchangé.

$\Delta log(w\hat{a}ge) = 0.004121\Delta exper + 0.02207\Delta tenure = 0.0261$

Quelques propriétés alg. des MCO

- $\hat{u}_i = y_i - \hat{y}_i$.
- 1) $\sum_{i=1}^n \hat{u}_i = 0 \leftarrow$ condition de premier ordre première: $\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$.
- 2) $\forall i = 1, \ldots, k : \sum_{i=1}^n x_{ij} \hat{u}_i = 0 \leftarrow$ condition de premier ordre première: $\sum_{i=1}^n x_{ij}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$.
- 3) $(\overline{x}_1, \ldots, \overline{x}_k, \overline{y})$ est toujours sur la droite de régression.

Autre interprétation

Exemple: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$.
$\rightarrow$ Comment obtenir $\hat{\beta}_1$?

- Résoudre le système de $k+1$ équations. Ou ...
- Régresser x_1 sur x_2: $\hat{x}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 x_2$.
- Calculer $\hat{r}_1 = x_1 - \hat{x}_1 \rightarrow$ ôter de x_1 l'effet lié à x_2.
- $\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1} y_i}{\sum_{i=1}^n \hat{r}_{i1}^2}$.
- β_1 est bien l'effet net de x_1 sur y, où net signifie qu'on a tenu compte de l'effet des autres variables explicatives.
- $\rightarrow y = \beta_0 + \beta_1 x_1 + u$ donnera le même estimateur de $\hat{\beta}_1$ que dans la régression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ ssi $\hat{\alpha}_1 = 0$, c-à-d $Corr(x_1, x_2) = 0$.

Goodness-of-Fit (GoF)

Il est possible de décomposer la variabilité observée sur y, c-à-d $SST \equiv \sum_{i=1}^{n}(y_i - \overline{y})^2$, en 2 quantités:

- $SSE \equiv \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$: variabilité expliquée par le modèle;
- $SSR \equiv \sum_{i=1}^{n} \hat{u}_i^2$: variabilité **non**-expliquée par le modèle.

$\rightarrow SST = SSE + SSR$.
On peut donc définir une mesure de “qualité” de la régression (GoF): $R^2 = SSE/SST$ compris entre 0 et 1.
Notons que $R^2 \uparrow$ quand $k \uparrow \rightarrow$ pas très utile pour voir si une nouvelle variables explicative apporte de l’information.

Propriétés des estimateurs MCO

MLR=Multiple Linear Regression.

- **MLR.1** $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \rightarrow$ linéaire en les paramètres.
- **MLR.2** $(x_{1i}, \ldots, x_{ki}, y_i) : i = 1, \ldots, n \rightarrow$ échantillon aléatoire de taille n tiré de la population.
- **MLR.3** $E(u|x_1, \ldots, x_k) = 0 \rightarrow$ moyenne conditionnelle nulle ou variables explicatives exogènes.
- **MLR.4** Aucune variable explicative $x_j \forall j = 1, \ldots, k$ n'est constante dans l'échantillon et pas de relation linéaire parfaite entre les x_j.
- **Théorème 3.1**: Sous les hypothèses **MLR.1 - MLR.4**, $E(\hat{\beta}_j) = \beta_k \; \forall j = 0, \ldots, k$.

Revenons sur MLR.3

- **MLR.3** $E(u|x_1, \ldots, x_k) = 0$.
- $\rightarrow$ problème si la forme fonctionnelle est mal spécifiée:
- Exemple 1) oubli de inc^2 dans $cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$.
- Exemple 2) on modélise $wage$ alors que le modèle décrit par **MLR.1** est sur $log(wage)$.
- $\rightarrow$ de manière générale, omettre une variable corrélée avec un des régresseurs.
- Exemple 3) $wage = \beta_0 + \beta_1 educ + \beta_2 abil + u$. Mais $abil$ non observé $\rightarrow$ on estime $wage = \beta_0^* + \beta_1^* educ + v (\equiv \beta_2 abil + u)$. Cependant $corr(v, educ) \neq 0$ car $corr(abil, educ) \neq 0 \Rightarrow$ **MLR.3** ne tient pas.

Revenons sur MLR.4

- **MLR.4**: pas de collinéarité parfaite entre les x_j.
- Il y a collinéarité parfaite lorsqu'une variable explicative est une combinaison linéaire parfaite des autres variables explicatives.
- $\rightarrow$ les variables peuvent être corrélées.
- Exemple: $cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$. $\rightarrow$ inc et inc^2 sont corrélés mais par parfaitement.
- Par contre $log(cons) = \beta_0 + \beta_1 log(inc) + \beta_2 log(inc^2) + u$. $\rightarrow log(inc)$ et $log(inc^2) = 2log(inc)$ sont parfaitement corrélés.
- Exemple: $cons = \beta_0 + \beta_1 inc + \beta_2 (inc * 1000)$.
- Exemple: $voteA = \beta_0 + \beta_1 expendA + \beta_2 expendB + \beta_3 totexpend \rightarrow$ problème car $expendA + expendB = totexpend$.

Inclusion de variables redondantes

- Supposons que le modèle de la population est $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$ et qu'il satisfait **MLR.1-4**.
- De plus $\beta_3 = 0 \longrightarrow E(y|x_1, x_2, x_3) = E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
- Mais comme nous ne le savons pas *a priori*, nous estimons le modèle $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$.
- $\longrightarrow E(\hat{\beta}_0) = \beta_0$, $E(\hat{\beta}_1) = \beta_1$, $E(\hat{\beta}_2) = \beta_2$ et $E(\hat{\beta}_3) = \beta_3$.
- Pas d'effet sur le caractère non-biaisé des MCO.
- Peut-être un effet sur la précision (variance) des estimateurs !

Omission de variables

- Supposons que le modèle de la population est $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ et qu'il satisfait **MLR.1-4**.
- Exemple: $wage = \beta_0 + \beta_1 educ + \beta_2 abil + u.$
- On n'observe pas $abil$ (ou x_2) et donc on estime $wage = \beta_0 + \beta_1 educ + v$, où $v = \beta_2 abil + u$
- L'estimation donne $\widetilde{wage} = \widetilde{\beta}_0 + \widetilde{\beta}_1 educ$, ou de manière générale $\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1$.
- $\rightarrow \widetilde{\beta}_1 = \frac{\sum_{i=1}^n (x_{i1} - \overline{x}_1) y_i}{\sum_{i=1}^n (x_{i1} - \overline{x}_1)^2} = \frac{\sum_{i=1}^n (x_{i1} - \overline{x}_1) y_i}{SST_1}$.

Omission de variables

- Notons que:

$$\begin{aligned}
\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)y_i &= \sum_{i=1}^{n}(x_{i1}-\overline{x}_1)(\beta_0+\beta_1x_1+\beta_2x_2+u) \\
&= \underbrace{\beta_1\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)}_{SST_1}+\beta_2\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)x_{i2} \\
&+ \sum_{i=1}^{n}(x_{i1}-\overline{x}_1)u_i
\end{aligned}$$

- $E(\widetilde{\beta}_1)=\beta_1+\beta_2\frac{\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)^2}\rightarrow$ non-biaisé uniquement si $\beta_2\neq 0$ ou $corr(x_1,x_2)=0$ (c-à-d **MLR.3**).

Propriétés des estimateurs MCO

- **MLR.5** $Var(u|x_1, \ldots, x_k) = \sigma^2 \rightarrow$ homoscédasticité.
- **Théorème 3.2**: Sous les hypothèses **MLR.1 - MLR.5** (hypothèses de Gauss-Markov) ,

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

où R_j^2 est le R^2 de le régression de x_j sur les autres x (+ une constante) et $SST_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$.

- 1. σ^2 grand $\rightarrow Var(\hat{\beta}_j)$ grand.
- 2. $SST_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$ grand $\rightarrow Var(\hat{\beta}_j)$ petit.
Corollaire: n petit peut avoir pour conséquence $Var(\hat{\beta}_j)$ grand.

Propriétés des estimateurs MCO

- 3. R_j^2 grand $\rightarrow Var(\hat{\beta}_j)$ grand.
- Exemple: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 u$. Dans le cas extrême ou $R_1^2 (\equiv R^2$ de la régression de x_1 sur $x_2)$ est proche de 1 $\rightarrow$ on ne peut quasiment pas distinguer β_1 et β_2 $\rightarrow Var(\hat{\beta}_1)$ et $Var(\hat{\beta}_2)$ grands.
- Dans ce cas on a de la multicolinéarité et son effet est que $Var(\hat{\beta}_j) \rightarrow \infty$ si $R_j^2 \rightarrow 1$.

$Var(\hat{\beta}_j)$ dans un modèle mal spécifié

- Modèle de la population: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ et **MLR.1 - MLR.5** vérifiés.
- On considère 2 estimateurs de β_1:

$$\begin{aligned}
\hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \\
\widetilde{y} &= \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1
\end{aligned}$$

- Si $\beta_2 \neq 0 \rightarrow \widetilde{\beta}_1$ est biaisé $\rightarrow \hat{\beta}_1$ est préférable en terme de biais.

$Var(\hat{\beta}_j)$ dans un modèle mal spécifié

- En terme de variance:

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1(1-R_1^2)} \text{ et } Var(\widetilde{\beta}_1) = \frac{\sigma^2}{SST_1}$$

- $\rightarrow Var(\widetilde{\beta}_1) < Var(\hat{\beta}_1)$ si $R_1^2 \neq 0$.
- $\rightarrow Var(\widetilde{\beta}_1) = Var(\hat{\beta}_1)$ si $R_1^2 = 0$.
- Par conséquent si $R_1^2 = 0$ (x_1 et x_2 sont non-corrélés):

Quand	$\widetilde{\beta}_1$	$\hat{\beta}_1$	et
$\beta_2 \neq 0$	biaisé	non biaisé	$Var(\widetilde{\beta}_1) < Var(\hat{\beta}_1)$
$\beta_2 = 0$	non biaisé	non biaisé	$Var(\widetilde{\beta}_1) < Var(\hat{\beta}_1)$

- $\rightarrow$ ajouter des variables explicatives redondantes diminue la précision des estimateurs.

Estimation de σ^2

- σ^2 est la variance de u, c-à-d la variance du terme d'erreur de la population $u_i = y_i - \beta_0 - \beta_1 x_i \to$ jamais observé.
- Ce que l'on observe après estimation c'est $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$.
- Ou encore,

$$\begin{aligned} \hat{u}_i &= (\beta_0 + \beta_1 x_i + u_i) - \hat{\beta}_0 - \hat{\beta}_1 x_i \\ &= u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i . \end{aligned}$$

- Comme $\sigma^2 = E(u^2)$, un estimateur non-biaisé de $\sigma^2 = \frac{1}{n}\sum_{i=1}^n u_i^2$. Mais u_i n'est pas observé.
- On serait tenté de remplacer u_i par $\hat{u}_i \to \sigma^2 = \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 = SSR/n$.

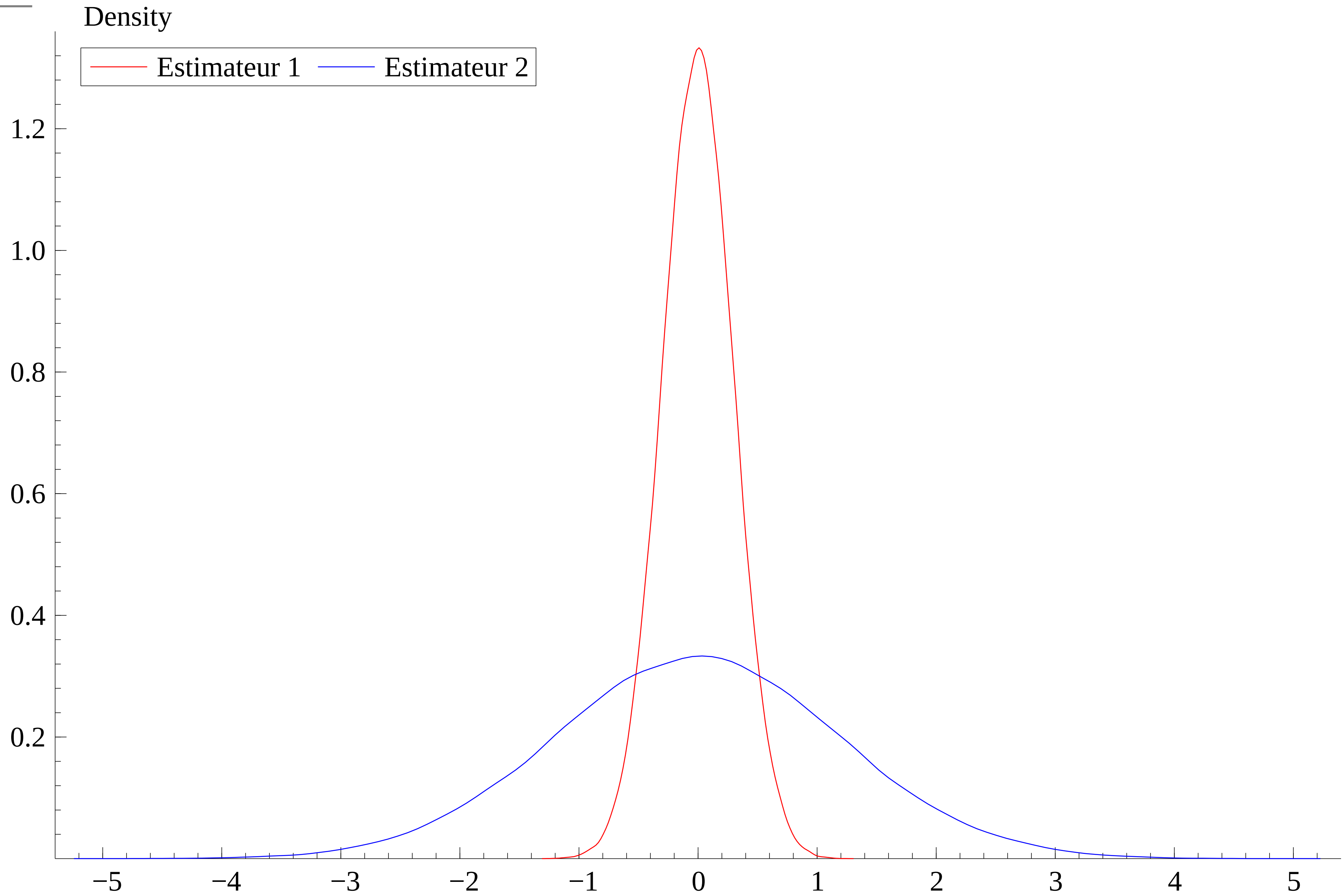
Estimation de σ^2

- Mais $E(SSR/n) \neq \sigma^2$.
- Ceci vient du fait que pour calculer $\hat{u}_i$ par MCO (c-à-d $\hat{\beta}_0$ et $\hat{\beta}_1$) on a utilisé $k+1$ restrictions.
- $\sum_{i=1}^{n} \hat{u}_i = 0$ et $\sum_{i=1}^{n} x_{ij}\hat{u}_i = 0 \ \forall j = 1, \ldots, k \rightarrow$ conditions de moment.
- $\rightarrow k+1$ valeurs de $\hat{u}_i$ sont contraintes.
- On a donc $(n-k-1)$ degrés de libertés.
- $\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \hat{u}_i^2 = SSR/(n-k-1)$ est un estimateur non-biaisé de $\sigma^2 \rightarrow E(\hat{\sigma}^2) = \sigma^2$.

En résumé

- **Théorème 3.3**: Sous les hypothèses **MLR.1 - MLR.5** (appelées hypothèses de Gauss-Markov), $E(\hat{\sigma}^2) = \sigma^2$, où $\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} \hat{u}_i^2$.
- **Théorème 3.4 où théorème de Gauss-Markov**: Sous les hypothèses **MLR.1 - MLR.5**, $\hat{\beta}_j$ est BLUE, $\forall i = 0, \ldots, k$.
- Linéaire: $\widetilde{\beta}_j$ est un estimateur linéaire de β_j s'il peut être exprimé comme une fonction linéaire des données $\rightarrow \widetilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$, où $w_{ij} = f(\mathbf{x})$.
- $\rightarrow$ entre 2 estimateurs non-biaisés il est logique de choisir celui qui a la plus petite variance.
- Preuve: 3.A.6 page 114.

Choix entre 2 estimateurs non-biaisés



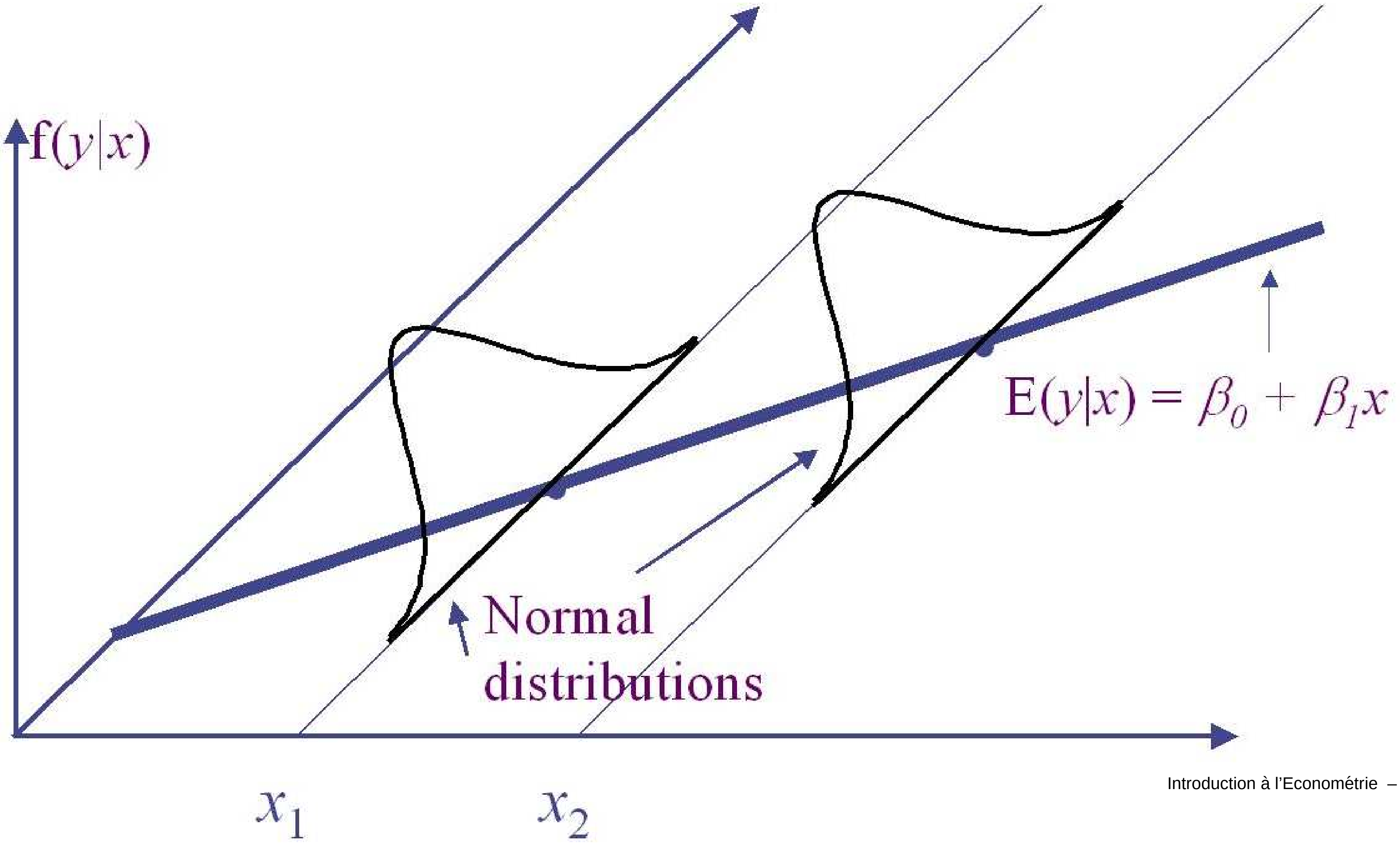
Chapitre 4: Modèle de régression multiple. Inférence

Distribution d'échantillonnage

Pour faire de l'inférence statistique, il faut ajouter une hypothèse supplémentaire:

- **MLR.6** $u \sim N(0, \sigma^2)$ et indépendant des $x_j \rightarrow$ normalité.
- **MLR.6** implique donc **MLR.3** et **MLR.5**.
- Les hypothèses **MLR.1 - MLR.6** sont appelées hypothèses *classiques* du modèle linéaire (CLM) = Gauss-Markov + normalité.
- CLM $\rightarrow$ MCO à la variance minimale (sans restreindre aux modèles linéaires).
- CLM $\rightarrow y|\mathbf{x} \sim N(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2)$.

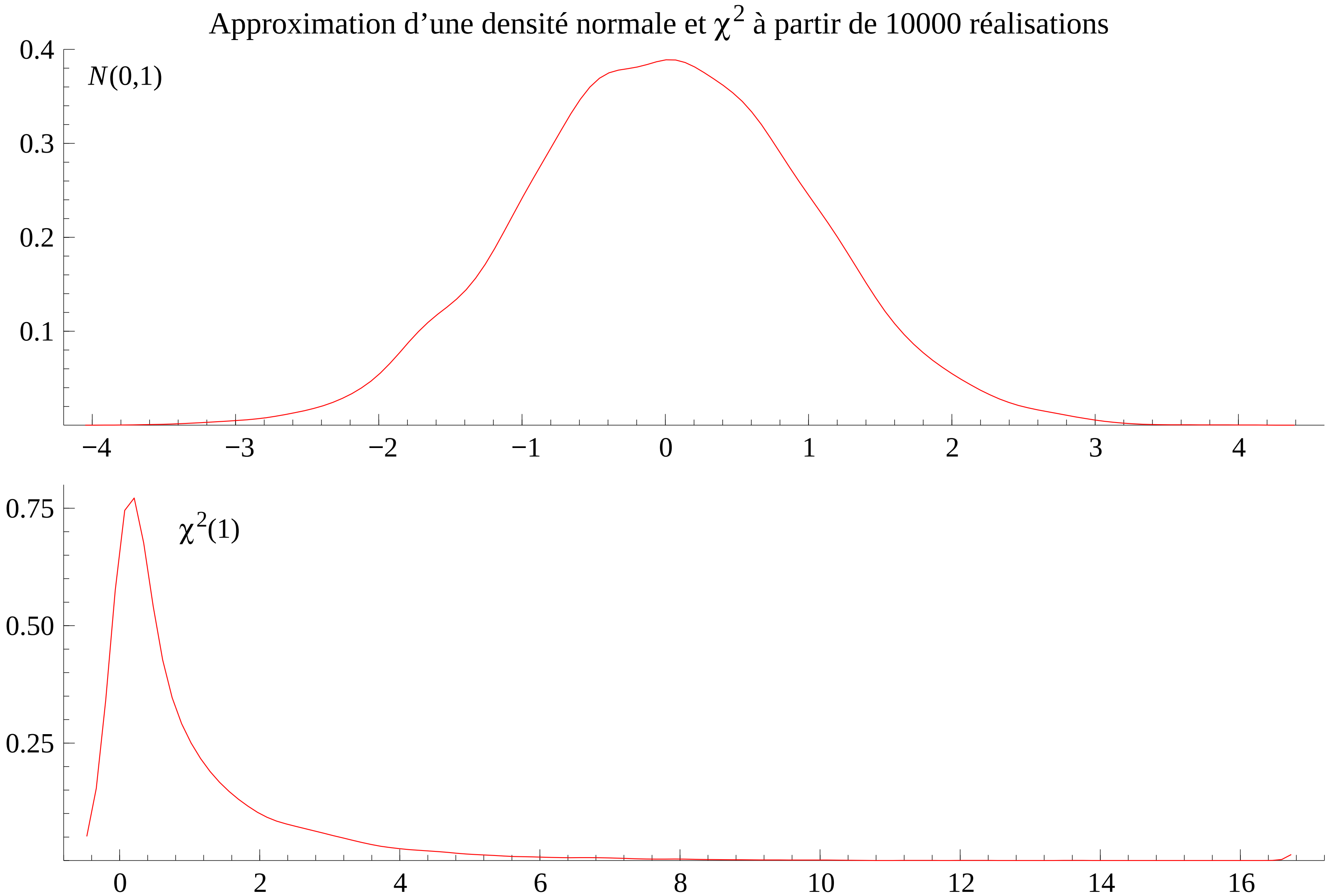
Normalité



Distribution d'échantillonnage

- Comment justifier cette hypothèse de normalité des résidus ?
- Si u est la somme de beaucoup de facteurs non-observés différents affectant y, on peut appeler le théorème central limite pour justifier la normalité.
- Théorème central limite : Soit $Y_1, Y_2, \ldots, Y_n$ un échantillon de variables aléatoires de moyenne μ et de variance σ^2.
- $\rightarrow Z_n = \frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}}$ suit asymptotiquement une distribution N(0,1).
- Exemple: $Y \sim \chi^2(1)$, $\mu = 1$ et $\sigma^2 = 2$.

$N(0,1)$ et $\chi^2(1)$



Approximation d'une densité normale et χ² à partir de 10000 réalisations
N(0,1)
0.4
0.3
0.2
0.1
−4
−3
−2
−1
0
1
2
3
4
χ²(1)
0.75
0.50
0.25
0
2
4
6
8
10
12
14
16

Théorème Central Limite

- Illustration.
- u est la somme de beaucoup de facteurs non-observés différents affectant y ?
- Faiblesse 1: les facteurs peuvent avoir des distributions différentes $\rightarrow$ le TCL marche moins bien.
- Faiblesse 2: le TCL suppose que les facteurs non-observés affectent y de manière séparée et additive $\rightarrow$ pas certain en pratique.
- Faiblesse 3: dans certains cas il est certain que l'hypothèse de normalité ne tient pas. Exemple: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$ $\rightarrow wage \sim N(\beta_0 + \beta_1 educ + \beta_2 exper, \sigma^2)$?
- Non car si $wage \sim N(\beta_0 + \beta_1 educ + \beta_2 exper, \sigma^2)$ $\rightarrow wage \in \Re$ alors que $wage > wage_{min} \geq 0$.

Non-normalité

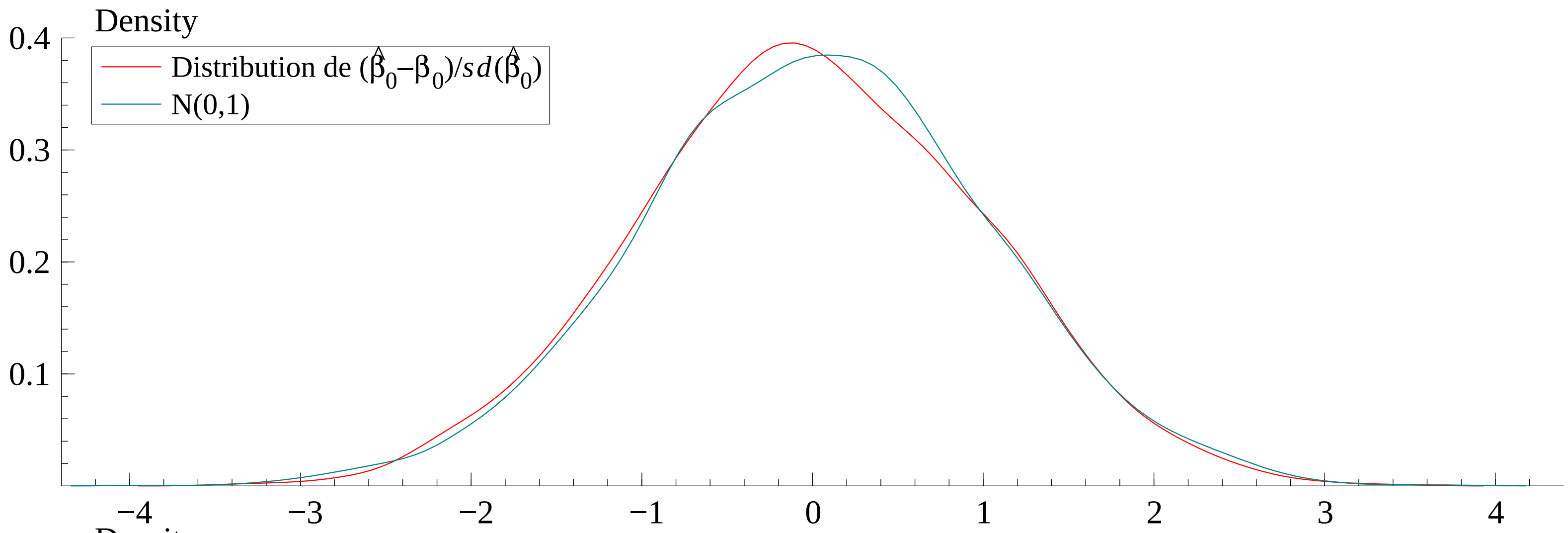
- $\longrightarrow$ il est désirable de tester cette hypothèse *ex-post*.
- Mais beaucoup d'études rejettent cette hypothèse.
- Est-ce dramatique ?

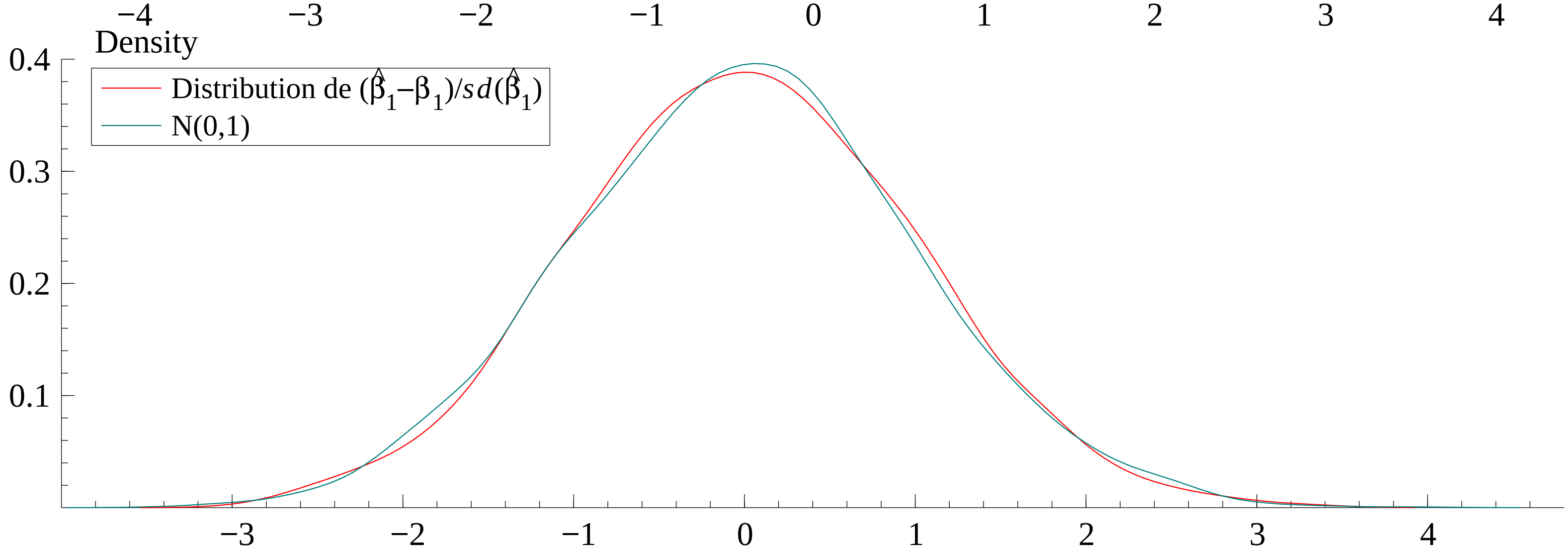
Inférence

- **Théorème 4.1**: Sous les hypothèses **MLR.1 - MLR.6**, $\hat{\beta}_j \sim N(\beta_j, Var(\beta_j))$, $\forall i = 0, \ldots, k$, où $Var(\beta_j) = \frac{\sigma^2}{(1-R_j^2)\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2}$.
- $\rightarrow \frac{\hat{\beta}_j-\beta_j}{\sqrt{Var(\beta_j)}} = \frac{\hat{\beta}_j-\beta_j}{sd(\beta_j)} \sim N(0,1)$.
- Illustration par une simulation.
$\rightarrow n = 20, y = 0.2 + 0.5x_1 + u$, et $u \sim N(0, 1.5)$.
Estimer $y = \beta_0 + \beta_1 x_1 + u$ 5000 fois et calculer après chaque estimation: $stat_{h,j} = \frac{\hat{\beta}_j-\beta_j}{sd(\beta_j)}$ $\forall h = 1, \ldots 5000$ et $j = 0, 1$. Notez qu'ici $\sigma^2 = 1.5$.

$stat_{h,j} \sim N(0,1)$?

$n = 20, y = 0.2 + 0.5x_1 + u$, $u \sim N(0, 1.5)$ et $\sigma^2 = 1.5$.





Tests d'hypothèse

- En pratique, le **Théorème 4.1** est inutile car on ne connaît pas σ^2.
- Si on remplace σ^2 par $\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \hat{u}_i^2 = SSR/(n-k-1)$, le **Théorème 4.1** n'est plus valide.
- **Théorème 4.2**: Sous les hypothèses **MLR.1 - MLR.6**, $\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim t_{n-k-1}$, $\forall i = 0, \ldots, k$, où $sd(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{(1-R_j^2)\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}}$.

Preuve

- Comme $sd(\beta_j) = \sqrt{\frac{\sigma^2}{(1-R_j^2)\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2}}$ et $sd(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{(1-R_j^2)\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2}}$.
- $\rightarrow sd(\beta_j) = \frac{\hat{s}d(\beta_j)}{\sqrt{\hat{\sigma}^2/\sigma^2}}$.
- Donc,

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} = \frac{\frac{\hat{\beta}_j - \beta_j}{sd(\beta_j)} \sim N(0,1)}{\sqrt{\hat{\sigma}^2/\sigma^2} \sim ?}$$

- Mais $\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2 = SSR/(n-k-1)$.
- $\rightarrow \sum_{i=1}^{n}\hat{u}_i^2 =$ somme de $(n-k-1)$ variables normales indépendantes au carré.

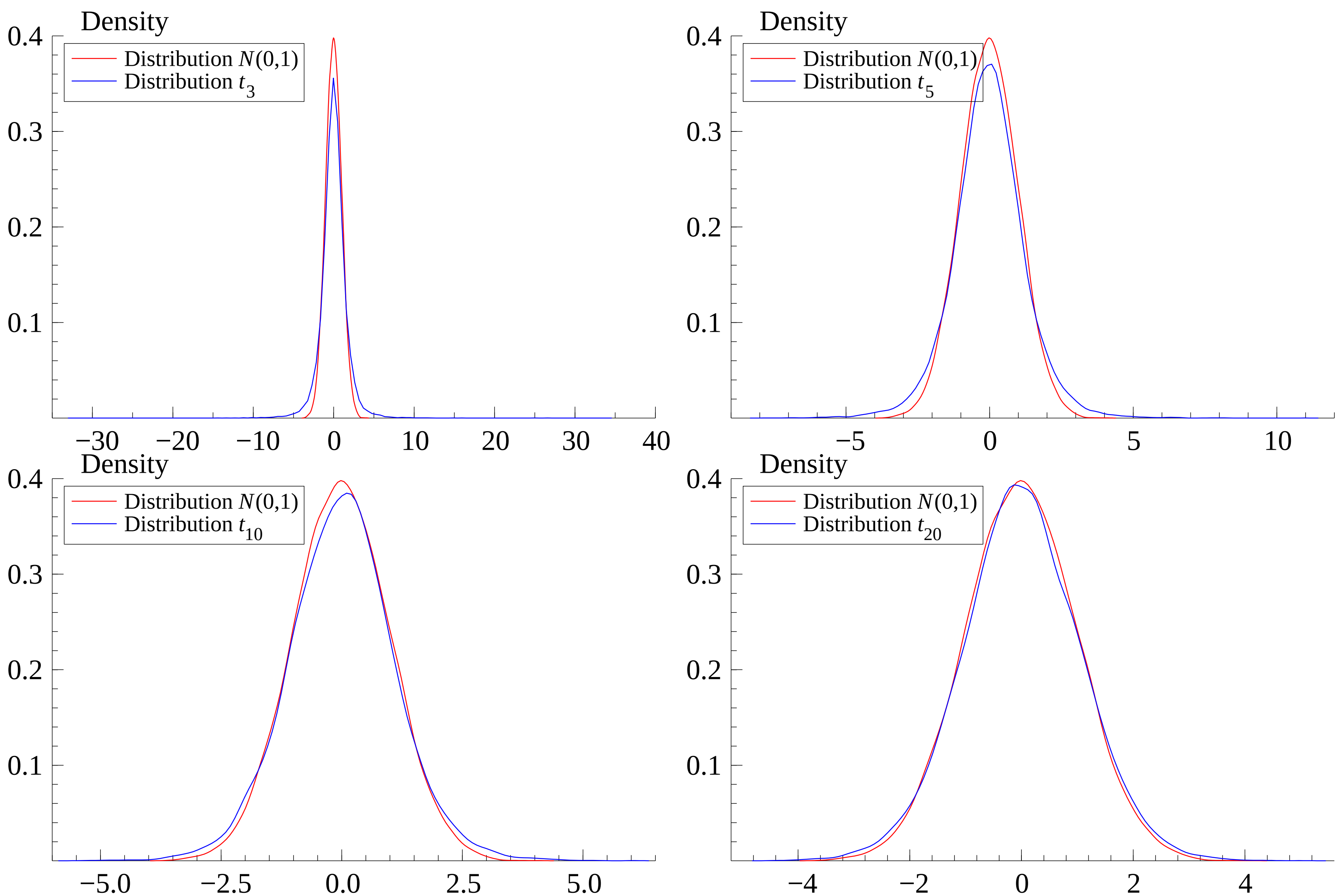
Preuve

- $\rightarrow \hat{\sigma}^2 \sim \frac{1}{n-k-1}\chi^2(n-k-1)$ (voir B.41 page 725).
- Donc,

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} = \frac{\frac{\hat{\beta}_j - \beta_j}{sd(\beta_j)} \sim N(0,1)}{\sqrt{\hat{\sigma}^2/\sigma^2} \sim \chi^2(n-k-1)} \sim t_{n-k-1},$$

  voir B.42 page 725. CQFD.
- Si $df \rightarrow \infty : t_{df} \rightarrow N(0,1)$.

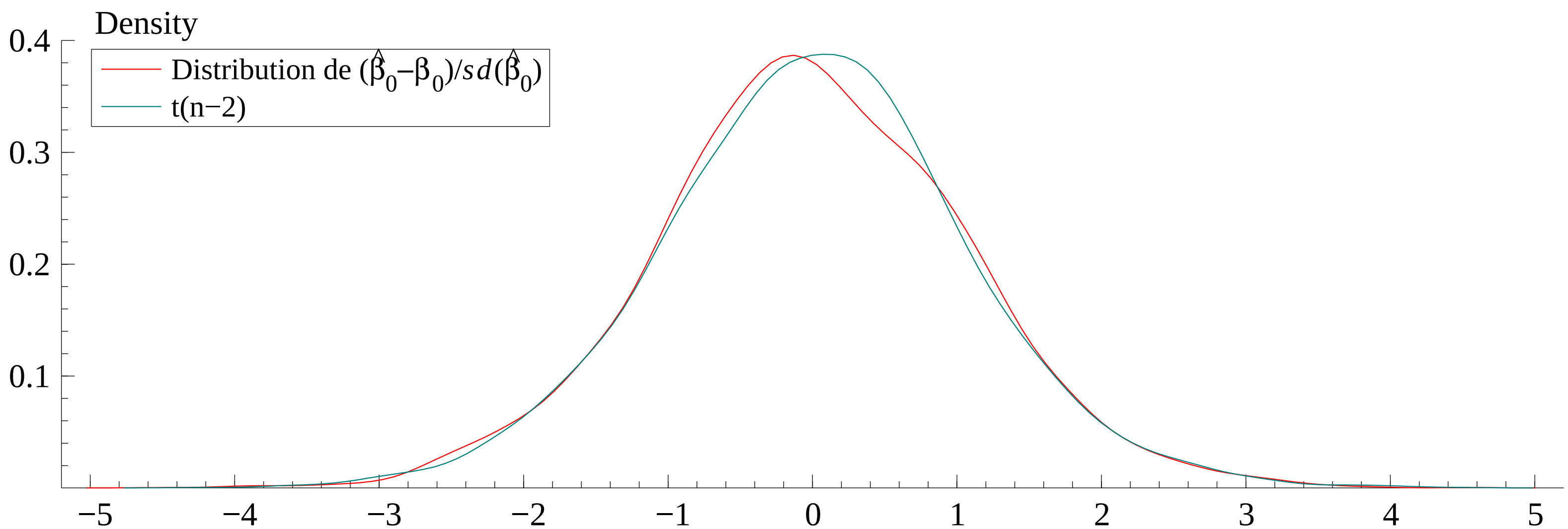
Graphiques

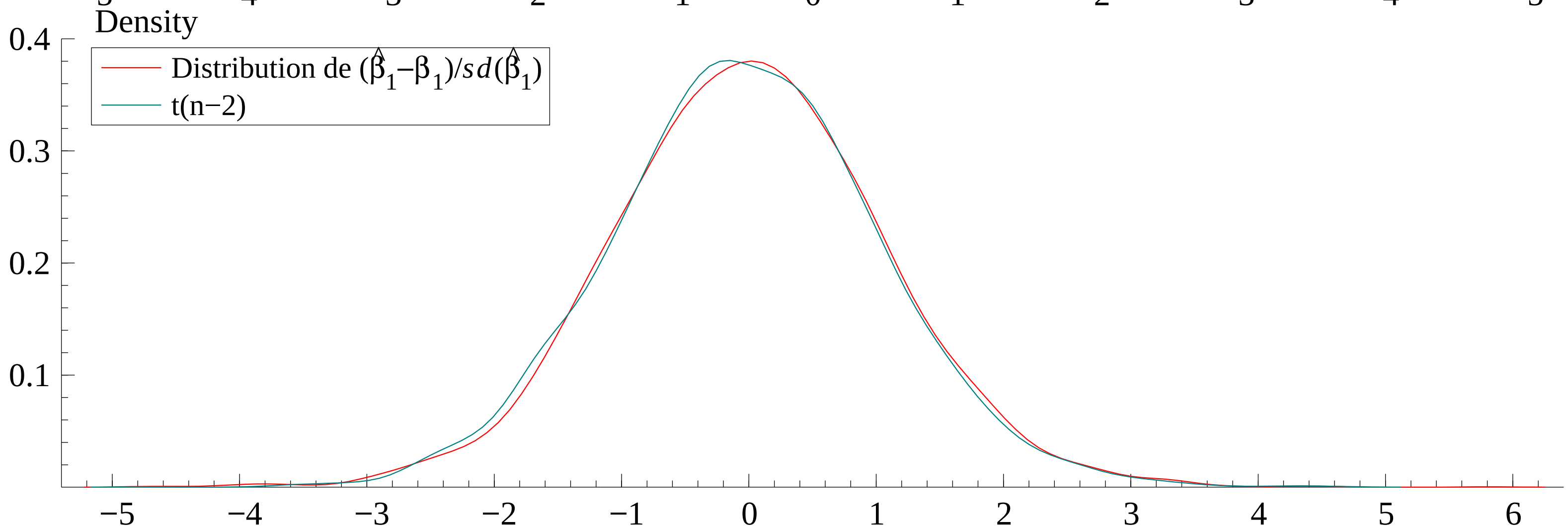


Density
Distribution $N(0,1)$
Distribution t_3
Distribution t_5
Distribution t_{10}
Distribution t_{20}
0.4
0.3
0.2
0.1
−30 −20 −10 0 10 20 30 40
−5 0 5 10
−5.0 −2.5 0.0 2.5 5.0
−4 −2 0 2 4

$stat_{h,j} \sim t(n-k)$?

$n = 20, y = 0.2 + 0.5x_1 + u$, $u \sim N(0, 1.5)$ et
$\hat{\sigma}^2 = SSR/(n-k-1)$.





Tests d'hypothèse

- Ceci nous permet par exemple de tester l'hypothèse nulle suivante:

$$H_0 : \beta_j = 0.$$

- $\rightarrow$ après avoir contrôlé l'effet des autres variables, est-ce que x_j a un effet sur y ?
- Exemple: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + u.$
  $H_0 : \beta_2 = 0 \rightarrow$ après avoir tenu compte de l'effet de l'éducation, est-ce que l'expérience explique le salaire horaire ? On parle des paramètres de la population.
- Ici, la t-stat est $t_{\hat{\beta}_2} \equiv \frac{\hat{\beta}_2 - 0}{sd(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{sd(\hat{\beta}_2)}$.

Exemple: WAGE

	Coefficient	Std.Error	t-value
Constant	-3.39054	0.7666	-4.42
EDUC	0.644272	0.05381	12.0
EXPER	0.0700954	0.01098	6.39

$t - stat$ est ici $t - value$.
$\rightarrow t - stat$ a le même signe que $\hat{\beta}_j$.
$\rightarrow$ Comment utiliser la t-stat ?

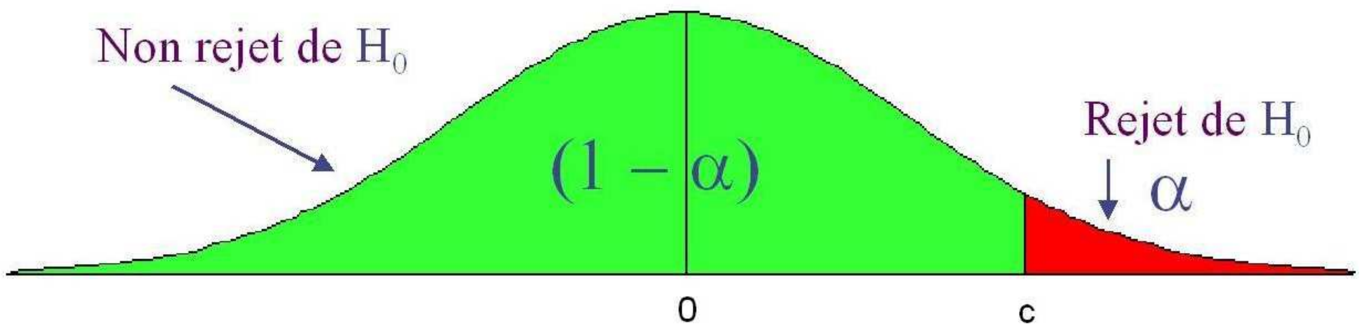
Comment utiliser la t-stat ?

- Même sous $H_0 : \beta_j = 0$, $\hat{\beta}_j$ ne sera jamais égal à zero.
- Ce que nous cherchons est une règle de décision pour rejeter H_0 avec un certain niveau de significativité.
- $\rightarrow$ probabilité de rejeter H_0 alors que H_0 est vraie.
- $\rightarrow$ requiert la distribution de $t_{\hat{\beta}_j}$ $\Rightarrow$ **Théorème 4.2**.

Alternative: H_1 ?

- Si H_0 est rejetée, en faveur de quoi est-elle rejetée ?
- $\rightarrow$ il faut spécifier une hypothèse alternative H_1.
- $\rightarrow$ 1. Alternative unilatérale: $H_1 > 0$ ou $H_1 < 0$.
- $\rightarrow$ 2. Alternative bilatérale: $H_1 \neq 0$.
- Il faut décider d'un niveau de significativité.
- $\rightarrow$ La plupart du temps 5%.
- Si on spécifie $H_1 > 0 : E(t_{\hat{\beta}_j}) = 0$ sous H_0, alors que sous H_1, $E(t_{\hat{\beta}_j}) > 0 \rightarrow$ rejet de H_0 uniquement si $t_{\hat{\beta}_j}$ est suffisamment > 0.
- $\rightarrow$ Si $t_{\hat{\beta}_j} > c$.
- $\rightarrow$ choisir c de telle manière à rejeter H_0 en faveur de H_1 alors que H_0 est vraie dans 5% des cas.

Test unilatéral: $H_1 > 0$

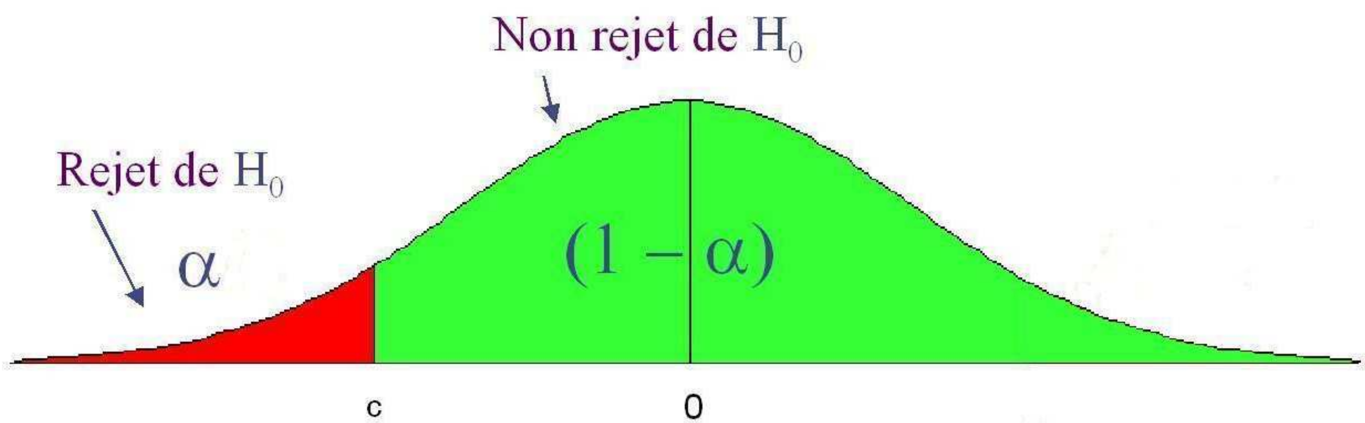


Exemple: $\alpha = 5\%$ et $n - k - 1 = 28 \rightarrow c = 1.701$.
Si $t_{\hat{\beta}_j} > 1.701 \Rightarrow$ rejet de H_0.
$\rightarrow$ Voir Table G.2 page 817: 1-tailed.
$\rightarrow$ Si rejet de H_0 pour $\alpha = 5\% \rightarrow$ rejet pour $\alpha = 10\%$.
$\rightarrow$ Si $df > 100, t_{\hat{\beta}_j}(5\%) \simeq 1.645$ et $t_{\hat{\beta}_j}(1\%) \simeq 2.326$.

Exemple: WAGE

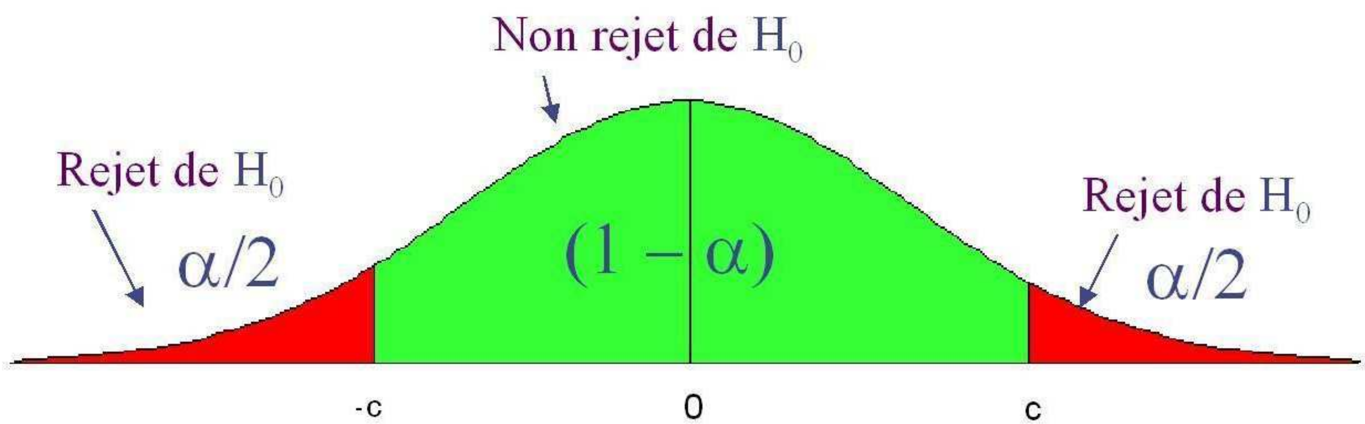
	Coefficient	Std.Error	t-value
Constant	-3.39054	0.7666	-4.42
EDUC	0.644272	0.05381	12.0
EXPER	0.0700954	0.01098	6.39

Test unilatéral: $H_1 < 0$



Exemple: $\alpha = 5\%$ et $n - k - 1 = 28 \rightarrow c = -1.701$.
Si $t_{\hat{\beta}_j} < -1.701 \Rightarrow$ rejet de H_0.
$\rightarrow$ Voir Table G.2 page 817: $-1\times$ valeur 1-tailed.
$\rightarrow$ Si rejet de H_0 pour $\alpha = 5\% \rightarrow$ rejet pour $\alpha = 10\%$.
$\rightarrow$ Pour tout valeur $t_{\hat{\beta}_j} > 0$ on ne rejette pas H_0.

Test bilatéral: $H_1 \neq 0$



Exemple: $\alpha = 5\%$ et $n - k - 1 = 28 \rightarrow c = 2.048$.
Si $t_{\hat{\beta}_j} > 2.048$ ou $t_{\hat{\beta}_j} < -2.048 \Rightarrow$ rejet de H_0.
$\rightarrow$ c-à-d si $|t_{\hat{\beta}_j}| > 2.048$.
$\rightarrow$ Voir Table G.2 page 817: 2-tailed.
$\rightarrow$ Si $df > 100, t_{\hat{\beta}_j}(5\%) \simeq 1.960$ et $t_{\hat{\beta}_j}(1\%) \simeq 2.576$.

Exemple: Crimes

$log(crime) = \beta_0 + \beta_1 log(enroll) + u$, où $crime$ est le nombre annuel de crimes par campus (de collège) et $enroll$ est le nombre d'étudiants par collège.

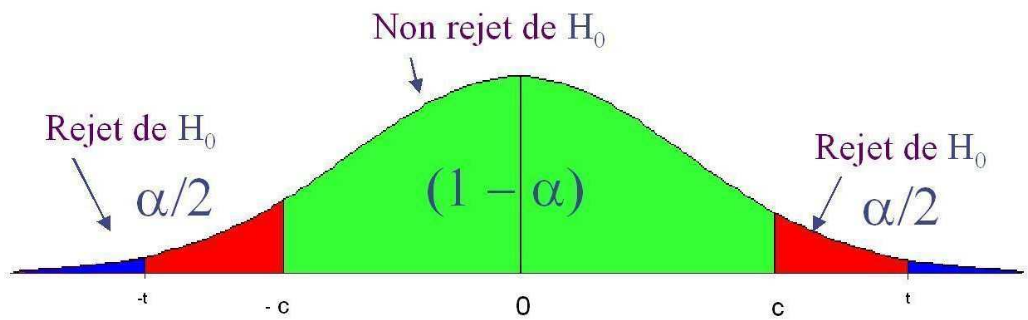
```
    Modelling LCRIME by OLS-CS (using campus)
         Coefficient Std.Error t-value t-prob
Constant    -6.63137     1.034   -6.42  0.000
LENROLL      1.26976    0.1098    11.6  0.000

no. of obs. 97           no. of parameters    2
```

Tester une valeur précise

- $H_0 : \beta_j = a_j \rightarrow t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j - a_j}{sd(\hat{\beta}_j)}$.
- $H_1 : \beta_j >=< a_j$.
- Exemple: $H_0 : \beta_1 = 1$ (c-à-d élasticité $crime/enroll = 1$) contre $H_1 : \beta_1 > 1$.
- $t_{\hat{\beta}_1} \equiv \frac{1.27-1}{0.11} \simeq 2.45$.
- $c(5\%) = 1.66$ et $c(1\%) = 2.37(df = 97 - 2 = 95 \simeq 90$ ou 120 pour la table).
- $\rightarrow$ rejet de H_0 même à 1%.
- ... pour autant que les hypothèses des MCO soient vérifiées.

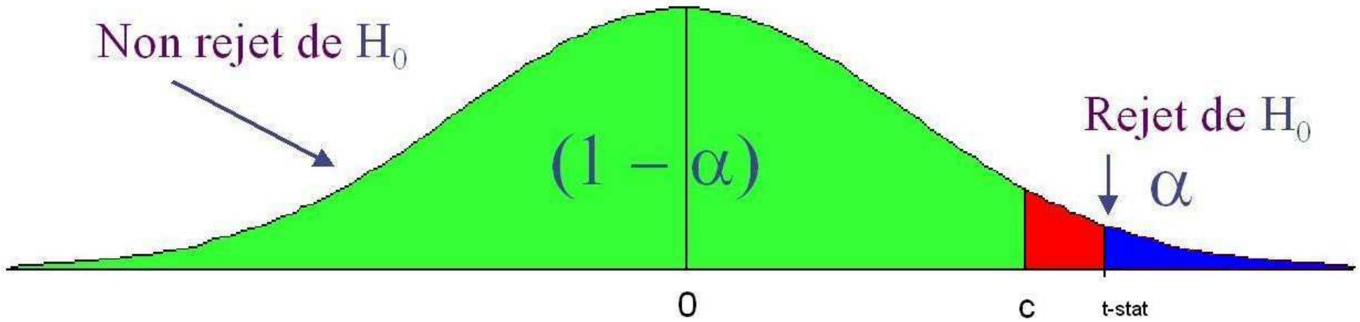
P-valeur

- P-valeur: Quelle est le plus petit niveau de significativité auquel H_0 serait rejeté.
- Exemple: $H_1 : \beta_j \neq 0$: P-valeur $= P(|T| > |t|) = 2P(T > |t|)$, où $T \sim t_{n-k-1}$ et t est la valeur observée de la t-stat sous H_0.



-
- → Rejeter H_0 si la P-Valeur < au seuil fixé (ex: 5%).

P-valeur d'un test $H_1 > 0$



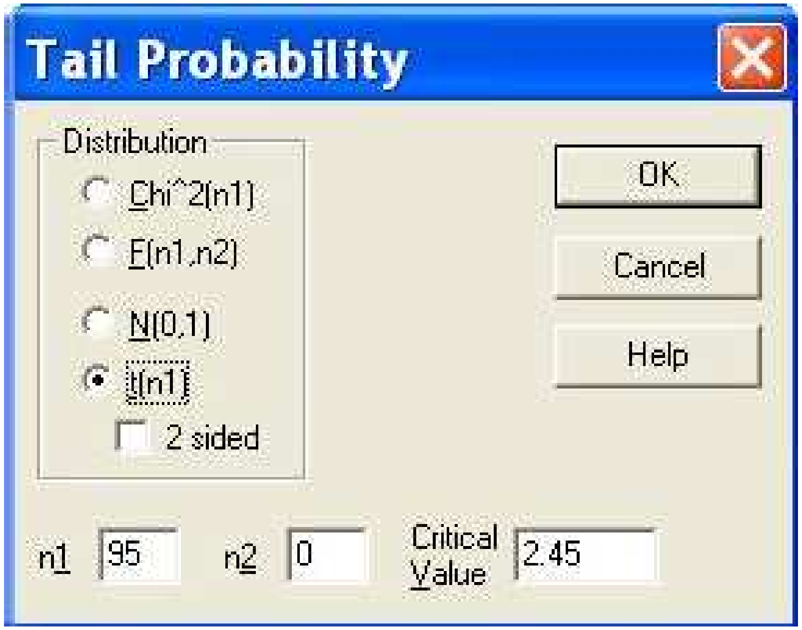
$H_1 : \beta_j > 0$: P-valeur $= P(T > t)$
$H_1 : \beta_j < 0$: P-valeur $= P(T < t) = P(T > |t|)$

Exemple: Crimes

```
Modelling LCRIME by OLS-CS (using campus)
           Coefficient Std.Error t-value
Constant      -6.63137     1.034   -6.42
LENROLL        1.26976    0.1098    11.6

no. of observations 97 R^2 0.584775
no. of parameters    2
```

$H_0 : \beta_1 = 1$ contre $H_1 : \beta_1 > 1$.
$\rightarrow t_{\hat{\beta}_1} \equiv \frac{1.27-1}{0.11} \simeq 2.45$.
$\rightarrow$ P-valeur $= P(T > t) = P(T > 2.45) = 0.0081$
$H_0 : \beta_1 = 1$ contre $H_1 : \beta_1 < 1$. $\rightarrow$ pas besoin car $t_{\hat{\beta}_1} > 0$.

Givewin




```
Output: t(95,1-sided) = 2.45 [0.0081] **
```

Language

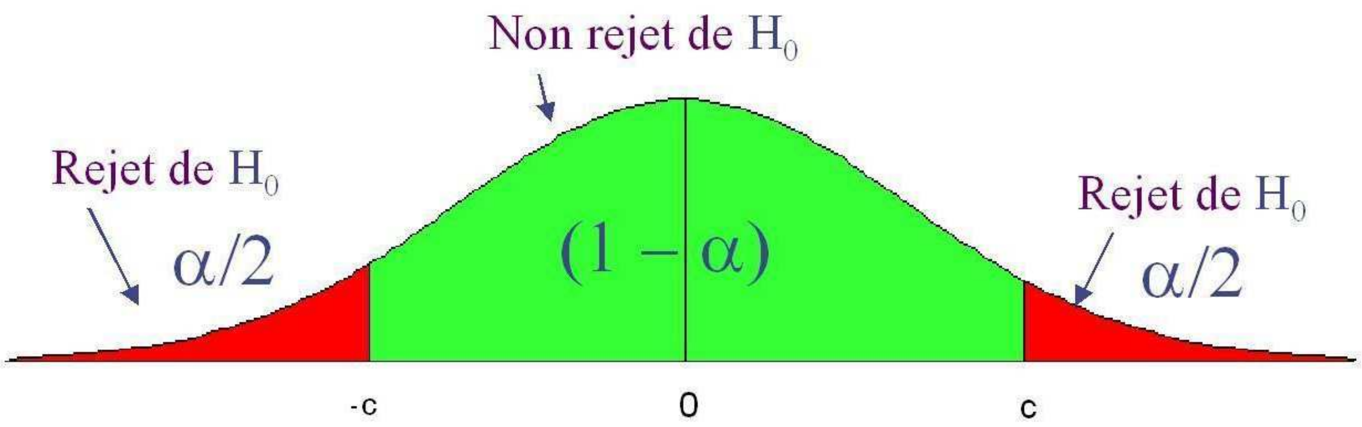
- Quand H_0 n'est pas rejeté on dit: *on ne rejette pas H_0 au seuil de $\alpha\%$* et non on accepte H_0 au seuil de $\alpha\%$.
- Quand on ne rejette pas H_0 au seuil de $\alpha\% \rightarrow$ sens statistique et non économique.
- $\rightarrow$ au sens économique, la valeur de $\hat{\beta}_j$ peut toutefois être non significative (car de faible ampleur).

Intervalle de confiance

- Rappelons que $t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim t_{n-k-1}$.
-

$$Pr\left[-c_{\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \leq c_{\alpha/2}\right] = 1 - \alpha$$

$$\Leftrightarrow \quad Pr\left[\hat{\beta}_j - c_{\alpha/2} sd(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_{\alpha/2} sd(\hat{\beta}_j)\right] = 1 - \alpha$$



Intervalle de confiance

- Intervalle de confiance à 95% de β_j si $df = 25$:
  $[\hat{\beta}_j - 2.06 sd(\hat{\beta}_j); \hat{\beta}_j + 2.06 sd(\hat{\beta}_j)]$.
- Intervalle de confiance à 95% de β_j si $df > 120$:
  $[\hat{\beta}_j - 2 sd(\hat{\beta}_j); \hat{\beta}_j + 2 sd(\hat{\beta}_j)]$.
- $H_0 : \beta_j = a_j$ contre $H_1 : \beta_j \neq a_j \rightarrow$ si a_j n'est pas dans l'intervalle de confiance à 95%, on rejette H_0 au seuil de 5%.

$H_0 : \beta_1 = \beta_2$

- De manière générale: test d'hypothèse sur une combinaison linéaire de paramètres.
- Exemple: $log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$, où jc et $univ$ sont le nombre d'années passées dans une école supérieur de type court et de type long.
- $H_0 : \beta_1 = \beta_2$ contre $H_1 : \beta_1 < \beta_2$.
- $\rightarrow H_0 : \beta_1 - \beta_2 = 0$ contre $H_1 : \beta_1 - \beta_2 < 0$.
- $\rightarrow t = \frac{\hat{\beta}_1 - \beta_j}{sd(\hat{\beta}_1 - \hat{\beta}_2)}$.
- $\rightarrow$ rejet de H_0 si $t < -c$.

Exemple: TWOYEAR

Modelling LWAGE by OLS-CS (using twoyear)

	Coefficient	Std.Error	t-value	t-prob
Constant	1.47233	0.02106	69.9	0.000
JC	0.0666967	0.006829	9.77	0.000
UNIV	0.0768762	0.002309	33.3	0.000
EXPER	0.00494422	0.0001575	31.4	0.000

R^2 0.222442 no. of observations 6763

$$
\begin{aligned}
sd(\hat{\beta}_1 - \hat{\beta}_2) &= \sqrt{Var(\hat{\beta}_1 - \hat{\beta}_2)} \\
&= \sqrt{Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)}
\end{aligned}
$$

Matrice de variance covariance

```
Covariance matrix of estimated parameters:

                Cons            JC           UNIV ...
Cons      0.00044353 -1.7414e-005 -1.5735e-005
JC      -1.7414e-005  4.6632e-005  1.9279e-006
UNIV    -1.5735e-005  1.9279e-006  5.3302e-006
EXPER   -3.1048e-006 -1.7183e-008  3.9335e-008
```

$\rightarrow$ Test/Further Output/Covariance matrix of estimated parameters.

$\rightarrow sd(\hat{\beta}_1 - \hat{\beta}_2) = 0.006935878$

$\rightarrow \frac{0.0666967 - 0.0768762}{0.006935878} = -1.467658512 > -1.645 \Rightarrow$ non-rejet de H_0 au seuil de 5%.

Autre méthode

- $log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$,
- $\rightarrow H_0 : \beta_1 - \beta_2 = 0$ contre $H_1 : \beta_1 - \beta_2 < 0$.
- On peut définir $\theta_1 = \beta_1 - \beta_2$ et tester $\theta_1 = 0$ contre $H_1 : \theta_1 < 0$.
- Sous H_0, le modèle se réécrit:

$$\begin{aligned} log(wage) &= \beta_0 + (\theta_1 + \beta_2) jc + \beta_2 univ + \beta_3 exper + u \\ &= \beta_0 + \theta_1 jc + \beta_2 \underbrace{(univ + jc)}_{totcoll} + \beta_3 exper + u \end{aligned}$$

Résultat

	Coefficient	Std.Error	t-value	t-prob
Constant	1.47233	0.02106	69.9	0.000
JC	-0.0101795	0.006936	-1.47	0.142
EXPER	0.00494422	0.0001575	31.4	0.000
TOTCOLL	0.0768762	0.002309	33.3	0.000

$\rightarrow$ Le résultat est interpretable directement.

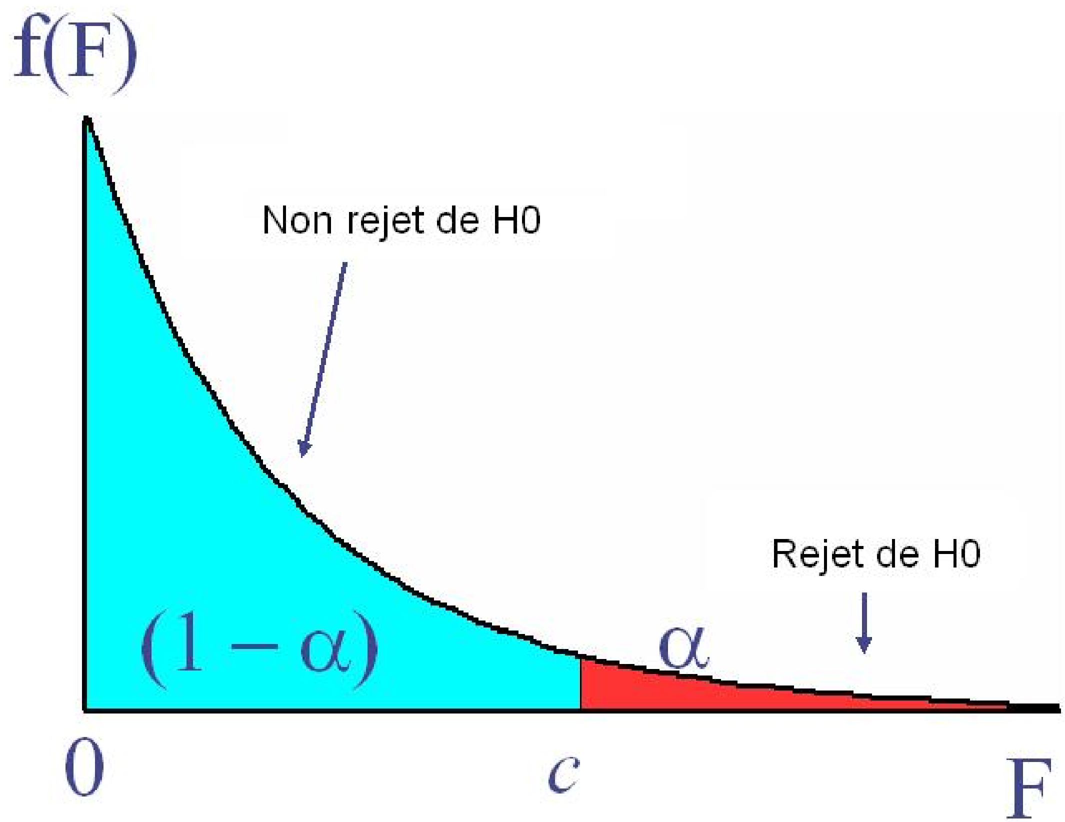
$\rightarrow$ La p-valeur est pour $H_0 : \theta_1 \neq 0$.

$\rightarrow$ Comme $\hat{\theta}_1 < 0$, on peut la diviser par 2 pour obtenir $P(T < t) \Rightarrow 0.071$.

Tests multiples de restrictions linéaires

- Exemple: $H_0 : \beta_1 = 0, \beta_2 = 0$.
- Ici, sous $H_1 : \beta_1 = 0$ ou $\beta_2 = 0 \rightarrow H_0$ n'est pas vrai.
- $\rightarrow 1)$ Estimer le modèle non contraint $\rightarrow k+1$ paramètres.
- $\rightarrow 2)$ Estimer le modèle contraint $\rightarrow k+1-q$ paramètres.
- $\rightarrow 3)$ Calculer la statistique $F \equiv \frac{(SSR_c - SSR_{nc})/q}{SSR_{nc}/(n-k-1)}$,où SSR dénote la somme des carrés des résidus et les indices c et nc signifient respectivement contraint et non-contraint.
- Sous H_0, $F \sim F_{q,n-k-1}$, où F =Fisher.
- On a aussi: $F \equiv \frac{(R^2_{nc} - R^2_c)/q}{(1-R^2_{nc})/(n-k-1)}$.

Distribution de Fisher



Rejet H_0, si $F > c_\alpha$, où $c = F^{\alpha}_{q,n-k-1}$.

Exemple: Salaire de joueurs de baseball

- $$\begin{aligned} log(salary) &= \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg \\ &+ \beta_4 hrunsyr + \beta_5 rbisyr + u. \end{aligned}$$
- $log(salary) =$ salaire total en 1993 en \$.
- $years =$ nombre d'années dans la league.
- $gamesyr =$ nombre moyen de parties jouées par an.
- $bavg =$ score moyen à la batte.
- $hrunsyr =$ nombre de "home run" par an (coup de batte qui permet au batteur de marquer un point en faisant un tour complet en une seule fois).
- $rbisyr$ nombre de "home run" tentés.

Exemple: MLB1

```
EQ( 1) Modelling LSALARY by OLS-CS (using mlb1)

                Coefficient  Std.Error  t-value  t-prob
Constant            11.1924     0.2888     38.8   0.000
YEARS             0.0688626    0.01211     5.68   0.000
GAMESYR           0.0125521   0.002647     4.74   0.000
BAVG            0.000978604   0.001104    0.887   0.376
HRUNSYR           0.0144295    0.01606    0.899   0.369
RBISYR            0.0107657   0.007175     1.50   0.134

sigma             0.726577  RSS                183.186321
R^2               0.627803  F(5,347) =   117.1 [0.000]**
no. of obs.            353  no. of parameters           6
```

Exemple: MLB1

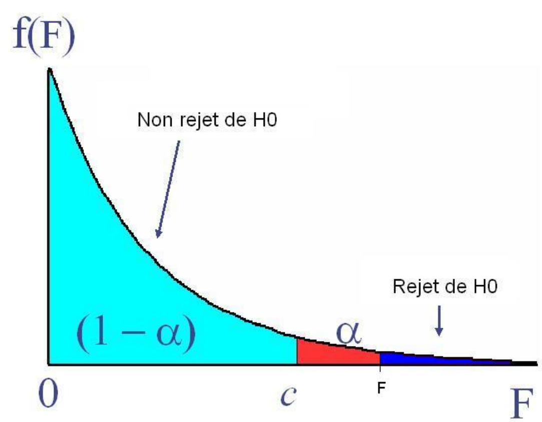
```
EQ( 2) Modelling LSALARY by OLS-CS (using mlb1)

              Coefficient  Std.Error  t-value  t-prob
Constant          11.2238     0.1083     104.   0.000
YEARS           0.0713180    0.01251     5.70   0.000
GAMESYR         0.0201745   0.001343     15.0   0.000

sigma            0.752731  RSS               198.311468
R^2              0.597072  F(5,347) =  259.3 [0.000]**
no. of obs.           353  no. of parameters          3
```

Exemple: Salaire de joueurs de baseball

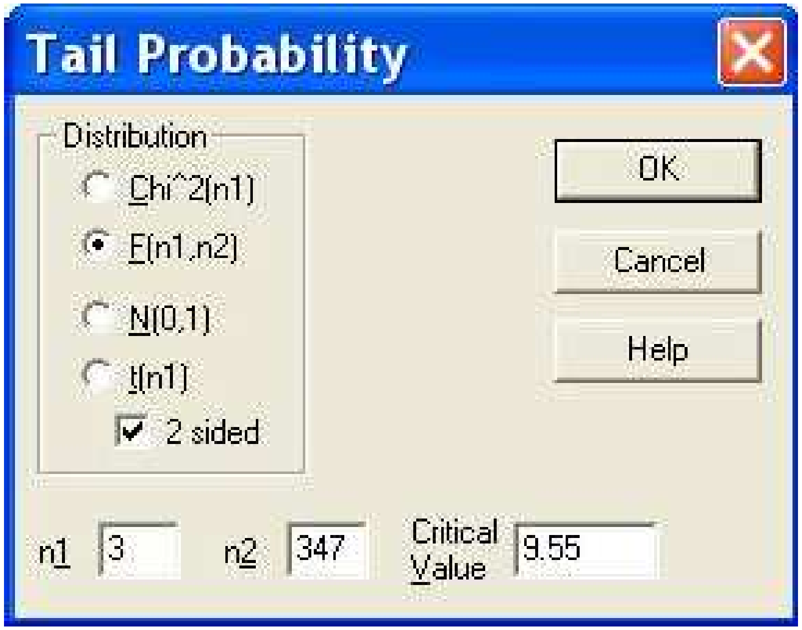
- $F \equiv \frac{(198.311-183.186)/3}{183.186/347} \approx 9.55$.
- $F \equiv \frac{(0.6278-0.5971)/3}{(1-0.6278)/347} \approx 9.55$.
- Tables G.3a-b-c:
  $c(10\%) = 2.08, c(5\%) = 2.60, c(1\%) = 3.78$.
- $\rightarrow F >> c(1\%) \Rightarrow$ rejet de $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$ en faveur de H_1 : au moins un des paramètres est significatif.
- Or, individuellement les 3 paramètres sont non significatifs. Pourquoi ?

Relation entre F et t tests

- Si on effectue un F test du style $H_0 : \beta_j = a_j$ contre $H_1 : \beta_j \neq a_j$, on peut montrer que $F = t^2$.
- Comme $F_{1,n_k-1} = t^2_{n_k-1} \rightarrow$ résultat identique (même p-valeur).
- Par contre le t-test est plus flexible car il permet des alternatives du type $\beta_j <$ ou $> a_j$.
- P-valeur du test $F = P(\Im > F)$, où $\Im \sim F_{q,n-k-1}$.



Givewin



- Output: F(3, 347) = 9.55 [0.0000] **

Test F reporté par défaut

- $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$.
- $\longrightarrow$ test de significativité global.
- Modèle contraint: $y = \beta_0 + u$.
- $F \equiv \frac{R^2/q}{(1-R^2)/(n-k-1)}$.
- Reporté par défaut à chaque régression incluant β_0.

Tester $H_0 : \beta_1 = 1, \beta_2 = 0$?

- Modèle de la population: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$
- Sous $H_0 : y = \beta_0 + 1x_1 + 0x_2 + \beta_3 x_3 + u$.
- $\rightarrow y - x_1 = \beta_0 + \beta_3 x_3 + u$.
- $\rightarrow w = \beta_0 + \beta_3 x_3 + u$, où $w = y - x_1$.
- Sous $H_0 : F \sim F_{2,n-3-1}$.

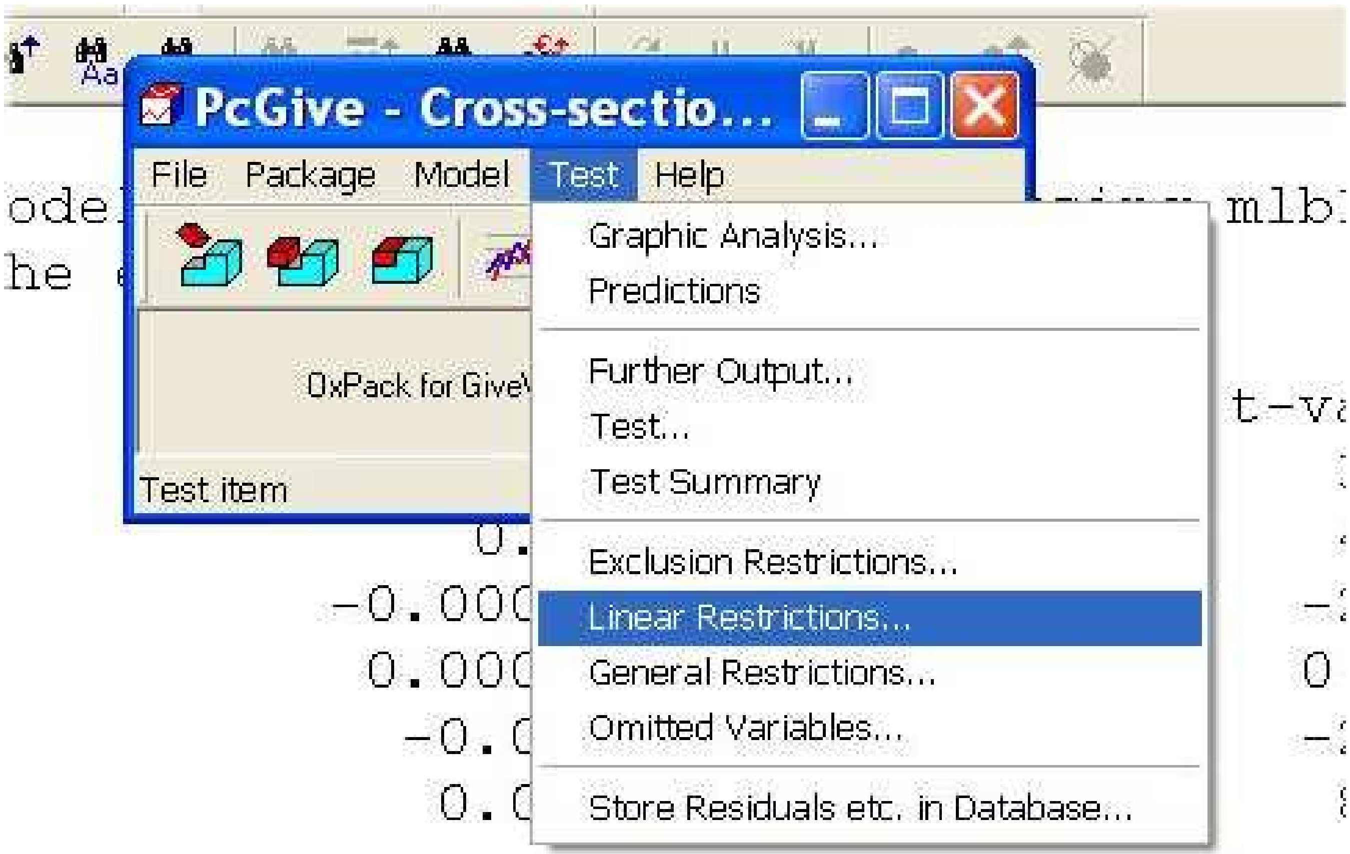
De manière générale $\mathbf{RB} = \mathbf{r}$

- On peut écrire ces restrictions linéaires de manière plus générale.
- H_0 : $\mathbf{RB} = \mathbf{r}$.
- H_1 : $\mathbf{RB} \neq \mathbf{r}$.
- $\underset{(k+1\times 1)}{\mathbf{B}} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$, $\underset{(q\times 1)}{\mathbf{r}} = \begin{bmatrix} r_1 \\ \vdots \\ r_q \end{bmatrix}$ et $\underset{(q\times k+1)}{\mathbf{R}} =$ matrice de constantes.
- Exemple: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \rightarrow k = 2$.
- $\rightarrow \underset{(3\times 1)}{\mathbf{B}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$.

Autres exemples

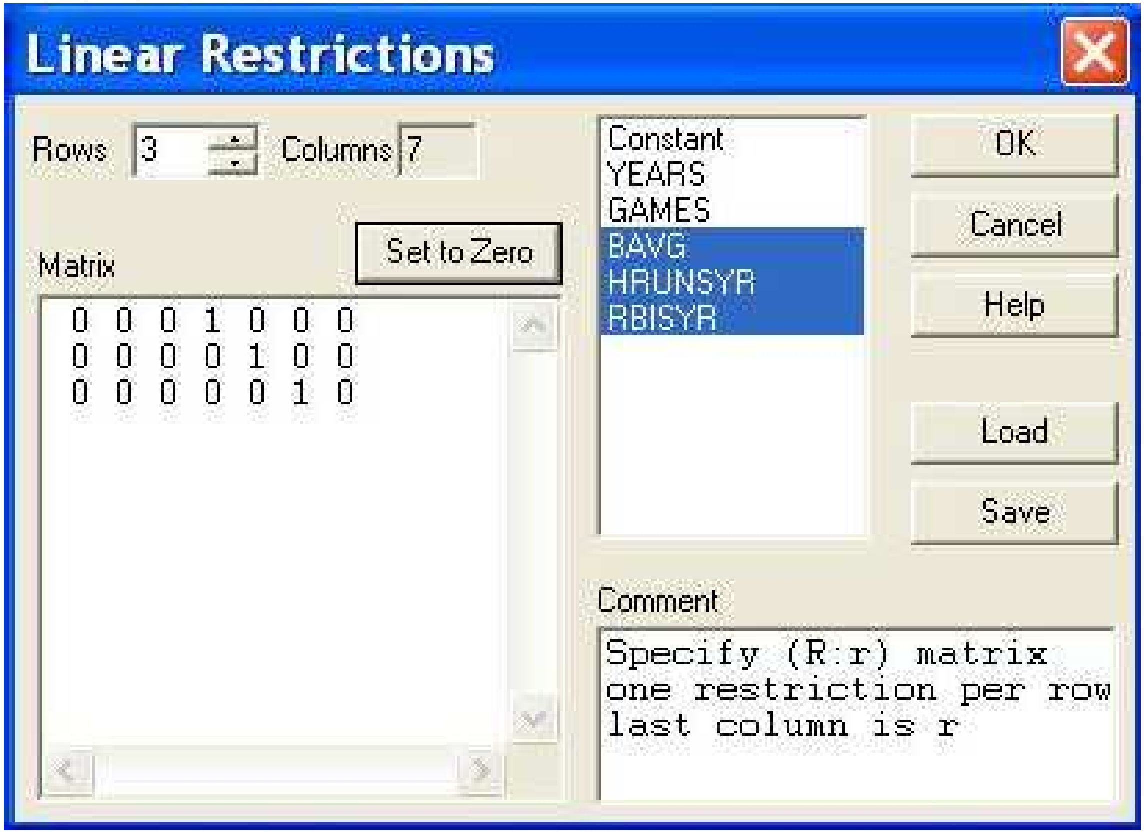
- $H_0 : \beta_1 = 0 \rightarrow \mathbf{R} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}, r = 0.$
- $H_0 : \beta_1 = \beta_2 = 0 \rightarrow \mathbf{R} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$
- $H_0 : \beta_1 + \beta_2 = 1 \rightarrow \mathbf{R} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}, r = 0.$
- $H_0 : \beta_1 + \beta_2 = 1 \rightarrow \mathbf{R} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}, r = 1.$
- $H_0 : \beta_1 - \beta_2 = 0 \rightarrow \mathbf{R} = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}, r = 0.$

GiveWin



PcGive - Cross-sectio...
File
Package
Model
Test
Help
Graphic Analysis...
Predictions
Further Output...
Test...
Test Summary
Exclusion Restrictions...
Linear Restrictions...
General Restrictions...
Omitted Variables...
Store Residuals etc. in Database...
Test item

GiveWin


Linear Restrictions
Rows 3
Columns 7
Set to Zero
Matrix
0 0 0 1 0 0 0
0 0 0 0 1 0 0
0 0 0 0 0 1 0
Constant
YEARS
GAMES
BAVG
HRUNSYR
RBISYR
OK
Cancel
Help
Load
Save
Comment
Specify (R:r) matrix
one restriction per row
last column is r

Exemple: LSALARY

```
Test for linear restrictions (Rb=r):
R matrix
      Constant        YEARS        GAMES
       0.00000      0.00000      0.00000
       0.00000      0.00000      0.00000
       0.00000      0.00000      0.00000
          BAVG      HRUNSYR       RBISYR
        1.0000      0.00000      0.00000
       0.00000       1.0000      0.00000
       0.00000      0.00000       1.0000
r vector
       0.00000      0.00000      0.00000
LinRes F(3,347) =    50.800 [0.0000]**
```

Comment reporter des résultats ?

Tableau 1: Résultats du modèle XYZ

	DEM	FRF
μ	0.0140	0.0102
	(0.038)	(0.014)
ψ_1	-0.0472	-0.0406
	(0.002)	(-0.047)
ω	-0.0246	-0.0134
	(0.006)	(0.068)
BDS(6)	-0.719	-0.825
S(5)	4.723	4.330

Note: les écart-types sont reportés entre parenthèses. $BDS(k)$ est le test BDS avec k degrés de libertés, ...

Chapitre 5: Propriétés asymptotiques des MCO

Propriétés exactes des MCO

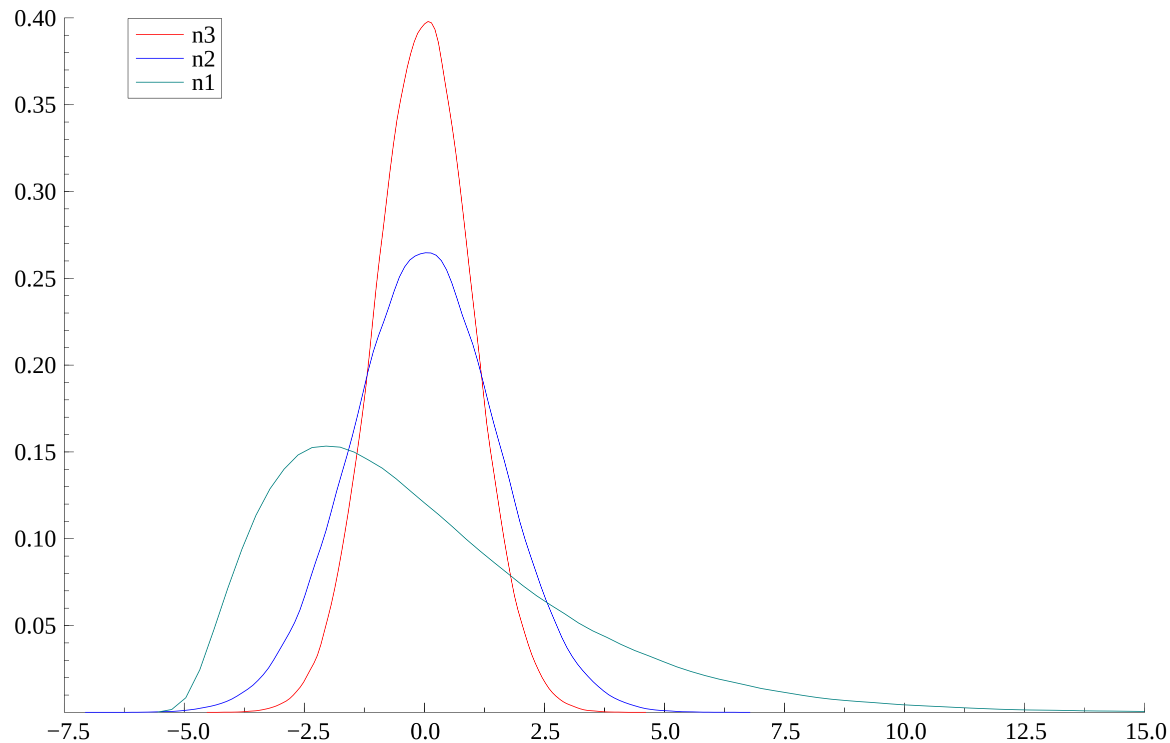
- Pour prouver le caractère non-biaisé et BLUE des MCO nous avons imposé des hypothèses assez fortes (**MLR.1-MLR.5**).
- Idem pour effectuer de l'inférence statistique (**MLR.6**: normalité).
- Ces propriétés sont vraies quel que soit la taille de l'échantillon ($\forall n$).
- $\rightarrow$ On parle dès lors de propriétés exactes, d'échantillon fini, ou encore de petit échantillon.

Propriétés asymptotiques des MCO

- Dans certains cas, le rejet de certaines hypothèses ne signifie pas que les MCO sont non valides.
- Exemple: non-normalité de u.
- $\rightarrow$ En effet, les MCO peuvent être encore valides en grand échantillon *sous des hypothèses plus faibles*.
- $\rightarrow$ Étudier les propriétés statistiques pour n grand = étudier les propriétés asymptotiques.
- Nouveau concept: CONVERGENCE.

Illustration de la consistence

Pour chaque taille d'échantillon $n_1 < n_2 < n_3$ on a une distribution de probabilité de $\hat{\beta}_j$ différente. Si **MLR.1 - MLR.4** sont vérifiées les MCO donnent $E(\hat{\beta}_j) = \beta_j \rightarrow$ pdf centrées en 0. Si $\hat{\beta}_j$ est consistent:



Consistence

- Nouveaux concepts: $plim$ et Loi des grands nombres.
- **Définition**: *plim = Limite en probabilité = valeur vers laquelle un estimateur converge lorsque la taille de l'échantillon tend vers l'infini.* Voir page 741.
- Loi des grands nombres: Soit $Y_1, Y_2, \ldots, Y_n$ des variables aléatoires indépendantes de moyenne μ. Alors:

$$plim(\overline{Y}_n) = \mu.$$

- $\rightarrow$ on peut obtenir une estimation de μ aussi précise que possible en calculant la moyenne empirique d'un échantillon de taille très grande.

Consistence

- **Théorème 5.1**: Sous les hypothèses **MLR.1 - MLR.4**, l'estimateur des MCO $\hat{\beta}_j$ est un estimateur consistent de $\beta_j, \forall i = 0, 1, \ldots, k$.
- **Preuve:** Prenons le modèle simple $y_i = \beta_0 + \beta_1 x_{i1} + u_i$.

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_{i1} - \overline{x}_1) y_i}{\sum_{i=1}^n (x_{i1} - \overline{x}_1)^2} \\ &= \beta_1 + \frac{n^{-1} \sum_{i=1}^n (x_{i1} - \overline{x}_1) u_i}{n^{-1} \sum_{i=1}^n (x_{i1} - \overline{x}_1)^2} \end{aligned}$$

On peut appliquer la loi des grands nombres et les propriétés des $plim$ pour montrer que pour autant que $Var(x_1) \neq 0$,

Suite de la preuve

$$
\begin{aligned}
plim\ \hat{\beta}_1 &= \beta_1 + \frac{Cov(x_1,u)}{Var(x_1)} \\
&= \beta_1, \text{si } Cov(u,x_1) = 0 \text{ CQFD.}
\end{aligned}
$$

- Pour rappel. **MLR.3** $E(u|x_1,\ldots,x_k) = 0 \rightarrow$ moyenne conditionnelle nulle.
- On a vu que $E(u|x_1) = 0$ implique $Cov(u,x_1) = 0$.
- $\rightarrow$ Il est possible de relâcher **MLR.3** pour prouver tout de même que les MCO sont convergents.
- **MLR.3'** $E(u) = 0$ et $Cov(x_j,u) = 0\ \forall j = 1,\ldots,k \rightarrow$ moyenne nulle et correlation nulle.
- $\rightarrow$ Sous les hypothèses **MLR.1 - MLR.2 - MLR.3' - MLR.4**, $\hat{\beta}_j$ est un estimateur convergent de β_j, $\forall j = 1,\ldots,k$.

Normalité asymptotique

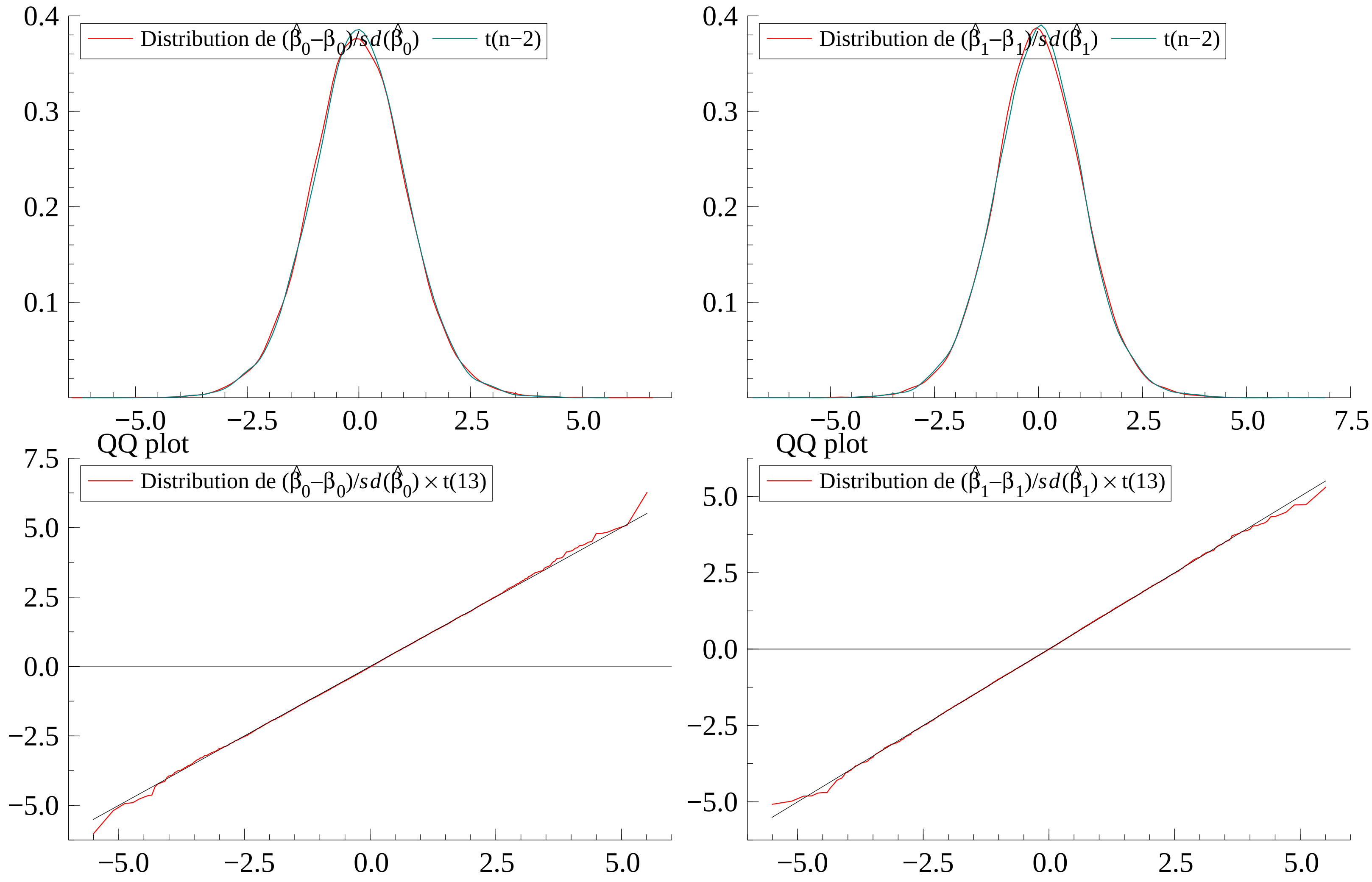
- La normalité ne joue aucun rôle dans le caractère non-biaisé des MCO ni dans leur caractère BLUE.
- Par contre, pour effectuer de l'inférence statistique nous avons supposé que $u \sim N(0, \sigma^2)$ et donc que la distribution de $y|x_1, \ldots, x_k$ est normale.
- $\rightarrow$ distribution symétrique autour de sa moyenne.
- $\rightarrow$ peut prendre des valeurs sur $\Re$.
- $\rightarrow$ plus de 95% des observations sont comprises entre 2 écart-types.

Normalité asymptotique

- Devons nous abandonner les t-stats si u n'est pas normalement distribué ?
- Non: on fait appel au théorème central limite pour conclure que les MCO satisfont la normalité asymptotique.
- Si n grand, $\hat{\beta}_j$ est *approximativement* $N(\beta_j, Var(\hat{\beta}_j))$.
- Illustration par une simulation.
  $\rightarrow n = 15, y = 0.2 + 0.5x_1 + u$, et $u \sim N(0, 1.5)$.
  Estimer $y = \beta_0 + \beta_1 x_1 + u$ 20000 fois et calculer après chaque estimation: $stat_{h,j} = \frac{\hat{\beta}_j - \beta_j}{sd(\beta_j)}$ $\forall h = 1, \ldots 20000$ et $j = 0, 1$. Notez qu'ici $\sigma^2 = 1.5$.

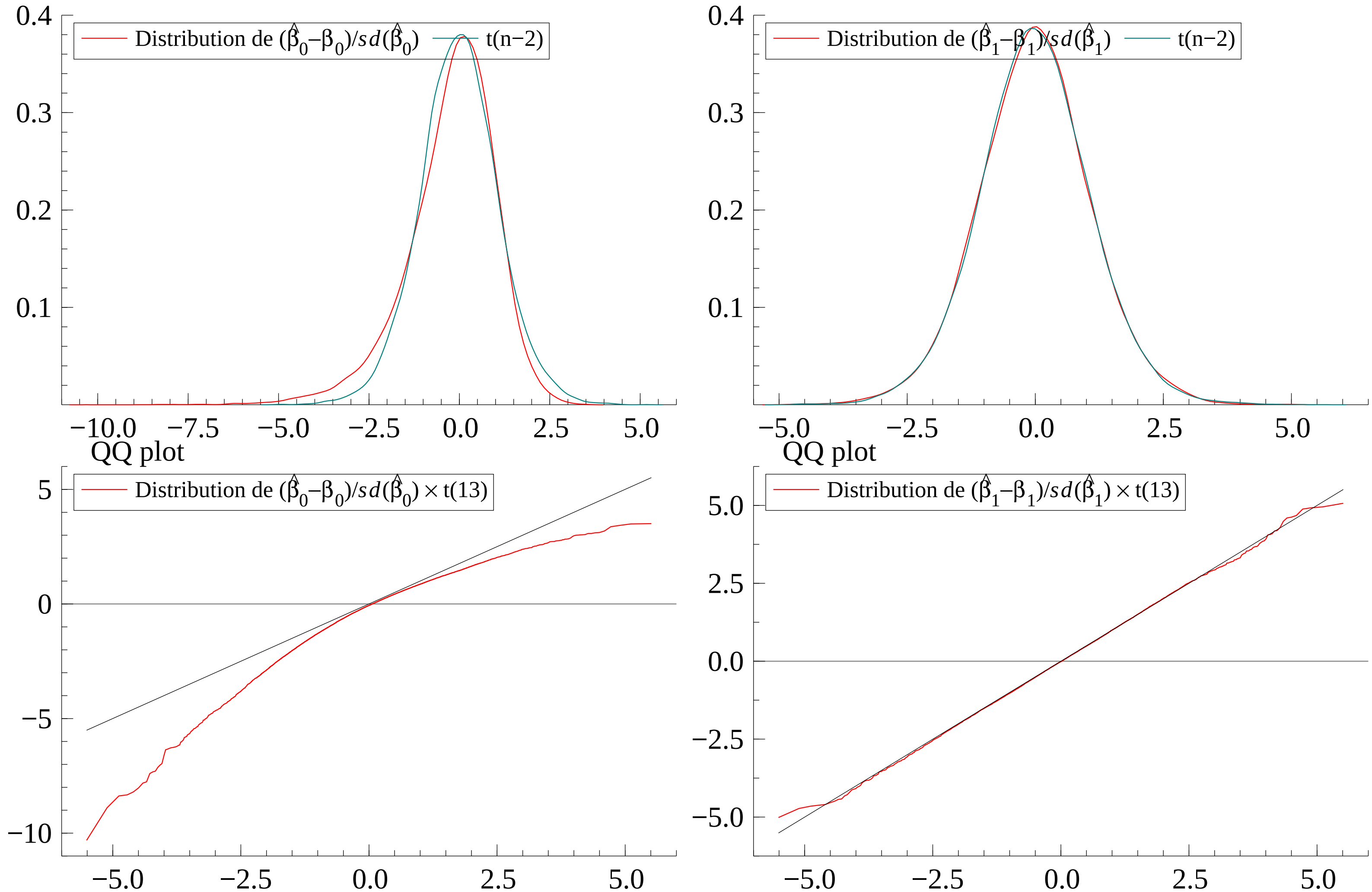
$stat_{h,j} \sim t(n-2)$?

$n = 15, y = 0.2 + 0.5x_1 + u$, $u \sim N(0, 1.5)$ et $\sigma^2 = 1.5$.



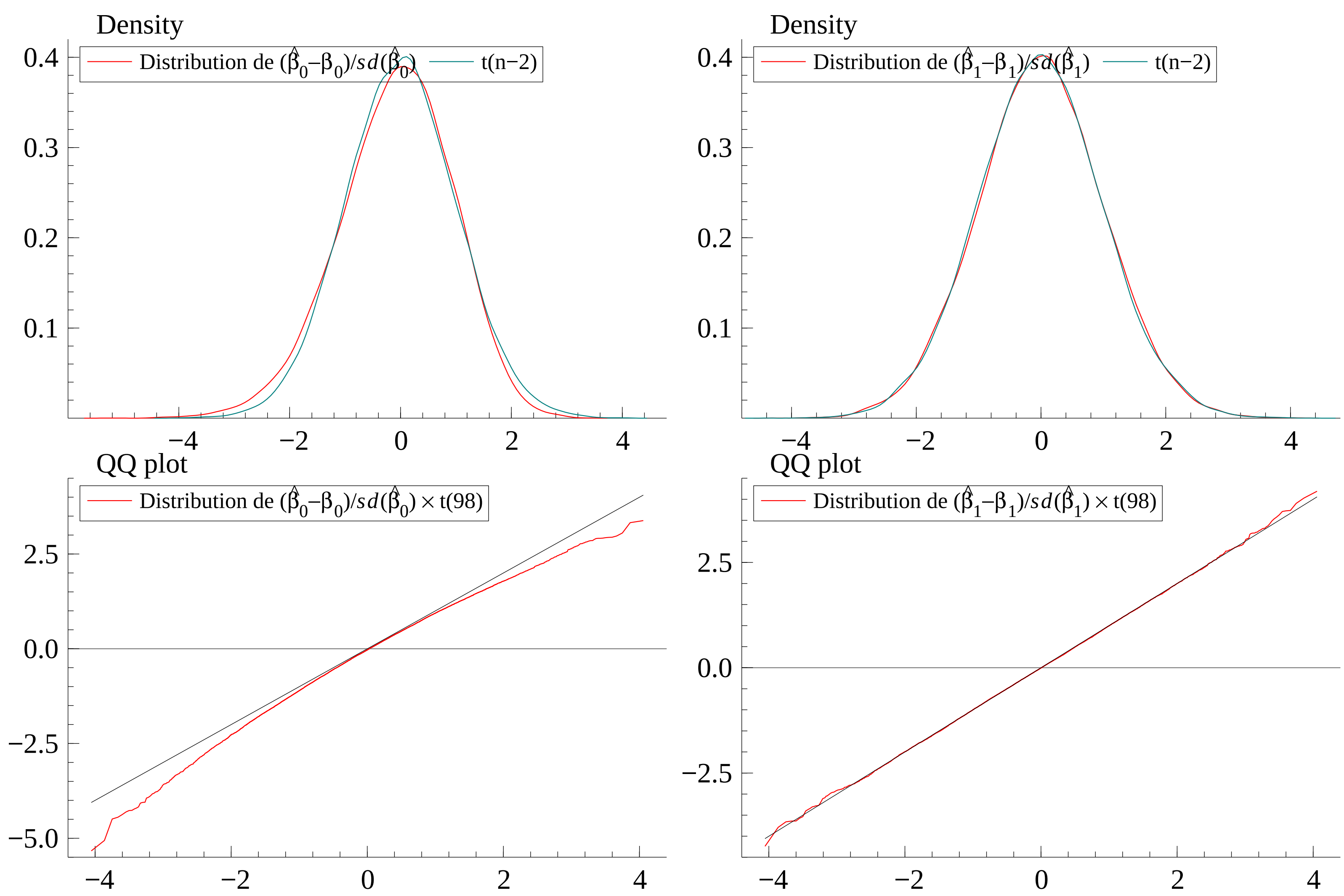
$stat_{h,j} \sim t(n-2)$?

$$n = 15, y = 0.2 + 0.5x_1 + u,\ u \sim \frac{\chi^2(3)-3}{\sqrt{2*3}}.$$



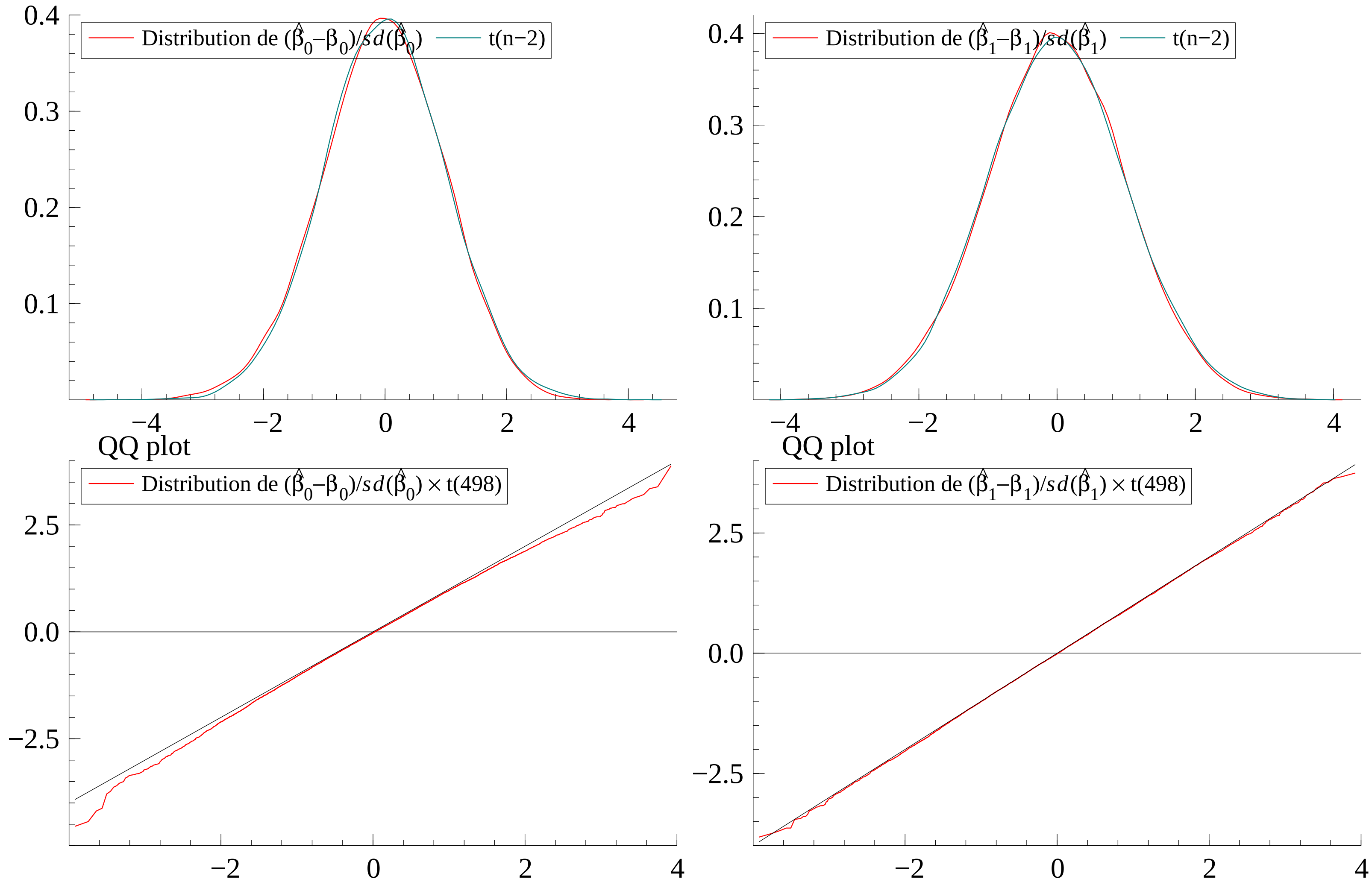
$stat_{h,j} \sim t(n-2)$?

$n = 100, y = 0.2 + 0.5x_1 + u$, $u \sim \frac{\chi^2(3)-3}{\sqrt{2*3}}$.



$stat_{h,j} \sim t(n-2)$?

$n = 500, y = 0.2 + 0.5x_1 + u$, $u \sim \frac{\chi^2(3)-3}{\sqrt{2*3}}$.



Test du multiplicateur de Lagrange

- $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u.$
- $H_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0 \rightarrow q$ restrictions.
- H_1 : au moins un des q paramètres est $\neq 0$.
- Contrairement au F-test, le LM test ne requiert que l'estimation du modèle contraint:

$$y = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1 + \ldots + \widetilde{\beta}_{k-q} x_{k-q} + \widetilde{u}.$$

- Sous $H_0, \widetilde{u}$ devrait être non corrélé avec $x_{k-q+1}, \ldots, x_k$.
- $\rightarrow$ régression auxiliaire:
$\widetilde{u} = \gamma_0 + \gamma_1 x_1 + \ldots + \gamma_{k-q} x_{k-q} + \gamma_{k-q+1} x_{k-q+1} + \ldots + \gamma_k x_k + v.$
- Sous H_0, le R^2 de cette régression auxiliaire devrait être proche de 0.
- $\rightarrow$ Sous H_0, $nR^2 \sim \chi^2(q)$.

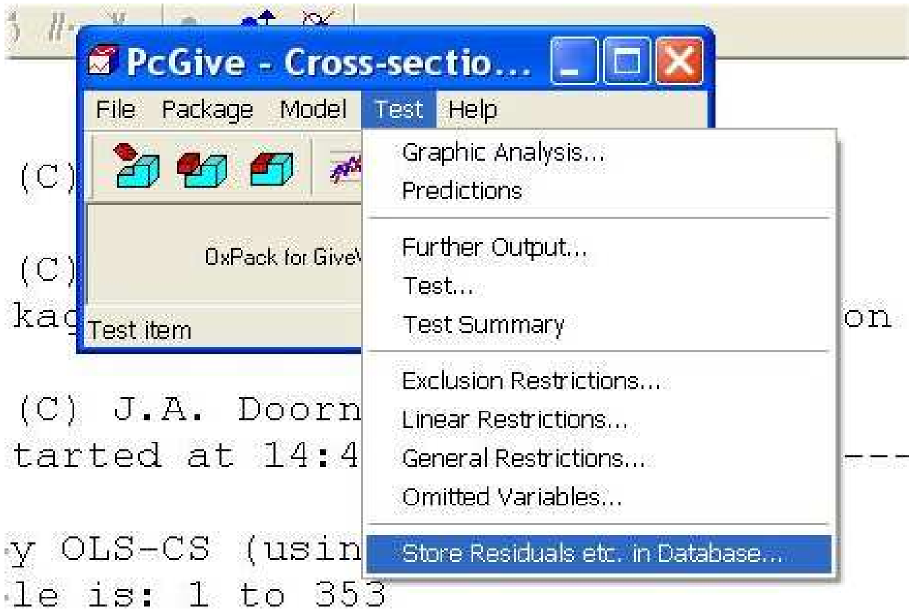
Exemple: LSALARY

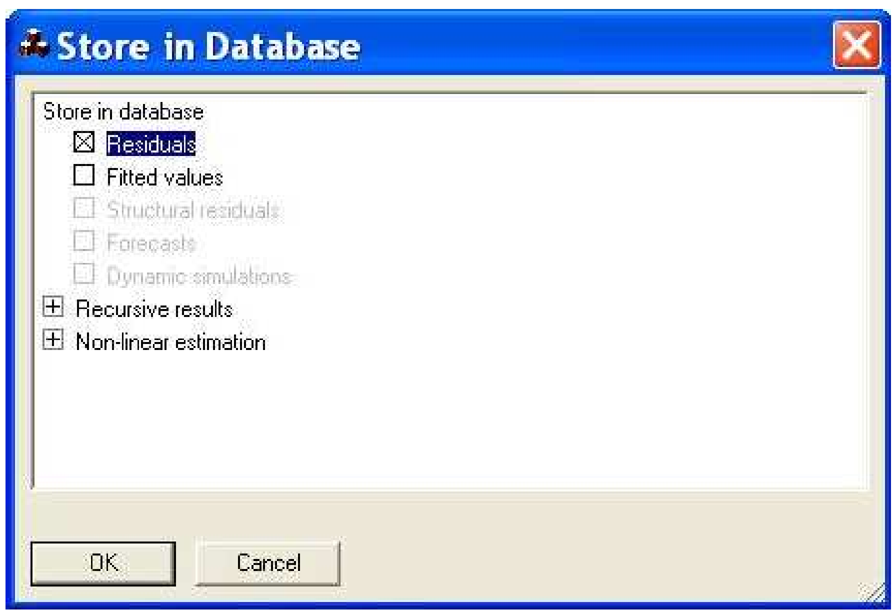
EQ(1) Modelling LSALARY by OLS-CS (using mlb1)

	Coefficient	Std.Error	t-value	t-prob
Constant	11.1924	0.2888	38.8	0.000
YEARS	0.0688626	0.01211	5.68	0.000
GAMESYR	0.0125521	0.002647	4.74	0.000
BAVG	0.000978604	0.001104	0.887	0.376
HRUNSYR	0.0144295	0.01606	0.899	0.369
RBISYR	0.0107657	0.007175	1.50	0.134

	Coefficient	Std.Error	t-value	t-prob
Constant	11.2238	0.1083	104.	0.000
YEARS	0.0713180	0.01251	5.70	0.000
GAMESYR	0.0201745	0.001343	15.0	0.000

GiveWin


PcGive - Cross-sectio...
File
Package
Model
Test
Help
Graphic Analysis...
Predictions
Further Output...
Test...
Test Summary
Exclusion Restrictions...
Linear Restrictions...
General Restrictions...
Omitted Variables...
Store Residuals etc. in Database...
Test item
(C) J.A. Doorn
tarted at 14:4
y OLS-CS (usin
le is: 1 to 353


Store in Database
Store in database
Residuals
Fitted values
Structural residuals
Forecasts
Dynamic simulations
Recursive results
Non-linear estimation
OK
Cancel

GiveWin

Exemple: LSALARY

```
EQ( 3) Modelling res by OLS-CS (using mlb1.in7)

                Coefficient  Std.Error  t-value  t-prob
Constant           -1.29330     0.2940    -4.40   0.000
YEARS              0.262056    0.03726     7.03   0.000
GAMES           -0.00281058  0.0003187    -8.82   0.000
BAVG            0.000551725   0.001126    0.490   0.624
HRUNSYR          -0.0401661    0.01401    -2.87   0.004
RBISYR            0.0456800   0.005380     8.49   0.000

R^2  0.305166       no. of observations         353
```

$nR^2 = 353 * 0.305 = 107.665 >> 11.34$ c-à-d $\chi^2(3)$ avec $\alpha = 1\%$.

$\longrightarrow$ Rejet $H_0 \longrightarrow$ p-valeur $= 0.0000$.

Chapitre 6: Problèmes additionnels dans la régression multiple

Impact des unités de mesure

Effets des changements d'échelle

- Dans le chapitre 2, on a brièvement analysé l'effet d'un changement d'unité de mesure.
  $\longrightarrow$ effet sur les coefficients estimés (intercept et pente).
- Un changement d'unité de mesure n'a pas d'effet sur les statistiques de qualité d'ajustement (R^2).
- Quid sur les autres statistiques telles que écart-types des coefficients, t ou F ?

Exemple préliminaire

- Effet du fait de fumer dans le chef de la maman sur le poids de l'enfant à la naissance.
- $bwght =$ poids du bébés en once. $\rightarrow bwghtlbs =$ poids du bébé en livres (1 livre=16 onces).
- $cigs =$ nombre de cigarettes fumées par jour $\rightarrow packs =$ nombre de paquets fumés par jour (1 paquet $= 20$ cigarettes).
- $faminc =$ revenu de la famille (variable de contrôle).
- On étudie la régression suivante:

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 faminc + u.$$

Résultats de la régression

Var. dépendante	$bwght$	$bwghtlbs$	$bwght$
Var. explicatives			
$cigs$	-0.4634	-0.0289	-
	(0.092)	(0.006)	
$packs$	-	-	-9.268
			(1.832)
$faminc$	0.0927	0.0058	0.0927
	(0.021)	(0.043)	
$Intercept$	116.974	7.311	116.974
	(1.049)	(0.066)	(1.049)
n	1388	1388	1388
R^2	0.0298	0.0298	0.0298
SCR	557.48	2177.67	557.48
SCE	20.063	1.254	20.063

Enseignements sur les coefficients

- Si une femme fume 5 cigarettes par jour, le poids diminue de $5 \times 0.4634 = 2.317$ onces.
- Le $t - stat$ est égal à $-5.06 \rightarrow$ effet très significatif.
- Si on fait la même régression en remplaçant $bwght$ par $bwghtlbs$, c-à-d poids en livre.
- On obtient la régression suivante (2ième colonne du tableau):

$$\frac{\hat{bwght}}{16} = \frac{\hat{\beta}_0}{16} + \frac{\hat{\beta}_1}{16} cigs + \frac{\hat{\beta}_2}{16} faminc.$$

- Chaque coefficient est divisé par 16 $\rightarrow$ le changement d'unité ne change rien du point de vue économique.

Effets en termes de significativité

- Effets sur les $t - stats$: chaque $t - stat$ est identique entre les 2 régressions car les écarts-types ont également été divisés par 16.
- Le changement d'unité n'a pas d'effet en terme de significativité et d'inférence.
- Les R^2 sont également identiques.
- Le changement d'unité n'a pas d'effet sur la qualité d'ajustement du modèle.

Effets d'un changement de mesure

- Quid si on mesure le fait de fumer en terme de paquets plutôt que de cigarettes ?
- Pour les variables inchangées, rien ne change.
- Pour les variables dont on change l'unité de mesure $\rightarrow$ les coefficients et les écart-types sont multipliés.
- $$\hat{bwght} = \hat{\beta}_0 + (20\hat{\beta}_1)cigs + \hat{\beta}_2 faminc.$$

  $\rightarrow$ Rien ne change en terme d'interprétation économique.
- $\rightarrow$ Rien ne change en terme de significativité et d'ajustement du modèle.

Transformation logarithmique

- Lorsque les variables sont exprimées en logarithmes $\rightarrow$ Les coefficients de pente estimés ne sont pas sensibles à des changements d'unité de mesure. Pourquoi ?
- Dans ce cas, les coefficients sont des élasticités $\rightarrow$ la valeur d'une élasticité ne depend pas de l'unité de mesure de y ou de x.
- Par contre, l'intercept se modifie. Exemple: si on multiplie y_i par c_1, alors, l'intercept après transformation devient $log(c_1) + \hat{\beta}_0$.

Coefficients betas

- Parfois, il est intéressant de standardiser les variables du modèle de régression de base pour 2 raisons différentes.
- Certaines variables ont une unité de mesure difficilement interprétable. Exemple: résultats de tests scolaires $\rightarrow$ cela depend de la difficulté de la matière, de l'enseignant, etc.
- On peut alors comparer les effets de différentes variables exprimées dans des unités de mesure complètement différentes.
- Le but est alors de calculer l'effet d'une augmentation d'un écart-type de x en terme de variation d'écart-type de y.

Coefficients betas

- Soit le modèle de régression de base:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + ... + \hat{\beta}_k x_{ki} + \hat{u}_i.$$

- En soustrayant la moyenne, on obtient:

$$y_i - \overline{y} = \hat{\beta}_1(x_{1i} - \overline{x}_1) + \hat{\beta}_2(x_{2i} - \overline{x}_2) + ... + \hat{\beta}_k(x_{ki} - \overline{x}_k) + \hat{u}_i.$$

- En divisant par l'écart-type de y, on obtient:
$\frac{y_i - \overline{y}}{\hat{\sigma}_y} = \frac{\hat{\sigma}_1}{\hat{\sigma}_y}\hat{\beta}_1\frac{(x_{1i} - \overline{x}_1)}{\hat{\sigma}_1} + \frac{\hat{\sigma}_2}{\hat{\sigma}_y}\hat{\beta}_2\frac{(x_{2i} - \overline{x}_2)}{\hat{\sigma}_2} + ... + \frac{\hat{\sigma}_k}{\hat{\sigma}_y}\hat{\beta}_k\frac{(x_{ki} - \overline{x}_k)}{\hat{\sigma}_k} + \frac{\hat{u}_i}{\hat{\sigma}_y}$.

Coefficients betas- suite

- Chaque variable est remplacée par sa valeur standardisée z, avec $z_i = \frac{(x_i - \overline{x}_i)}{\hat{\sigma}_i}$.
- On obtient donc:

$$z_y = \hat{b}_1 z_1 + \hat{b}_2 z_2 + ... + \hat{b}_k z_k + error.$$

- Les coefficients standardisés sont donc:

$$\hat{b}_j = \frac{\hat{\sigma}_i}{\hat{\sigma}_y} \hat{\beta}_j.$$

- $\rightarrow$ Si x_j augmente d'un écart-type, alors $\hat{y}$ augmente de $\hat{b}_j$ écart-type(s) $\rightarrow$ les $\hat{b}_j$ sont directement comparables.
- Par contre, les $t - stats$ des coefficients restent inchangés dans la régression standardisée.

Coefficients betas - Exemple

- Soit l'impact de différentes variables sur le prix des maison $(price)$, dont le taux de pollution dans l'air (nox).
- Avec une régression standardisée, on a:

$$\begin{aligned}\hat{z}price &= -.340znox - .143zcrime + .514zrooms \\ &\quad -.235zdist - .270zratio.\end{aligned}$$

- Une augmentation de la pollution d'1 écart-type par rapport à la population des villes entraîne une diminution de 0.34 écart-type du prix des maisons.
- En termes relatifs par rapport à la population, on voit que la pollution exerce une effet plus important que le taux de criminalité.

Formes fonctionnelles

Formes fonctionnelles

Il existe différentes formes du modèle de régression.

- Certaines variables sont mesurées en logarithmes.
- Certaines variables apparaissent avec des exposants différents de 1. Exemple: $x_i^2 \rightarrow$ forme quadratique.
- On peut insérer dans le modèle des termes d'interaction, c-à-d des variables qui sont le produit de plusieurs variables explicatives.
- Comment interpréter les coefficients estimés ?

Transformation logarithmique

- Prenons l'exemple du modèle de régression suivant:

$$log(y) = \hat{\beta}_0 + \hat{\beta}_1 log(x_1) + \hat{\beta}_2 x_2.$$

- $\hat{\beta}_1$ s'interprète comme une élasticité: l'effet d'1 variation en pourcentage de x_1 en terme de variation en pourcentage de y.
- $\hat{\beta}_2$ est une semi-élasticité: l'effet approximatif d'1 augmentation d'une unite de x_2 en terme de pourcentage de y. L'effet exact en pourcentage pour $\Delta x_2 = 1$ est donné par: $\%\hat{\Delta} y = 100 \times [\exp(\hat{\beta}_2) - 1]$.

Transformation logarithmique

- Prenons l'exemple du modèle de régression expliquant le prix des habitations ($price$) par le taux de pollution (nox) et le nombre de chambres ($rooms$):

$$log(price) = \hat{\beta}_0 + \hat{\beta}_1 log(nox) + \hat{\beta}_2 rooms.$$

- On obtient: $\hat{\beta}_0 = 9.23(0.19)$; $\hat{\beta}_1 = -0.718(0.066)$; $\hat{\beta}_2 = 0.306(0.019)$; $n = 506$;$R^2 = 0.514$.
- Lorsque le taux de pollution augmente de 1%, le prix des maison baisse de 0.718%.
- Lorsqu'une maison comporte 1 chambre en plus, l'effet approximatif sur le prix est de $100 \times 0.306 = 30.6\%$. L'effet exact est donné par $100 \times [exp(0.306) - 1] = 35.8\%$.

Avantage des logarithmes

- Quand $y > 0$, le fait de prendre $log(y)$ induit une distribution plus gaussienne, ce qui permet de se rapprocher des hypothèses du modèle linéaire classique; Exemple: salaires.
- Le logarithme réduit l'écart maximal des valeurs possibles des variables $\rightarrow$ rend les estimations moins sensibles aux valeurs extrêmes. Parfois, le logarithme atténue les problèmes d'hétéroscédasticité.
- En pratique, on prend souvent le logarithmes des variables exprimées en unités monétaires (Ex.: salaire), des variables faisant référence à un nombre de personnes (Ex.: population). Par contre, les variables exprimées en nombre d'années demeurent dans leur forme originelle (Ex.: éducation).

Inconvénients des logarithmes

- Lorsqu'une variable y peut prendre des valeurs nulles, on prend alors $log(1+y) \rightarrow$ Attention à l'interprétation !!!.
- On ne peut pas comparer le R^2 d'un modèle du type $y = \ldots$ et $log(y) = \ldots$.
  $\rightarrow$ le R^2 dépend de l'expression de y.
- Pour les variables déjà exprimées en pourcentage (Ex.: chômage), ne pas confondre changement en pourcentage et en point de pourcentage. Exemple: si y augmente de 8 à 9% $\rightarrow log(9) - log(8) = 0.118$ c-à-d augmentation de 1 point de %age $=$ augmentation de 12.5 % par rapport au niveau initial ou encore en prenant l'approximation logarithmique: augmentation de environ 11.8%.

Fonctions quadratiques

Termes quadratiques

- On peut introduire des termes quadratiques de manière à appréhender des effets marginaux croissant ou décroissant.
- $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2$.
- L'effet marginal de x_1 sur $\hat{y}$ est le suivant:

$$\frac{\Delta \hat{y}}{\Delta x_1} = \hat{\beta}_1 + 2\hat{\beta}_2 x_1.$$

- $\rightarrow$ Si $\hat{\beta}_1 > 0$ et si $\hat{\beta}_2 < 0$, l'effet de x_1 sur $\hat{y}$ prend une forme concave: l'effet de x_1 est positif mais décroissant au fur et à mesure que x_1 augmente. Exemple: effet de l'expérience professionnelle sur le salaire.

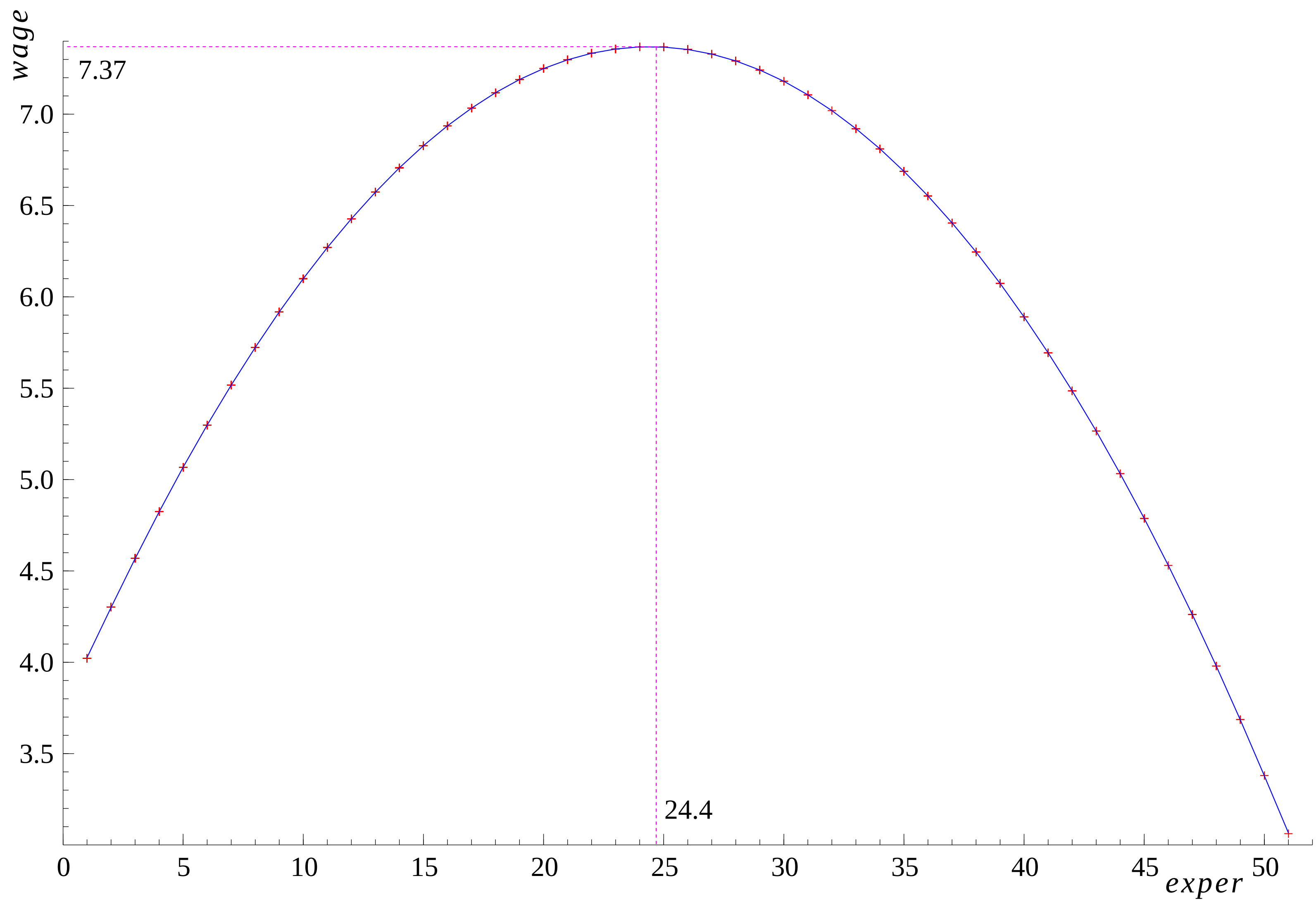
Exemple d'estimation

$$wage = \hat{\beta}_0 + \hat{\beta}_1 exper + \hat{\beta}_2 exper^2 + u.$$

	Coefficient	Std.Error	t-value	t-prob
Constant	3.72541	0.3459	10.8	0.000
EXPER	0.298100	0.04097	7.28	0.000
EXPERSQ	-0.00612989	0.0009025	-6.79	0.000

```
R^2        0.0927693    F(2,523) =  26.74 [0.000]**
no. obs.         526    no. of par.                3
```

Relation quadratique



Termes quadratiques-suite

- $\rightarrow$ L'effet de l'experience n'est pas constant.
  Exemple:
  si $exper = 1, \frac{\Delta \hat{y}}{\Delta exper} = 0.298 - 2(0.0061)(1) \simeq 0.286$;
  si $exper = 10, \frac{\Delta \hat{y}}{\Delta exper} = 0.298 - 2(0.0061)(10) \simeq 0.176$.
- Si $\hat{\beta}_1 < 0$ et si $\hat{\beta}_2 < 0$, l'effet de x_1 sur y prend une forme convexe: l'effet de x_1 est négatif mais croissant au fur et à mesure que x_1 augmente. Exemple: effet de la pollution d'une ville sur le prix des maisons.

Termes d'interaction

Termes d'interaction

- Soit l'équation du prix des maisons:

$$\begin{aligned} price &= \beta_0 + \beta_1 sqrft + \beta_2 bdrms \\ &\quad + \beta_3 (sqrft \times bdrms) + \beta_4 bthrms + u. \end{aligned}$$

- Le terme d'interaction permet d'appréhender un effet non linéaire de $bdrms$ (nombre de chambres) et de faire dépendre (une partie de) cet effet d'une autre variable explicative, c-à-d la surface de la maison ($sqrft$).
- $\rightarrow$ L'effet d'1 chambre supplémentaire sur le prix est: $\frac{\Delta \hat{price}}{\Delta bdrms} = \hat{\beta}_2 + \hat{\beta}_3 sqrft \rightarrow$ L'interprétation est plus délicate. Exemple: $\hat{\beta}_2$ capture l'effet d'une chambre supplémentaire dans une maison avec une surface nulle !

Qualité d'ajustement et selection de régresseurs

R-carré ajusté

- Le critère du R^2 comme critère de sélection de modèle n'est pas toujours un bon critère.
- Un faible R^2 ne signifie pas que les facteurs omis dans la régression et présents dans u sont corrélés avec les facteurs inclus dans la régression $\rightarrow$ Les estimateurs peuvent être non biaisés.
- Un faible R^2 implique que la variance de $\hat{u}$ est élevée par rapport à la variance de y mais ceci peut être compensé par un nombre d'observations assez élevé.
- Par contre, le changement du R^2 au fur et à mesure que des variables explicatives sont ajoutées est instructif.

R-carré ajusté

- $R^2 = 1 - \left[\left(\frac{SSR}{n}\right) / \left(\frac{SST}{n}\right)\right]$, où SSR: Somme des carrés des résidus et SST: Somme totale des carrés.
- Le R^2 est en fait un estimateur de $1 - (\frac{\sigma_u^2}{\sigma_y^2})$, c-à-d la proportion de la variance de y dans la population expliquée par les variables indépendantes du modèle.
- σ_u^2 est estimé de manière biaisée par $\frac{SSR}{n}$.
  $\rightarrow$ à remplacer par $\frac{SSR}{n-k-1}$.
- σ_y^2 est estimé de manière biaisée par $\frac{SST}{n}$.
  $\rightarrow$ à remplacer par $\frac{SST}{n-1}$.
- On obtient alors le R^2 ajusté :
  $\overline{R}^2 = 1 - \left[\left(\frac{SSR}{n-k-1}\right) / \left(\frac{SST}{n-1}\right)\right]$.

Avantage du R-carré ajusté

- Le R-carré ajusté est un estimateur corrigé du R-carré classique.
- On peut utiliser le $\overline{R}^2$ pour sélectionner de manière successive des régresseurs potentiels.
- Lorsque l'on ajoute un régresseur supplémentaire, le R^2 classique augmente toujours.
- Mais dans le $\overline{R}^2$, il y a une pénalité car son dénominateur inclut k, le nombre de régresseur.
- De manière nette, le $\overline{R}^2$ n'augmente que si le régresseur supplémentaire apporte un pouvoir explicatif suffisant ($|t - stat| > 1$).

Sélection avec le R-carré ajusté

- Le R-carré ajusté permet de comparer des modèles non-emboîtés. Un modèle est dit emboîté si un des modèles est un cas particulier de l'autre.
- Exemple de modèles non emboîtés: salaire des joueurs de base-ball.
- $log(salary) =$
  $\beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + u.$
  $\rightarrow \overline{R}^2 = .6211.$
- $log(salary) =$
  $\beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 rbisyr + u.$
  $\rightarrow \overline{R}^2 = .6226.$

Sélection avec le R-carré ajusté

- On peut comparer aussi des modèles avec des formes fonctionnelles différentes. Exemple: Dépenses R&D et ventes.
  $rdintens = \beta_0 + \beta_1 log(sales).$
  $rdintens = \beta_0 + \beta_1 sales$.
- Attention: on ne peut pas comparer les modèles avec une variables dépendantes différente. La valeur du $\overline{R}^2$ et du R^2 dependent directement de l'unité de mesure de la variable dépendante.
- Exemple: on ne peut pas comparer 2 fonctions de salaire du type $salaire = \ldots$ et $log(salaire) = \ldots$.

Prévision et analyse des résidus

Prévision de y

- Modèle de régression du style:
  $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$.
- Pour des valeurs précises de $x_1 = c_1, x_2 = c_2, ..., x_k = c_k$, on peut calculer la valeur prédite de y:

$$\begin{aligned} \hat{\theta}_0 &= E(y|x_1 = c_1, \ldots, x_k = c_k) \\ &= \hat{\beta}_0 + \hat{\beta}_1 c_1 + \hat{\beta}_2 c_2 + \ldots + \hat{\beta}_k c_k. \end{aligned}$$

- Comme les $\hat{\beta}_i$ sont estimés et donc aléatoires, il faut construire un IC pour $\hat{\theta}_0$ du type $\hat{\theta}_0 + 2se(\hat{\theta}_0)$.
  $\rightarrow$ il faut estimer $se(\hat{\theta}_0)$.
- On peut réécrire le modèle :

$$y = \theta_0 + \beta_1(x_1 - c_1) + \beta_2(x_2 - c_2) + ... + \beta_k(x_k - c_k) + u.$$

Exemple: Wage1

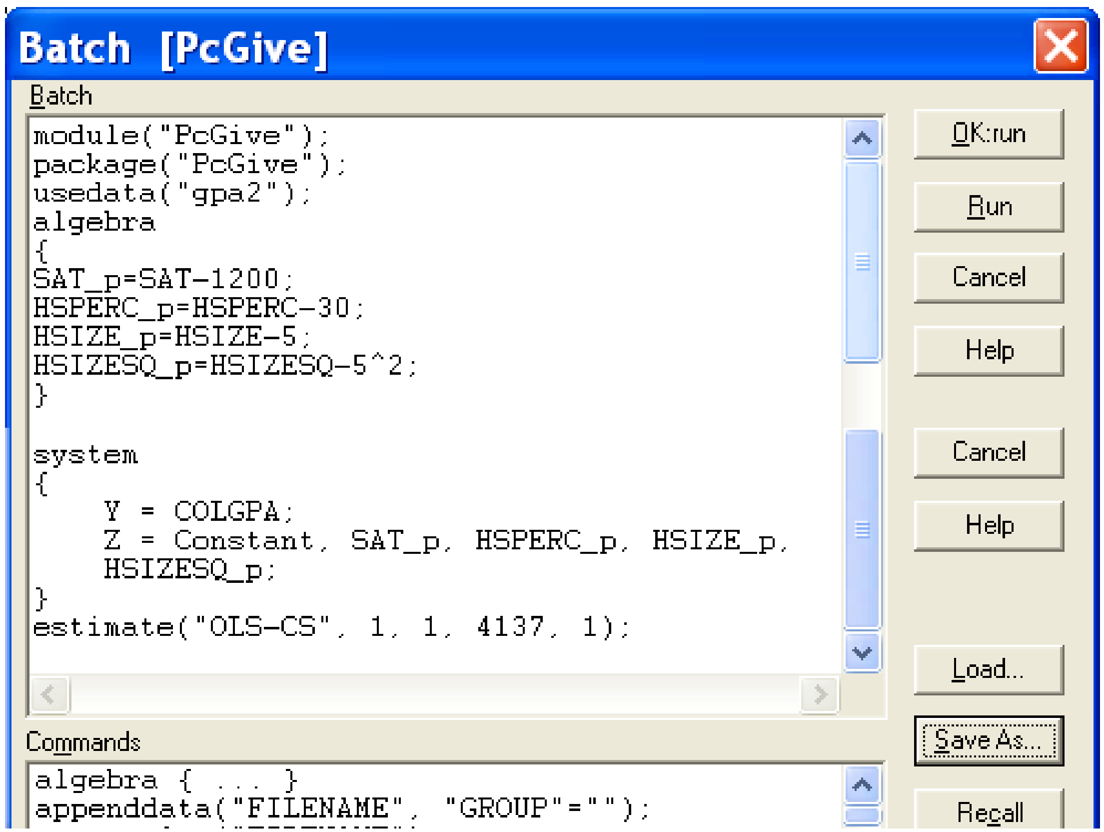
```
EQ( 1) Modelling COLGPA by OLS-CS (using gpa2)
(COLGPA=College Grade Point Average)
         Coefficient  t-prob
Constant      1.49265   0.000 (SAT=Exam. entrée)
SAT        0.00149250   0.000 (SAT score)
HSPERC     -0.0138558   0.000 (high sch. perc.)
HSIZE      -0.0608815   0.000 (size grad. class)
HSIZESQ    0.00546030   0.016 (HSIZE^2)

sigma               0.559864 RSS            1295.16517
R^2                 0.278136 F(4,4132) = 398[0.000]
log-likelihood     -3467.93 DW                   1.88
no. of observations 4137 no.                        5
```

Prévision de y

Prévision:

$y = \theta_0 + \beta_1(x_1 - 1200) + \beta_2(x_2 - 30) + \beta_3(x_3 - 5) + \beta_4(x_4 - 5^2) + u$

`ALT + B` ou `Tools/Batch Editor...`



Exemple

```
EQ( 4) Modelling COLGPA by OLS-CS (using gpa2)

                Coefficient  Std.Error
Constant            2.70008    0.01988
SAT_p            0.00149250 6.521e-005
HSPERC_p         -0.0138558  0.0005610
HSIZE_p          -0.0608815    0.01650
HSIZESQ_p        0.00546030   0.002270
```

$\rightarrow$ IC à 95 % pour $\hat{\theta}_0 = 2.70 \pm 1.96(0.020) = (2.66; 2.74)$.

Prévision de y

- On a donc calculé $E(y|x_1 = c_1, \ldots, x_k = c_k)$, c-à-d la valeur prévue de la moyenne de y pour des valeurs de x données ainsi que son intervalle de confiance (IC).
- L'IC est relatif à la moyenne à une personne moyenne, ce qui est différent d'un IC pour un individu donné.
- $\rightarrow$ pour cela il faut tenir compte d'une autre source d'incertitude: la variance de u.
- Soit y^0 la valeur de y pour un individu (hypothétique) hors de l'échantillon et $x_1^0, \ldots, x_k^0$ les valeurs des variables indépendantes de cet individu.
- $\rightarrow y^0 = \beta_0 + \beta_1 x_1^0 + \beta_2 x_2^0 + \ldots + \beta_k x_k^0 + u^0.$

Prévision de y

- L'erreur de prévision pour cet individu est $\hat{e}^0 = y^0 - \hat{y}^0 = (\beta_0 + \beta_1 x_1^0 + \beta_2 x_2^0 + \ldots + \beta_k x_k^0 + u^0) - \hat{y}^0$.
- Hors $E(\hat{\beta}_0 + \hat{\beta}_1 x_1^0 + \hat{\beta}_2 x_2^0 + \ldots + \hat{\beta}_k x_k^0 + \hat{u}^0) = \beta_0 + \beta_1 x_1^0 + \beta_2 x_2^0 + \ldots + \beta_k x_k^0$ et donc comme $E(u^0) = 0$, $\rightarrow E(\hat{e}^0) = 0$.
- Notons que $Corr(\hat{\beta}_j, u^0) = 0$, $\forall j = 1, \ldots, k$.
- $\rightarrow Var(\hat{e}^0) = Var(\hat{y}^0) + Var(u^0) = Var(\hat{y}^0) + \sigma^2$.
- $\rightarrow \hat{Var}(\hat{e}^0) = Var(\hat{y}^0) + \hat{\sigma}^2$ et donc $\hat{sd}(\hat{e}^0) = \sqrt{\hat{Var}(\hat{e}^0)}$.

Exemple

- Dans l'exemple précédant on a calculé un intervalle de prévision pour $COLGPA$ pour la moyenne des étudiants qui ont les caractéristiques suivantes: $SAT = 1200, \ldots$
- Nous voulons maintenant un IC à 95% pour un étudiant particulier qui a les mêmes caractéristiques.
- $\rightarrow \hat{\theta}_0 \pm 1.96 \hat{sd}(\hat{e}^0)$, où $\hat{sd}(\hat{e}^0) = \sqrt{0.020^2 + 0.560^2} \simeq 0.560$.
- $\rightarrow (1.60; 3.80)$ par rapport à $(2.66; 2.74)$.

Prévision de logarithmes

- Comment prédire y à partir de :
  $log(y) = \beta_0 + \beta_1 x_1^0 + \beta_2 x_2^0 + ... + \beta_k x_k^0 + u$?
- Une estimation de $\hat{y}$ par $\exp(\hat{log}(y))$ est incorrecte (sous-estimation).
- En fait, $\hat{y} = \exp(\hat{\sigma}^2/2)\exp(\hat{log}(y))$.
- Il faut donc utiliser $\hat{\sigma}^2$ pour ajuster la prévision.

Chapitre 7: Régression multiple avec variables binaires ou dummy

- Il existe 2 types de variables.
- Variables incorporant une information quantitative. Exemple: Education, salaire, emploi, ventes d'une firme.
- Variables incorporant une information qualitative. Exemple: Sexe d'une personne, race, type d'industrie, localisation.
- On va voir comment traiter les variables incorporant une information qualitative agissant comme des variables exogènes.
- Lorsqu'elles sont endogènes $\longrightarrow$ traitement économétrique différent (probit, logit, etc.).

Comment transcrire une information qualitative ?

Variable binaire ou dummy

- Parfois, l'information est binaire. Ex: homme ou femme, on possède un permis de conduire ou non, etc.
- $\rightarrow$ Variable dummy $=$ variable avec 2 valeurs possibles: 0 ou 1.
- Le choix du standard (valeur 0) est important pour l'interprétation. Exemple : variable $gender$ (genre): $gender = 1$ si femme, $gender = 0$ si homme.

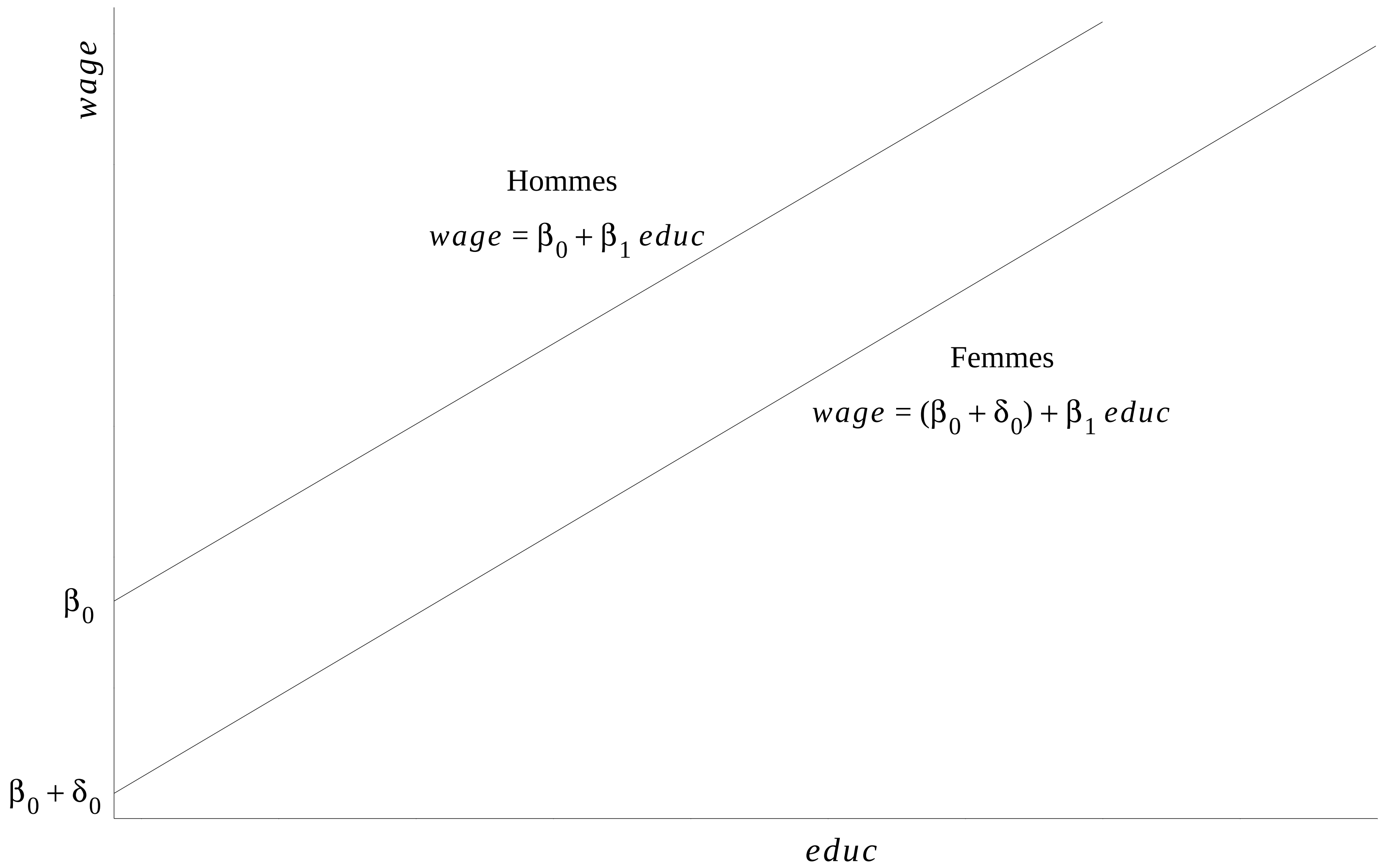
Cas d'une seule variable dummy

- Supposons la fonction de salaire suivante:

$$wage = \beta_0 + \beta_1 educ + \delta_0 female + u.$$

- $female = 1$ si l'individu est une femme, $female = 0$ si l'individu est un homme.
- δ_0 va capturer la différence de salaire entre femme et homme pour un même niveau d'éducation $\longrightarrow \delta_0 = E(wage|female = 1, educ) - E(wage|female = 0, educ) = (\beta_0 + \delta_0) - \beta_0$.
- $\delta_0 < 0 \longrightarrow$ discrimination défavorable aux femmes.

Exemple où $\delta < 0$



Piège à éviter

- On ne peut pas introduire simultanément 2 variables dummy mutuellement exclusives. Exemple $female$ et $male$.
- Pourquoi ? $female + male = 1 \longrightarrow male$ est une fonction linéaire parfaite de $female \Rightarrow$ rejet de **MLR.4**.
- Le choix de la variable standard n'importe que pour l'interprétation.

Exemple: SALARY

```
EQ( 1) Modelling WAGE by OLS-CS (using wage1.in7)

              Coefficient Std.Error  t-value  t-prob
Constant         -1.56794     0.7246   -2.16   0.031
EDUC             0.571505    0.04934    11.6   0.000
EXPER           0.0253959    0.01157    2.20   0.029
TENURE           0.141005    0.02116    6.66   0.000
FEMALE           -1.81085     0.2648   -6.84   0.000

sigma                2.95757 RSS              4557.30771
R^2                 0.363541 F(4,521) =   74.4 [0.000]
no. of obs               526 no. of par.               5
mean(WAGE)            5.8961 var(WAGE)            13.613
```

Variables dummy, suite

- Souvent, les variables dummy peuvent refléter un choix ou une action de la part d'un individu
$\rightarrow$ lien de causalité.
- Exemple 1: $PC = 1$ si l'étudiant détient un PC, 0 sinon.
$\rightarrow$ effet sur les résultats scolaires.
- Exemple 2: $grant = 1$ si une firme reçoit une bourse pour la formation de ses employés, 0 sinon.
$\rightarrow$ effet sur le nombre d'heures de formation des employés.

Modèles log-niveau

- Comment interpréter le coefficient relatif à une variable dummy quand la variable indépendante est exprimée en log ?
- $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + \delta_0 female + u$.
- Le différentiel de salaire moyen entre un homme et une femme est de $[exp(\delta_0) - 1]\%$.
- Exemple: $\delta_0 = -.297 \rightarrow$ une femme gagne en moyenne 25.7 % en moins qu'un homme à caractéristiques (présentes dans le modèle) identiques.

Plusieurs variables binaires

- On peut inclure plusieurs variables dummy binaires dans la même régression.
- Exemple: $female = 1$ si femme; $married = 1$ si marrié(e).
- Si introduites séparément, effet du marriage est identique entre hommes et femmes. Comment introduire de l'interaction ? $\rightarrow$ 4 groupes: femme celib, homme celib, femme mariée, homme marié.
- Si g catégories, il faut introduire $g - 1$ variables dummy $+$ un intercept qui capture l'effet pour la catégorie de base.
- $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \delta_1 marrfemale + \delta_2 marmale + \delta_3 singfemale + u \rightarrow \beta_0$ fait reference au salaire des hommes célibataires.

Information ordinale

- Parfois, la variable dummy est une variable ordinale .
- Exemple : rating de credit des villes : CR prenant 5 valeurs possibles de 0 (le plus mauvais débiteur) à 4 (meilleur débiteur).
- Modèle : $Txint = \beta_0 + \beta_1 CR + ... + u$.
  $\rightarrow$ problème : effet sur le taux d'intérêt $Txint$ d'1 passage de 0 à 1 pas équivalent au passage de 1 à 2, de 2 à 3, etc.

Information ordinale

- Solution : créer une variable dummy binaire par valeur: $CR1 = 1$ si $CR = 1$, $CR2 = 1$ si $CR = 2$ etc.
- $Txint = \beta_0 + \delta_1 CR_1 + \delta_2 CR_2 + \delta_3 CR_3 + \delta_4 CR_4 + ... + u$. $\rightarrow$ Tout s'interprète par rapport à $CR = 0$.
- Remarque : le premier modèle est un modèle restreint du second avec restriction suivante: $CR = CR_1 + 2CR_2 + 3CR_3 + 4CR_4$.

Pentes et intercepts différents

- On peut utiliser des variables dummy pour tester des pentes et des intercepts différents entre catégories. Exemple : salaire, education et sexe.
- $log(wage) = (\beta_0 + \delta_0 female) + (\beta_1 + \delta_1 female)educ + u$.
- Equation estimable correspondante: $log(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female \times educ + u.$
- Hypothèse : rendement de l'education identique selon le sexe: $\delta_1 = 0$.
- Hypothèse : salaire moyen identique selon le sexe: $\delta_0 = \delta_1 = 0 \longrightarrow$ test en F.

Exemple: SALARY

$log(wage) = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + \beta_4 exper + \beta_5 exper^2 + \beta_6 tenure + \beta_7 tenure^2 + u.$

EQ(1) Modelling LWAGE by OLS-CS

	Coefficient	Std.Error	t-value	t-prob
Constant	0.388806	0.1187	3.28	0.001
FEMALE	-0.226789	0.1675	-1.35	0.176
EDUC	0.0823692	0.0085	9.72	0.000
FEMALEEDUC	-0.0055645	0.0131	-0.43	0.670
EXPER	0.0293366	0.0050	5.89	0.000
EXPERSQ	-0.0005804	0.0001	-5.40	0.000
TENURE	0.0318967	0.0069	4.65	0.000
TENURSQ	-0.0005899	0.0002	-2.51	0.012

sigma 0.4001 RSS 82.9215951

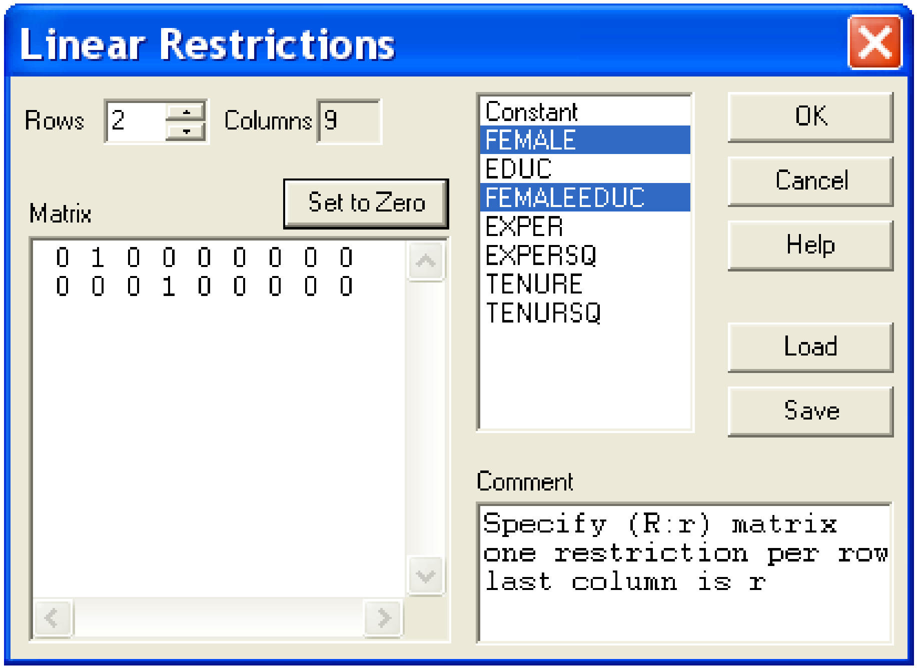
R^2 0.440964 F(7,518) = 58.37 [0.000]

Exemple

- Le rendement de l'education des femmes est de $\hat{\beta}_2 + \hat{\beta}_3$, c-à-d $0.082 - 0.0056 = 0.0764$. Il est non significativement différent de celui des hommes $(beta_2)$: $tstat = -0.0056/0.0131 = -0.43$.
- Le salaire moyen des femmes est de $exp(0.227) - 1 = 0.255\%$ inférieur à celui des hommes. La différence est non significative : $tstat = -0.227/0.168 = -1.35$.
- Cela ne signifie pas nécessairement qu'il n'y pas discrimination : les variables $female$ et $female \times educ$ sont très corrélées (0.96357) $\rightarrow$ individuellement moins significatives.

F Test

F test $H_0 : \beta_1 = \beta_3 = 0 \rightarrow F = 34.326 \rightarrow$ à comparer par rapport à une $F(2, 518)$.
$\rightarrow$ p-valeur $< 0.0001 \rightarrow$ discrimination.




```
LinRes F(2,518) =   34.326 [0.0000]**
```

Test de difference entre groupes

- Comment tester si le même modèle de regression (intercept et pentes) s'applique à 2 groupes différents $(g = 1, 2)$?
- $y = \beta_{g,0} + \beta_{g,1}x_1 + \beta_{g,2}x_2 + ... + \beta_{g,k}x_k + u_g.$
- Hommes: $wage = \beta_{h,0} + \beta_{h,1}educ + \beta_{h,2}exper + u_h.$
- Femmes: $wage = \beta_{f,0} + \beta_{f,1}educ + \beta_{f,2}exper + u_f.$
- Comment tester $H_0 : \beta_{h,0} = \beta_{f,0}, \beta_{h,1} = \beta_{f,1}, \beta_{h,2} = \beta_{f,2}$?

Test de Chow

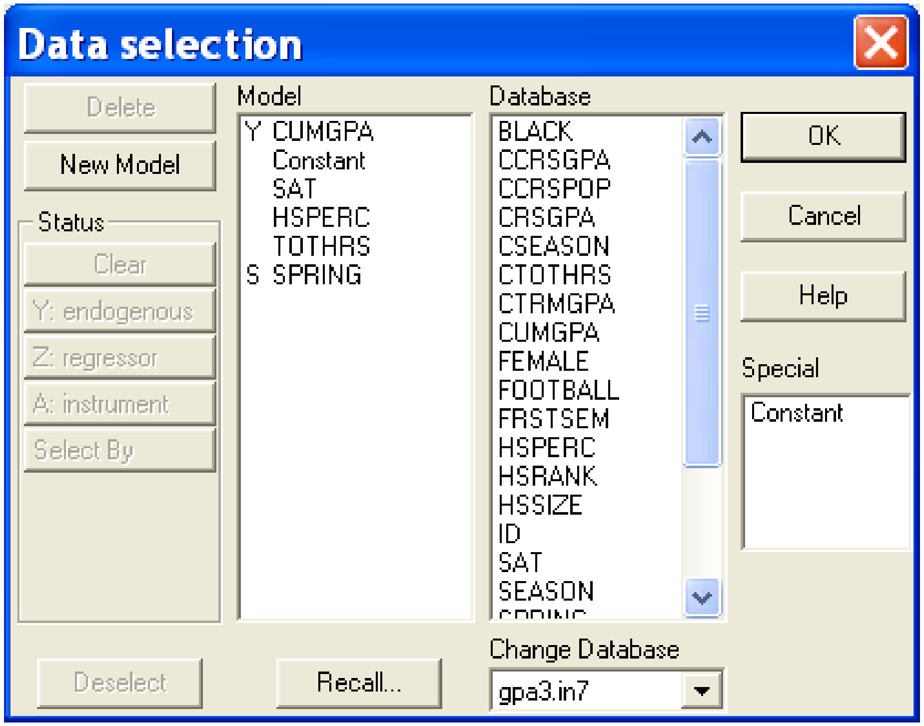
- Estimer le modèle contraint (sous H_0): $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$.
- $\rightarrow wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$.
- Calculer $SSR_p =$ somme des carrés des résidus du modèle contraint (n obs. au total).
- Estimer le modèle pour $g = 1$ (Hommes) et $g = 2$ (Femmes).
- Calculer $SSR_1 (n_1$ obs.$)$ et $SSR_2 (n_2$ obs.$)$.
- Sous $H_0 : F = \frac{SSR_p - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \times \frac{n - 2(k+1)}{k+1} \sim \Im(k+1, n-2(k+1))$.
- $\rightarrow$ Le modèle contraint implique $k+1$ restrictions.

Résultats scolaires au printemps

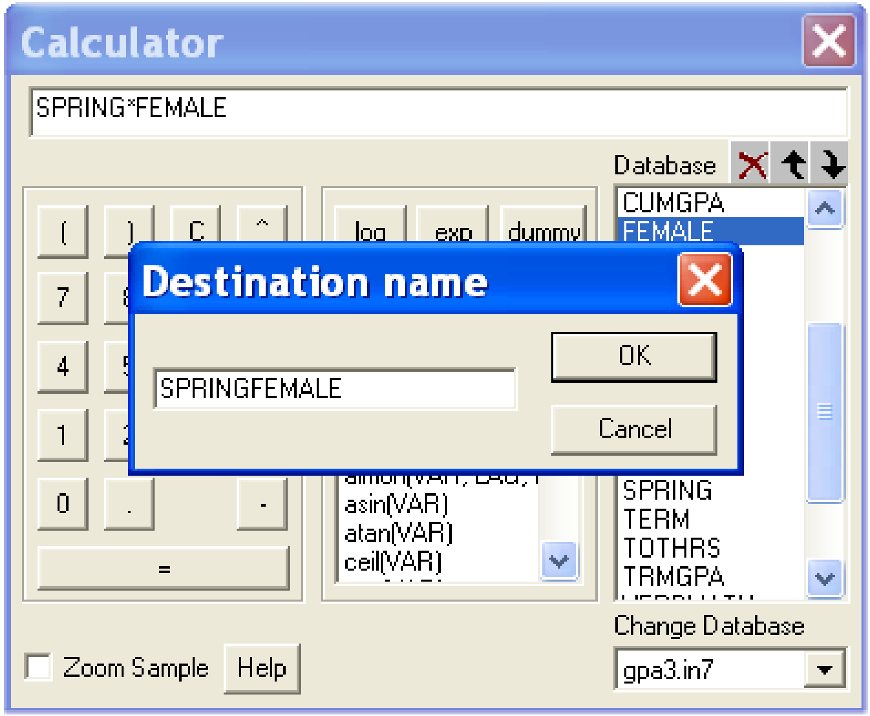
$CUMGPA = \beta_0 + \beta_1 SAT + \beta_2 HSPERC + \beta_3 TOTHRS + u$,
où $CUMGPA =$ College Grade Point Average cumulé.
$\rightarrow$ 90 filles et 276 garçons; 3 variables explicatives;

```
EQ(1) Modelling CUMGPA by OLS-CS
       Estimation sample selected
       by SPRING: 366 observations

       Coefficient t-prob
Constant   1.49085 0.000
SAT     0.00118496 0.000 sc. à l'ex. d'entrée
HSPERC -0.00995690 0.000 High school perc.
TOTHRS  0.00234295 0.002 # tot. d'h. de cours

sigma           0.486034  RSS              85.5150683
R^2             0.351636  F(3,362) = 65.44 [0.000]
no. of obs.          366  no. of parameters         4
```

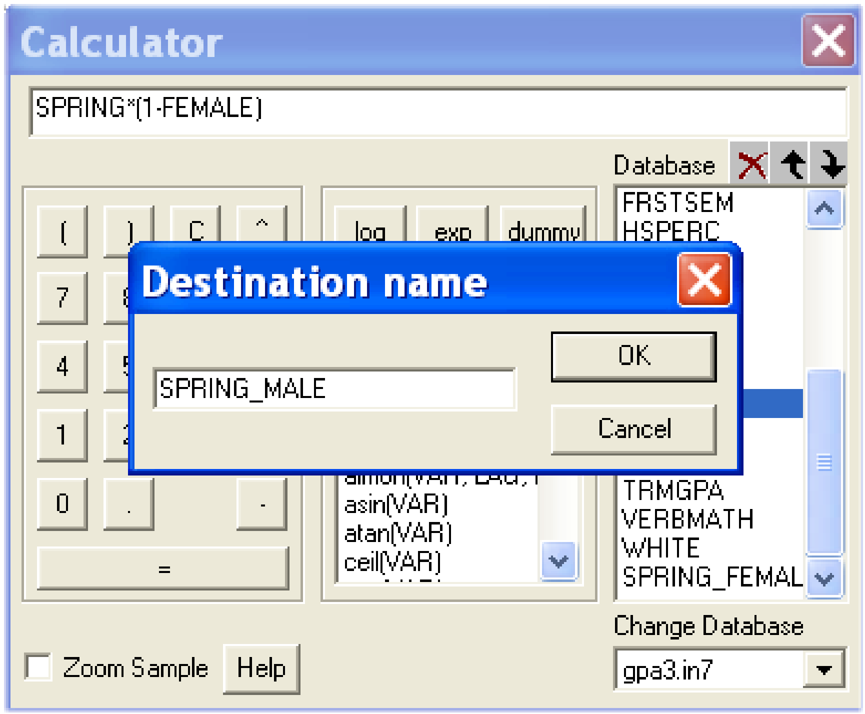
Sélection de $SPRING = 1$



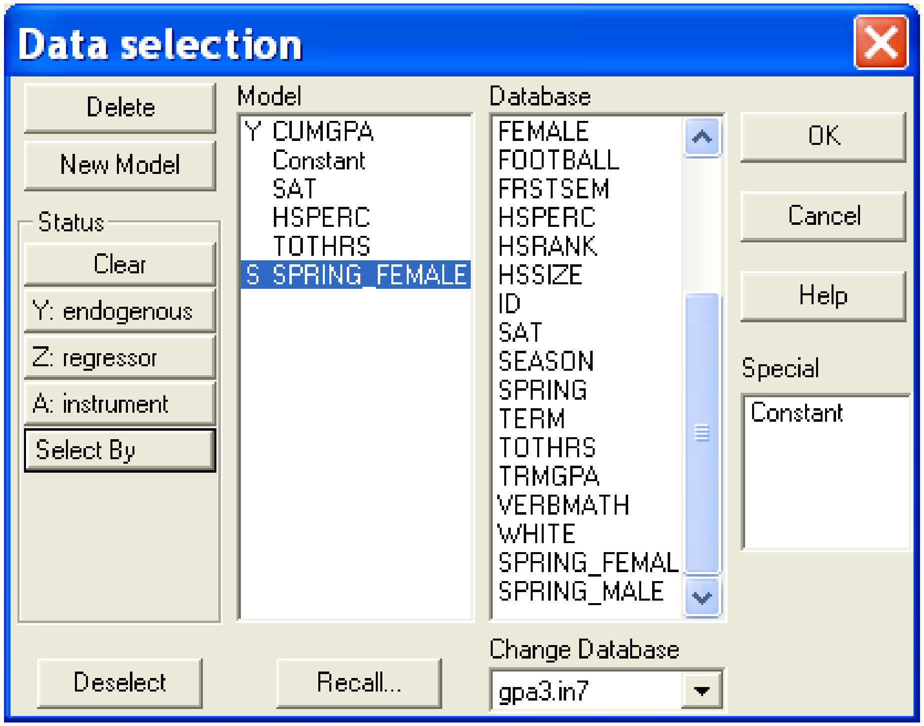
$SPRING = 1$ **et** $FEMALE = 1$



$SPRING = 1$ **et** $MALE = 1$

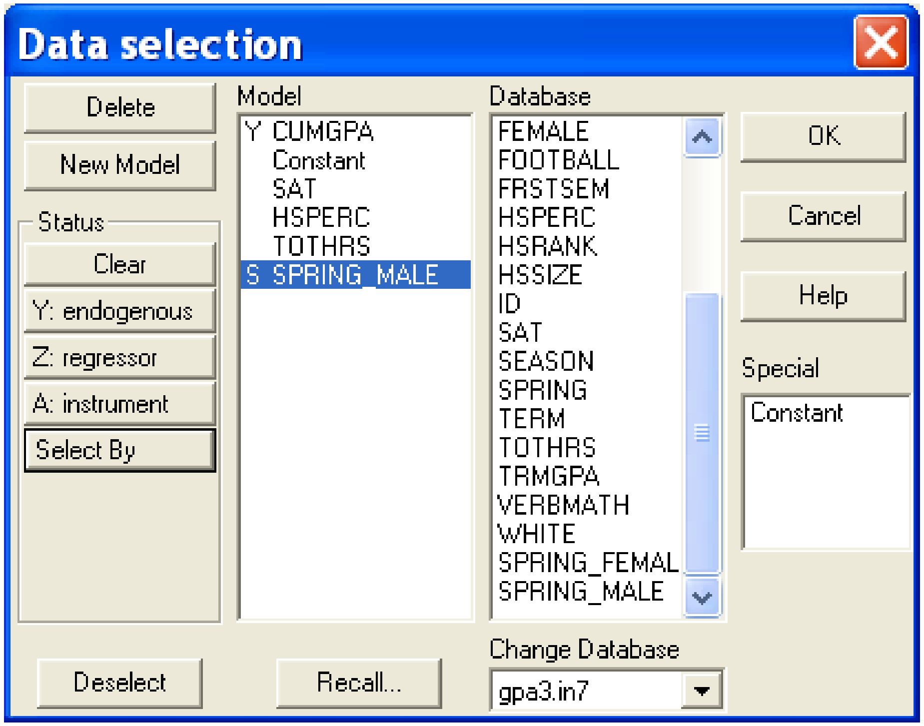


Modèle $SPRINGFEMALE = 1$



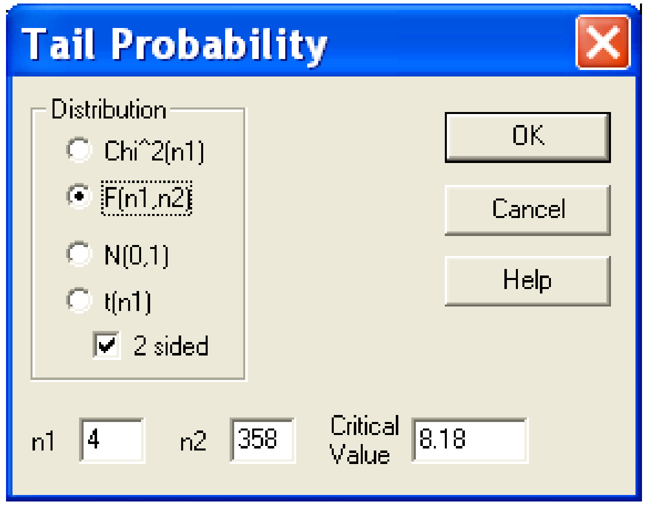
$\rightarrow RSS = 19.6027867, n_1 = 90.$

Modèle $SPRINGMALE = 1$



$\rightarrow RSS = 58.7517208, n_2 = 276.$

F Test



$$
\begin{aligned}
F &= \frac{SSR_p - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \times \frac{n - 2(k+1)}{k+1} \\
&= \frac{85.5150683 - (19.6027867 + 58.7517208)}{19.6027867 + 58.7517208} \times \frac{366 - 2(3+1)}{4} \\
&= \frac{85.5150683 - 78.355}{78.355} \times \frac{358}{4} \\
&\simeq 8.18 \gg 3.32 \text{ (valeur critique pour } \alpha = 1\%\text{).}
\end{aligned}
$$

$\rightarrow$ Rejet de H_0.

Variable binaire dépendante

- Prenons le cas d'une variable binaire dépendante. Exemple : individu hautement éduqué ou non; individu utilisant des drogues ou non.
- Modèle linéaire de régression multiple: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u.$
- Comme $y = 0$ ou $y = 1$, interprétation de β_j particulière: β_j ne capture pas l'effet d'une variation de 1 unité de x_j sur une variation de $y \rightarrow$ interprétation en terme de probabilité.
- En prenant l'espérance conditionnelle $E(y|x)$: $Prob(y = 1|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$ car dans ce cas $E(y|x) = Prob(y = 1|x) \rightarrow$ capture la probabilité conditionnelle de succès.
- $\rightarrow$ Modèle de probabilité linéaire (MPL).

- $\Delta Prob(y = 1|x) = \beta_j \Delta x_j$.
  $\rightarrow \beta_j$ mesure la modification de la probabilité de succès ($y = 1$) quand x_j augmente d'1 unité.
- $\hat{\beta}_0$ mesure la probabilité de succès quand tous les x sont égaux à zéro.
- Exemple : Participation au marché du travail des femmes mariées en 1975.
  $\hat{inlf} = \hat{\beta}_0 + \hat{\beta}_1 nwifeinc + \hat{\beta}_2 educ + \hat{\beta}_3 exper + \hat{\beta}_4 exper^2 + \hat{\beta}_5 age + \hat{\beta}_6 kidslt6 + \hat{\beta}_7 kidsgt6 + u.$

- $y = 1$ si femme mariée travaille.
  $\rightarrow$ 428 femmes sur 753.
- Interprétation de β_2 : impact d'une année d'étude supplémentaire sur la probabilité pour une femme mariée de travailler.
- Impact linéaire (irréaliste) : pour une série de valeurs données pour les autres variables, en dessous d'un certain nombre d'année d'études, la probabilité devient négative!!

Limites du MPL

- Prédictions de probabilité irréaliste. Pour certaines valeurs des variables, on peut obtenir $Prob(y = 1|x) < 0$ ou $Prob(y = 1|x) > 1$. $\rightarrow$ irréaliste.
- Exemple : pour 10, des 753 femmes, compte tenu des valeurs estimées par MCO et des valeurs des variables,$\hat{P}rob(y = 1|x) < 0$.
- Les valeurs irréalistes sont aussi liées à l'hypothèse de linéarité. Exemple : $\hat{\beta}_6 = -0.262$. Si une femme passe de 0 à 1 jeune enfant, sa probabilité de travailler baisse de 0.262.
  $\rightarrow$ Si une femme passe de 0 à 4 jeunes enfants, sa probabilité de travailler baisse de $4 \times 0.262 = 1.048$.
  $\rightarrow$ irréaliste.

Limites du MPL

- Le modèle MPL viole l'hypothèse d'homoscédasticité $\rightarrow$ inférence problématique.
- En effet, la variance conditionnelle de y dépend de x:
  $Var(y|x) = \frac{Prob(y=1|x)}{1-Prob(y=1|x)}$.
  $Var(y|x) = \frac{(\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_k x_k)}{1-(\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_k x_k)}$.
- En règle générale, si on omet les problèmes de prévision irréalistes et les problèmes d'hétéroscédasticité, le MPL est utilisé souvent dans l'analyse empirique préliminaire.

Problèmes d'auto-sélection

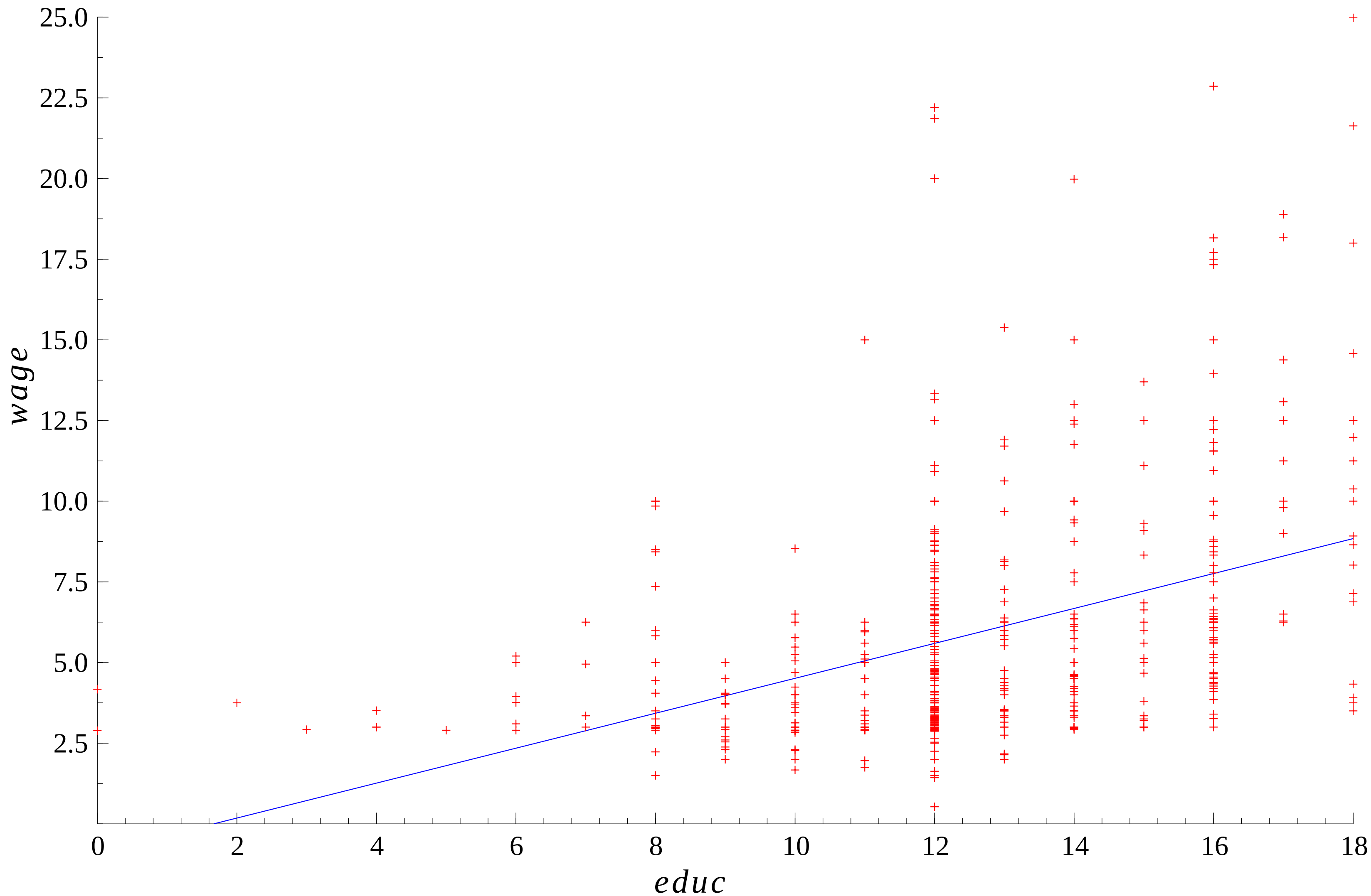
- $y = \beta_0 + \beta_1 partic + u$.
- Les variables binaires capturant la participation ($partic$) à des politiques doivent être exogènes par rapport au terme d'erreur u ou à y.
- Si $E(u|partic = 1) \neq E(u|partic = 0)$, il y a auto-selection $\rightarrow$ la participation dépend des facteurs non inclus (et présents dans u) $\rightarrow$ estimation biaisée de β_1.
- Exemple : impact des bourses de formation aux entreprises $(grant = 0, 1)$ sur le taux de déchet dans la production. Est que $grant$ est exogène ou déterminé aléatoirement. $\rightarrow$ il est possible que les firmes les moins productives demandent et reçoivent plus de bourses.$\rightarrow$ estimation biaisée de β_1.

- Une solution : inclure des facteurs observés qui son corrélés avec $partic$ et qui vont capturer une partie de la variabilité de u.
- Exemple : inclure ventes, $log(sale)$, et la taille de l'entreprise, $log(employ)$, pour diminuer ce problème d'auto-sélection.

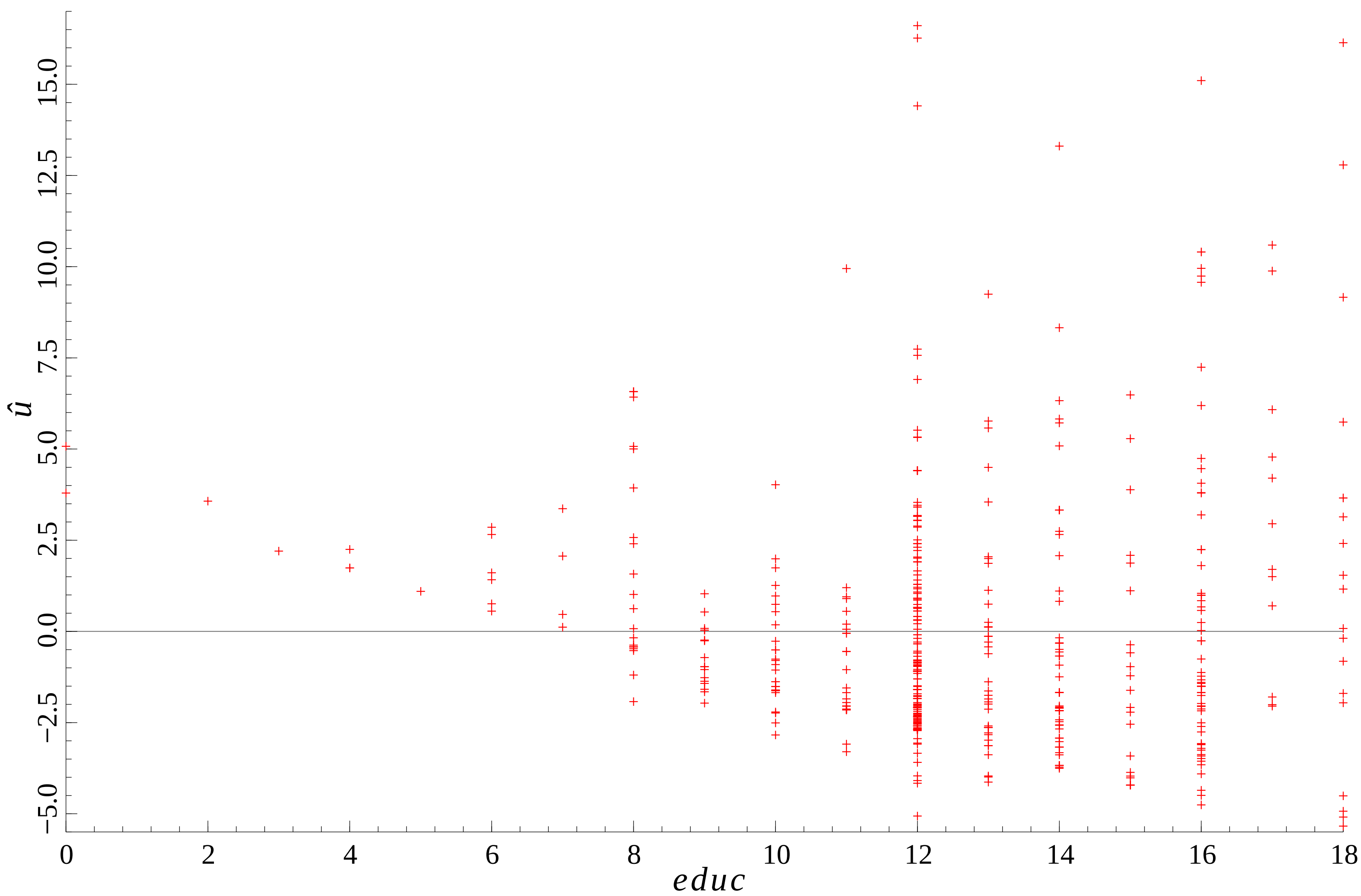
Chapitre 8: Hétéroscédasticité

- Dans le chapitre 3, nous avons introduit 5 hypothèses dont l'hypothèse d'homoscédasticité (**MLR.5**).
- **MLR.5** $Var(u|x_1, \ldots, x_k) = \sigma^2$.
  $\rightarrow$ homoscédasticité.
- **MLR.5** est violée si $Var(u|x_1, \ldots, x_k) = \sigma_i^2$.
  $\rightarrow$ hétéroscédasticité.
- Quelles sont les conséquences ?

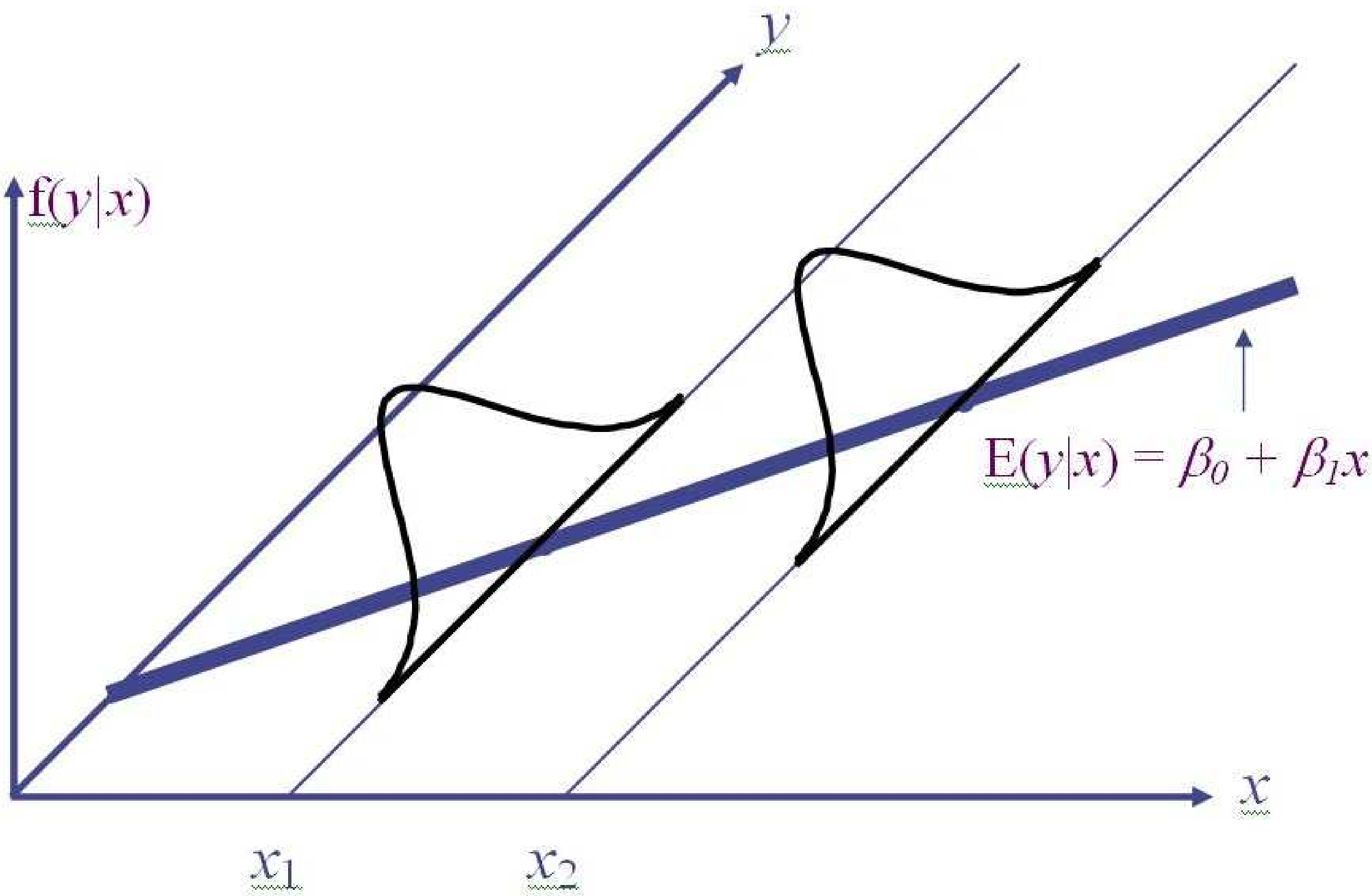
Exemple: $wage = \beta_0 + \beta_1 educ + u$



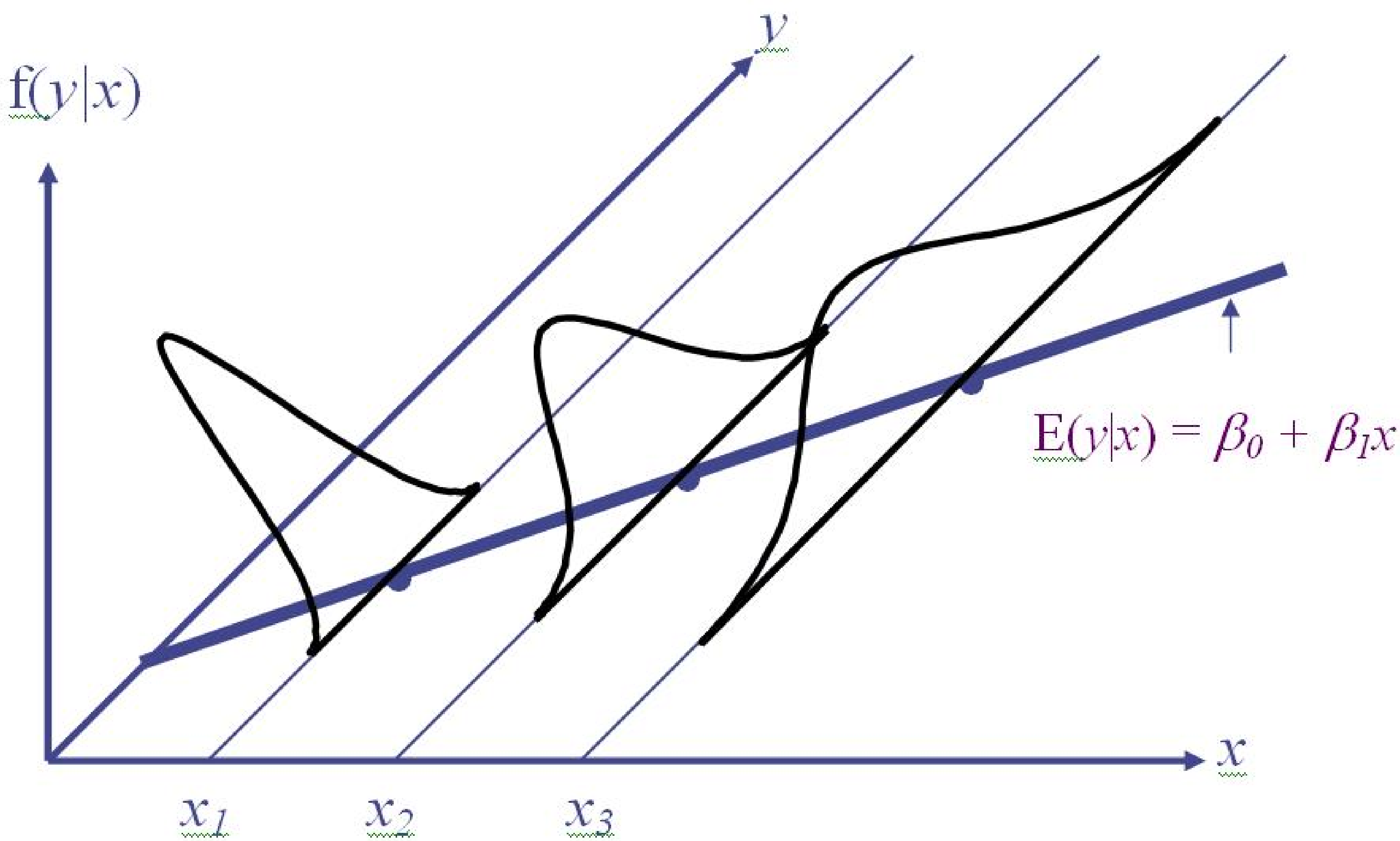
Lien entre $\hat{u}$ et $educ$



Rappel : Homoscédasticité



Rappel : Hétéroscédasticité



Conséquences de l’hétéroscédasticité pour les MCO

Violation sans conséquences

- Sous hypothèses **MLR.1-MLR.4**.
  $\rightarrow$ les MCO sont non biaisés.
- $\rightarrow$ L'hypothèse $Var(u|x_1, \ldots, x_k) = \sigma^2$ ne joue aucun rôle dans le caractère non biaisé.
- Le calcul et l'interprétation du R^2 et $\overline{R}^2$ ne dépendent pas de la validité de **MLR.5**.

Violation avec Conséquences

- Par contre, **MLR.5** est important pour l'estimation de la variance des estimateurs: $Var(\hat{\beta}_j)$.
- En présence d'hétéroscédasticité, la formule proposée de $\hat{Var}(\hat{\beta}_j)$ donne un estimateur biaisé.
- Conséquences ?
- $\rightarrow$ Les statistiques de test en t n'ont plus une distribution en t.
- $\rightarrow$ Les statistiques de test en F ne sont plus valides.
- $\rightarrow$ Les statistiques de test LM ne suivent plus une distribution du $\chi^2(.)$.

Inférence robuste à l'hétéroscédasticité

Forme inconnue d'hétéroscédasticité

- En présence d'hétéroscédasticité, les MCO ne sont plus efficients.
- Stratégie si la forme de l'hétéroscédasticité est inconnue (cas le plus fréquent) ?
- Garder les estimateurs $\hat{\beta}_j$ des MCO mais ajuster les statistiques nécessaires à l'inférence.
- $\rightarrow$ Ajuster les statistiques en t, F et LM.
- $\rightarrow$ Procédures robustes à l'hétéroscédasticité valides dans des grands échantillons.

Calcul de $Var(\hat{\beta}_j)$

- $y_i = \beta_0 + \beta_1 x_i + u_i$.
- Hétéroscédasticité : $Var(u_i|x_i) = \sigma_i^2$.
- $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \overline{x}) u_i}{\sum_{i=1}^n (x_i - \overline{x})^2} = \beta_1 + \frac{\sum_{i=1}^n (x_i - \overline{x}) u_i}{SST_x}$.
- $\rightarrow Var(\hat{\beta}_1) = \frac{\sum_{i=1}^n (x_i - \overline{x})^2 \sigma_i^2}{SST_x^2}$.
- Dans le cas homoscédastique ($\sigma_i^2 = \sigma^2$ pour tous les i).
- $\rightarrow$ On retrouve $Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$ et $\hat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{SST_x}$.
- Par contre si hétéroscédasticité $\rightarrow$ la formule de $Var(\hat{\beta}_1)$ dérivée dans le cas homoscédastique n'est plus valable.

Écart-types robustes

- White (1980) propose d'utiliser la formule théorique pour trouver un estimateur de $V\hat{a}r(\hat{\beta}_j)$ en cas d'hétéroscédasticité.
- Idée : estimer $\hat{\sigma}_i^2$ par les $\hat{u}_i^2$.
- $V\hat{a}r(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 \hat{u}_i^2}{SST_x^2}$.
- Dans le cas de régression multiple, la formule analogue est : $V\hat{a}r(\hat{\beta}_j) = \frac{\sum_{j=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$, où pour rappel $\hat{r}_{ij}$ est le $i$ème résidu de la régression de x_j sur les autres variables explicatives de la régression multiple.

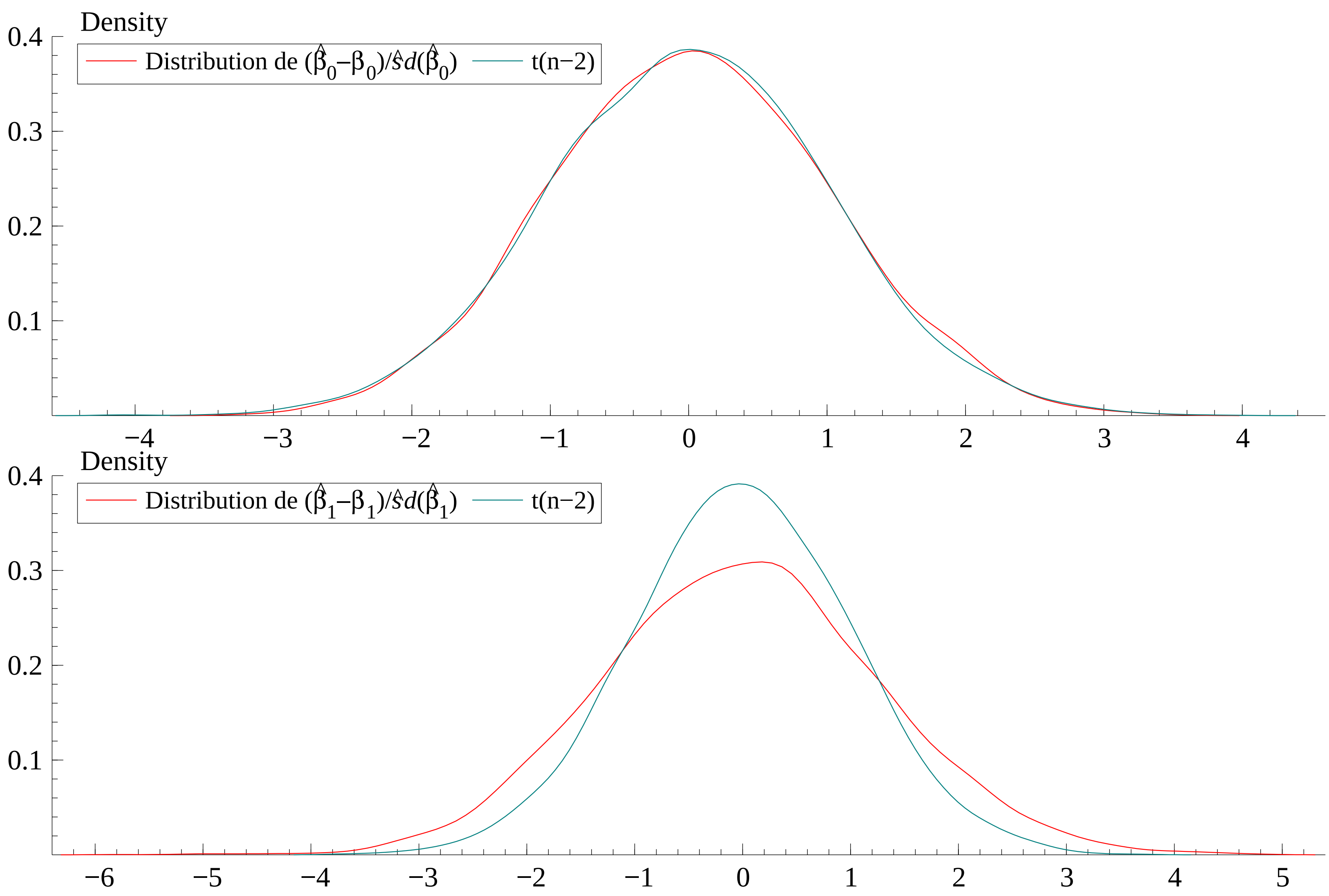
Écart-types robustes

- La racine carrée de cet estimateur donne un estimateur de l'écart-type de $\hat{\beta}_j$ robuste à la présence d'hétéroscédastique ou écart-type de White : noté HCSE.
- MacKinnon et White (1985) proposent d'ajuster la formule de $\hat{Var}(\hat{\beta}_j)$ par un facteur correctif $\frac{n}{n-k-1}$: noté JHCSE pour Jack-knife "heteroscedastic consistant SE".
- Remarque : si on a homoscédasticité, on retrouve les écart-types des MCO habituels.
- On peut alors construire des statistiques t robustes : $t_{robuste} = \frac{\hat{\beta}_j - \hat{\beta}_j^0}{\hat{sd}(\hat{\beta}_j)}$.

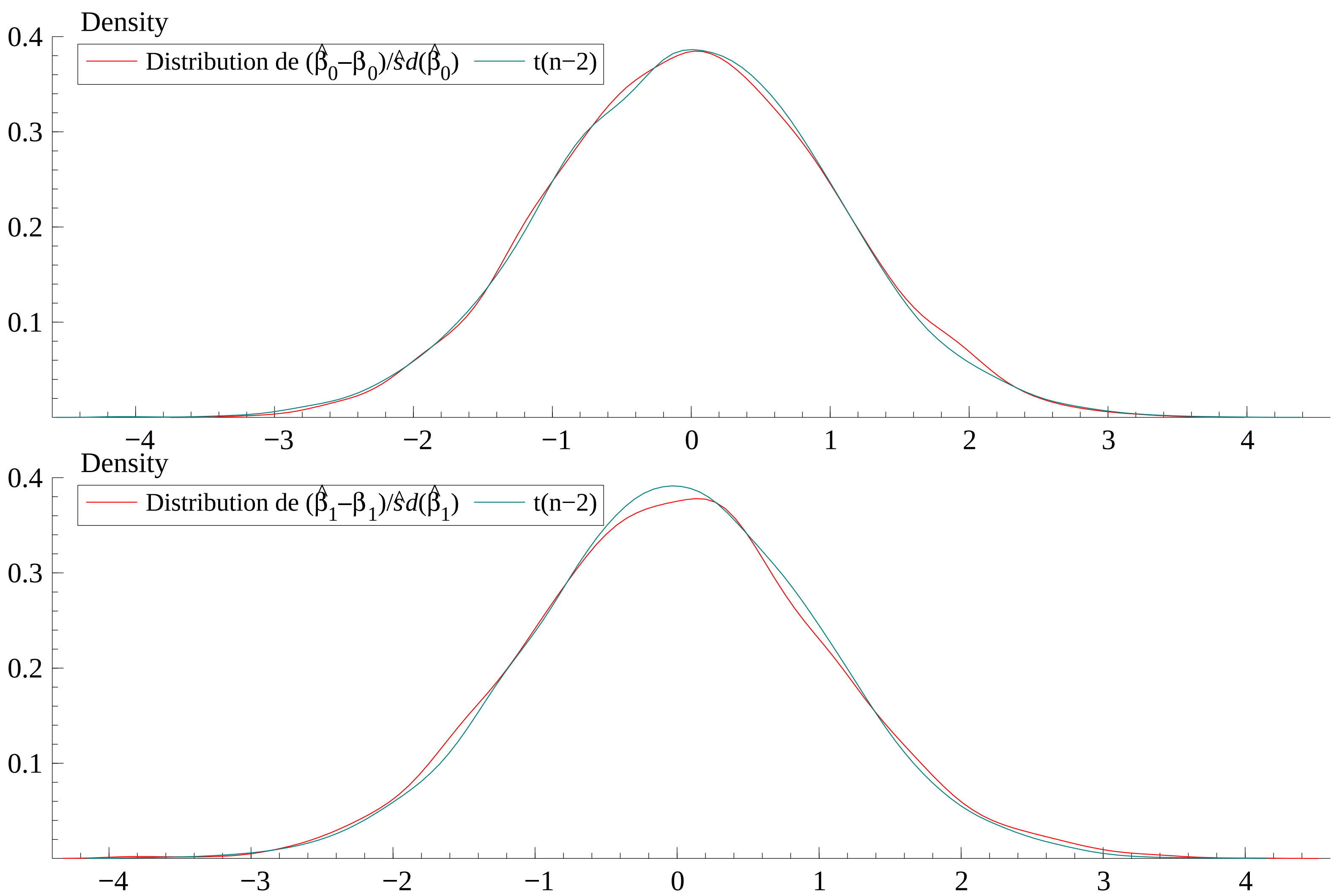
Illustration par une simulation.

$\rightarrow n = 200, y = 0.2 + 0.5x_1 + u$, et $u \sim N(0, 0.5exp(0.2x_1^2))$.
Estimer $y = \beta_0 + \beta_1 x_1 + u$ 5000 fois et calculer après chaque estimation: $stat_{h,j} = \frac{\hat{\beta}_j - \beta_j}{\hat{sd}(\hat{\beta}_j)}$ $\forall h = 1, \ldots, 5000$ et $j = 0, 1$, avec $\hat{sd}(\hat{\beta}_j)$ calculé selon la formule standard, HCSE et JHCSE.

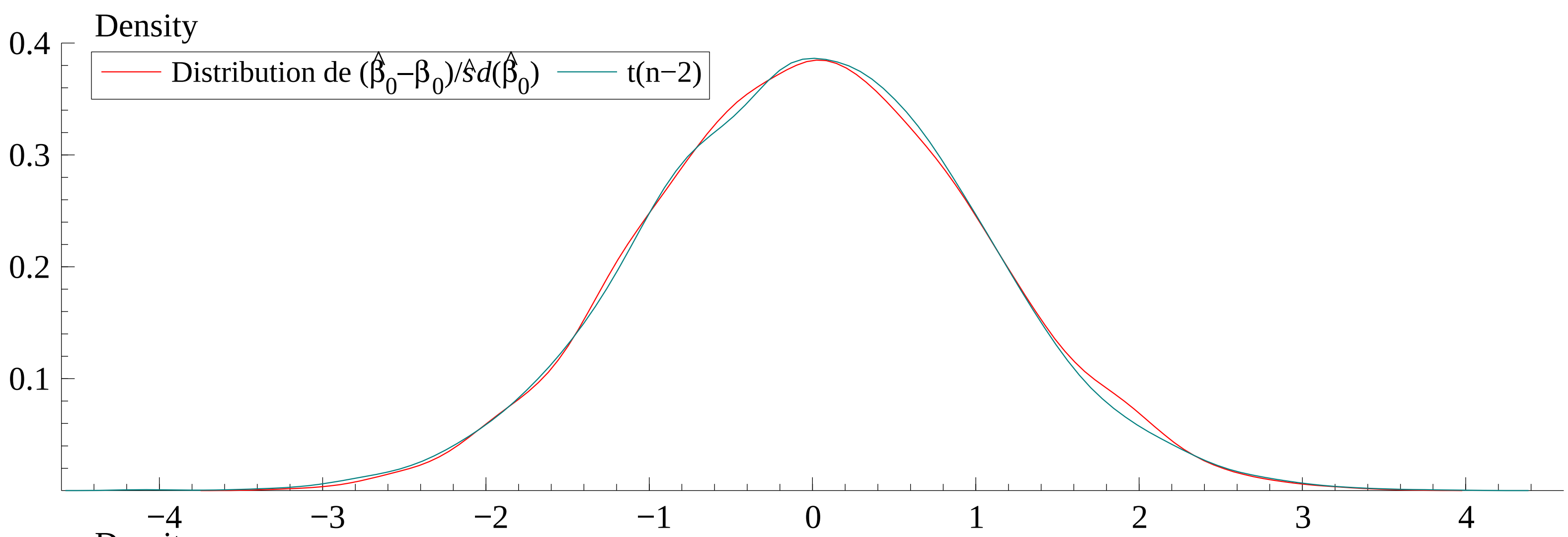
Écart-types standards

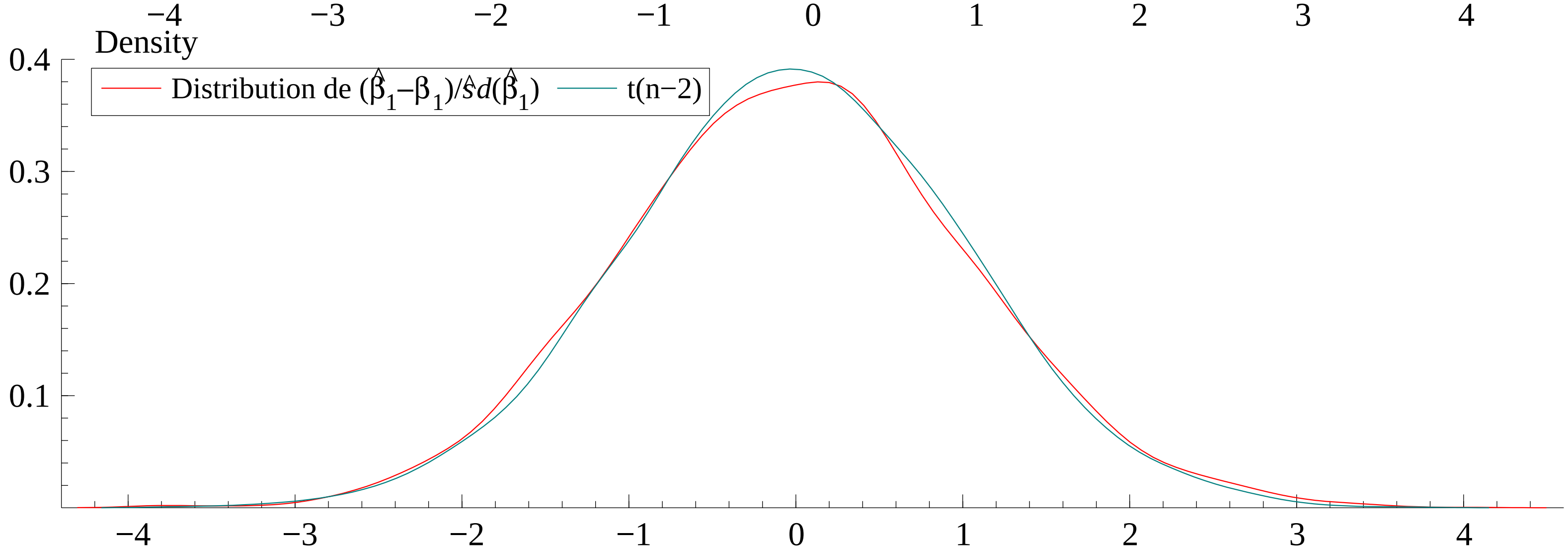


Écart-types HCSE

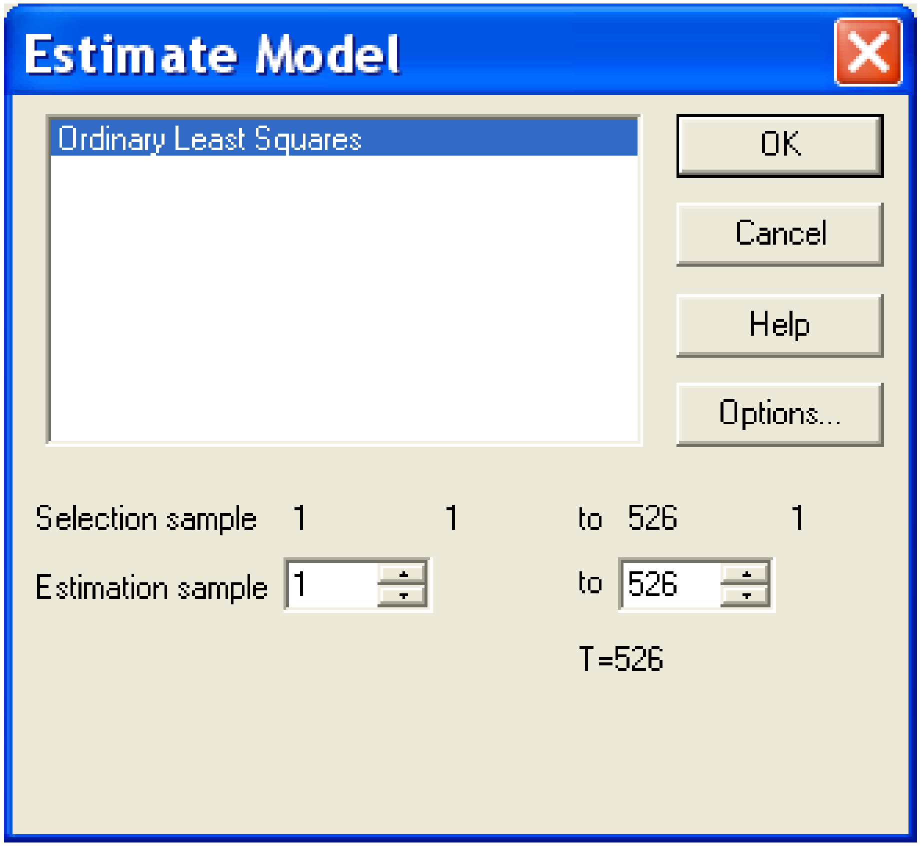


Écart-types JHCSE

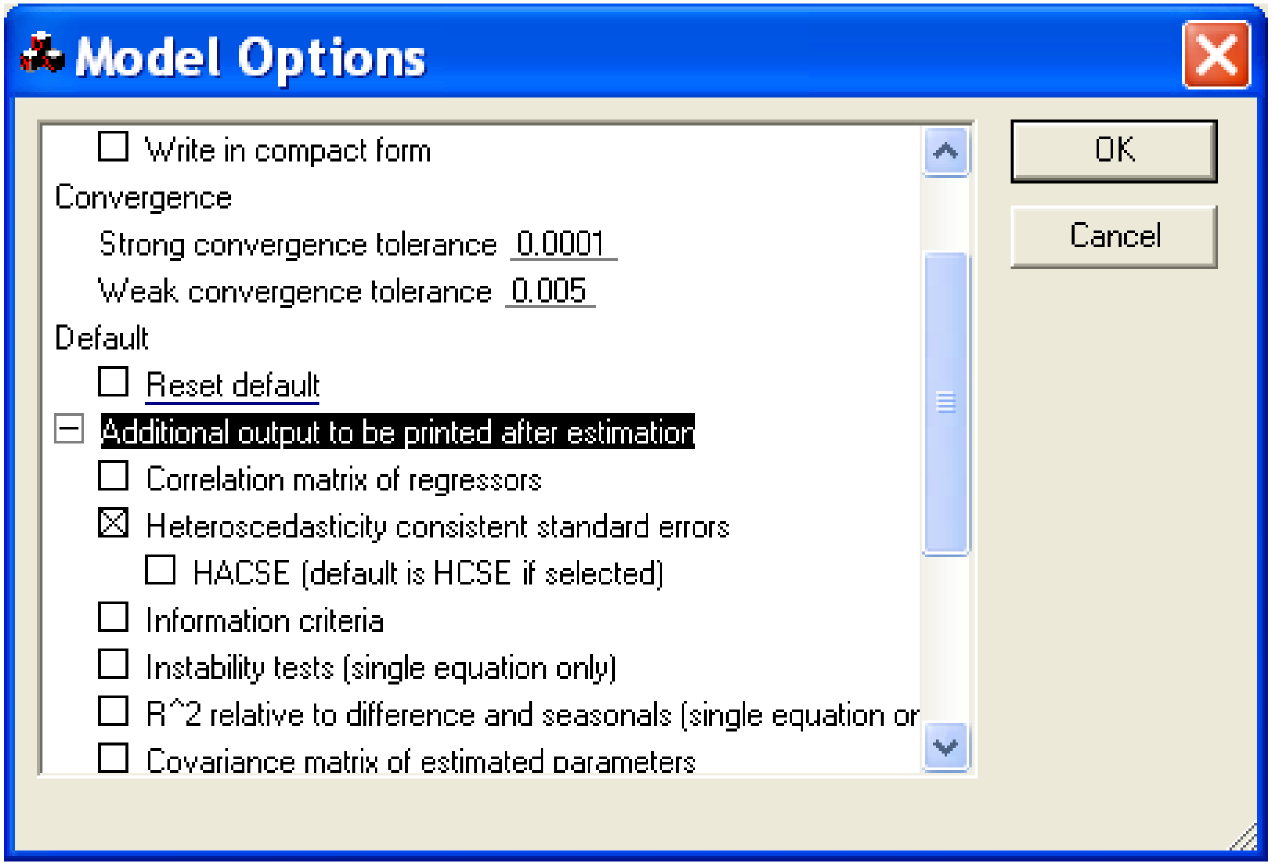




GiveWin


Estimate Model
Ordinary Least Squares
OK
Cancel
Help
Options...
Selection sample 1 1 to 526 1
Estimation sample 1 to 526
T=526

GiveWin



Model Options
Write in compact form
Convergence
Strong convergence tolerance 0.0001
Weak convergence tolerance 0.005
Default
Reset default
Additional output to be printed after estimation
Correlation matrix of regressors
Heteroscedasticity consistent standard errors
HACSE (default is HCSE if selected)
Information criteria
Instability tests (single equation only)
R^2 relative to difference and seasonals (single equation or
Covariance matrix of estimated parameters
OK
Cancel

Exemple: Salaire

	Coefficient	Std.Error	JHCSE
Constant	0.321378	0.1000	0.1114
MARRMALE	0.212676	0.05536	0.05811
MARRFEMALE	-0.198268	0.05784	0.05964
SINGFEMALE	-0.110350	0.05574	0.05780
EDUC	0.0789103	0.006694	0.007565
EXPER	0.0268006	0.005243	0.005200
EXPERSQ	-0.000535245	0.0001104	0.0001081
TENURE	0.0290875	0.006762	0.007372
TENURSQ	-0.000533142	0.0002312	0.0002719

Exemple: équation de salaire

- On voit ici que les t statistiques robustes sont très proches des t de départ.
$\rightarrow$ peu de présence d'hétéroscédasticité.
$\rightarrow$ rien ne change du point de vue de la significativité.
- Les écart-types robustes peuvent être soit plus élevés, soit plus faibles $\rightarrow$ le biais va dans les 2 sens.
- En pratique néanmoins, les écart-types robustes sont souvent plus grands que les écart-types non ajustés.

F et LM robustes

- On peut construire des statistiques en F robustes.
$\rightarrow$ utiles pour tester des hypothèses jointes.
- Certains mais pas tous les logiciels les construisent.
- Quand ce n'est pas disponibles, on peut construire des statistiques de test LM robustes à l'hétéroscédasticité.
- Exemple. Tester l'hypothèse $\beta_4 = \beta_5 = 0$ dans l'équation suivante : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$.
- Dans le cas homoscédastique, $LM = n \times R^2_{\widetilde{u}}$ où le $R^2_{\widetilde{u}}$ est le R^2 de la régression des résidus du modèle restreint $(\beta_4 = \beta_5 = 0)$ sur toutes les variables explicatives.

F et LM robustes

- La procédure pour construire des statistiques LM robustes en 3 étapes.
- Récupérer les résidus du modèle restreint $\widetilde{u}$.
- Régresser les q variables exclues $(x_4$ et $x_5)$ sur toutes les variables explicatives incluses $(x_1$, x_2 et $x_3)$ et récupérer les résidus $\rightarrow$ calculer les q produits : $\widetilde{r}_4$ et $\widetilde{r}_5$.
- Régresser des 1 sur $\widetilde{r}_4\widetilde{u}$ et $\widetilde{r}_5\widetilde{u} \rightarrow$ calculer la SSR.
- $LM = n - SSR$: sous H_0, LM suit une χ^2_q.

Exemple: arrestations criminelles

Var dépendante	$narr86$		
Var explicatives	coeff	ecart-type	ecart-type rob.
$Constant$	.567	.036	.040
$pcnv$	-.136	.040	.034
$avgsen$	.0178	.0097	.0101
$avgsen^2$	-.00052	.00030	.00021
$ptime86$	-.0394	.0087	.0062
$qemp86$	-.0505	.0144	.0142
$inc86$	-.00148	.00034	.00023
$black$	.325	.045	.058
$hispan$	.193	.040	.040
$R^2 = .0728$	$n = 2725$		

Exemple: arrestations criminelles

- Il y a des différences substantielles dans les écart-types. Exemple: $avgsen^2$: $t = -1.73$ vs $t = -2.48$.
- On veut tester si la sentence moyenne a une influence sur le nombre d'arrestations : $H_0 : \beta_{avgsen} = \beta_{avgsen^2} = 0$ $\rightarrow$ Test LM.
- Sans ajustement : $LM = 3.54 \rightarrow p - valeur = .170$.
- Avec ajustement : $LM = 4.00 \rightarrow p - valeur = .135$.
  $\rightarrow$ la sentence n'apparaît pas avoir un effet significatif.

Tests d'hétéroscédasticité

Test d'hétéroscédasticité

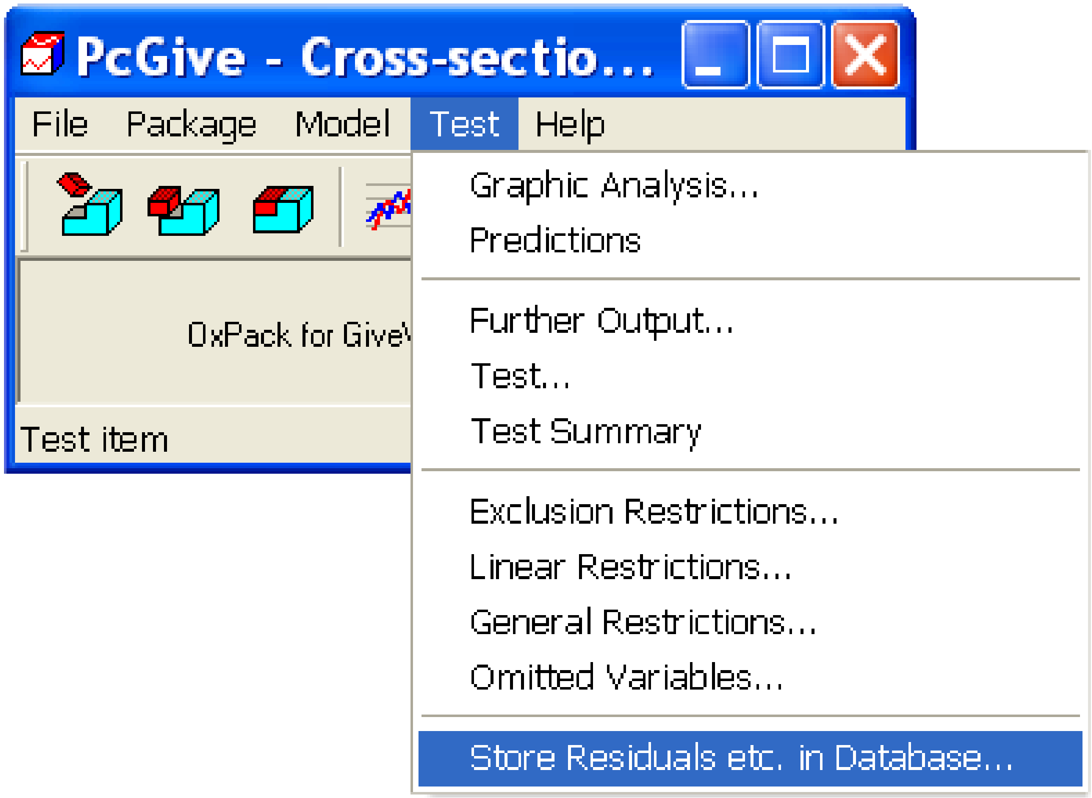
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u.$
- On suppose que les hypothèses **MLR.1-MLR.4** sont valides.
- Sous l'hypothèse nulle, cas homoscédastique: $H_0 : Var(u|x_1, \ldots, x_k) = \sigma^2$.
- Comme u a une espérance conditionnelle nulle (**MLR.2**), on a $Var(u|x) = E(u^2|x) \rightarrow$ on peut exploiter cela pour écrire l'hypothèse sous une forme testable : $H_0 : E(u^2|x_1, \ldots, x_k) = E(u^2) = \sigma^2$.
- Si l'hypothèse est violée, alors u^2 est lié à une des variables explicatives $\rightarrow$ on peut tester cela dans un modèle de régression linéaire.
- $u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \nu$, avec $H_0 : \delta_1 = \delta_2 = \ldots = \delta_k = 0$.

- Pour tester H_0, on peut utiliser soit F, soit LM, car asymptotiquement, ils ont distribués selon une χ^2.
- Problème : u est inobservé $\rightarrow$ on va utiliser les résidus $\hat{u} \rightarrow$ on estime : $\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + error$.
- Test en F: $F = \frac{R^2_{\hat{u}^2}/k}{(1-R^2_{\hat{u}^2})/(n-k-1)}$ où $R^2_{\hat{u}^2}$ est le R^2 de la régression précédente $\rightarrow$ sous H_0, F est distribué selon une $\Im_{k,n-k-1}$.
- Test LM (Test de Breusch-Pagan): $LM = n \times R^2_{\hat{u}^2}$. $\rightarrow$ suit une $\chi^2(k)$.
- Remarque 1 : $\hat{u}^2 > 0 \rightarrow$ rejet de **MLR.6**. Ce test n'est donc applicable qu'en grand échantillon.
- Remarque 2 : $\hat{u}^2$ estimé avec erreur. Cette erreur n'a pas de conséquence si n est grand.

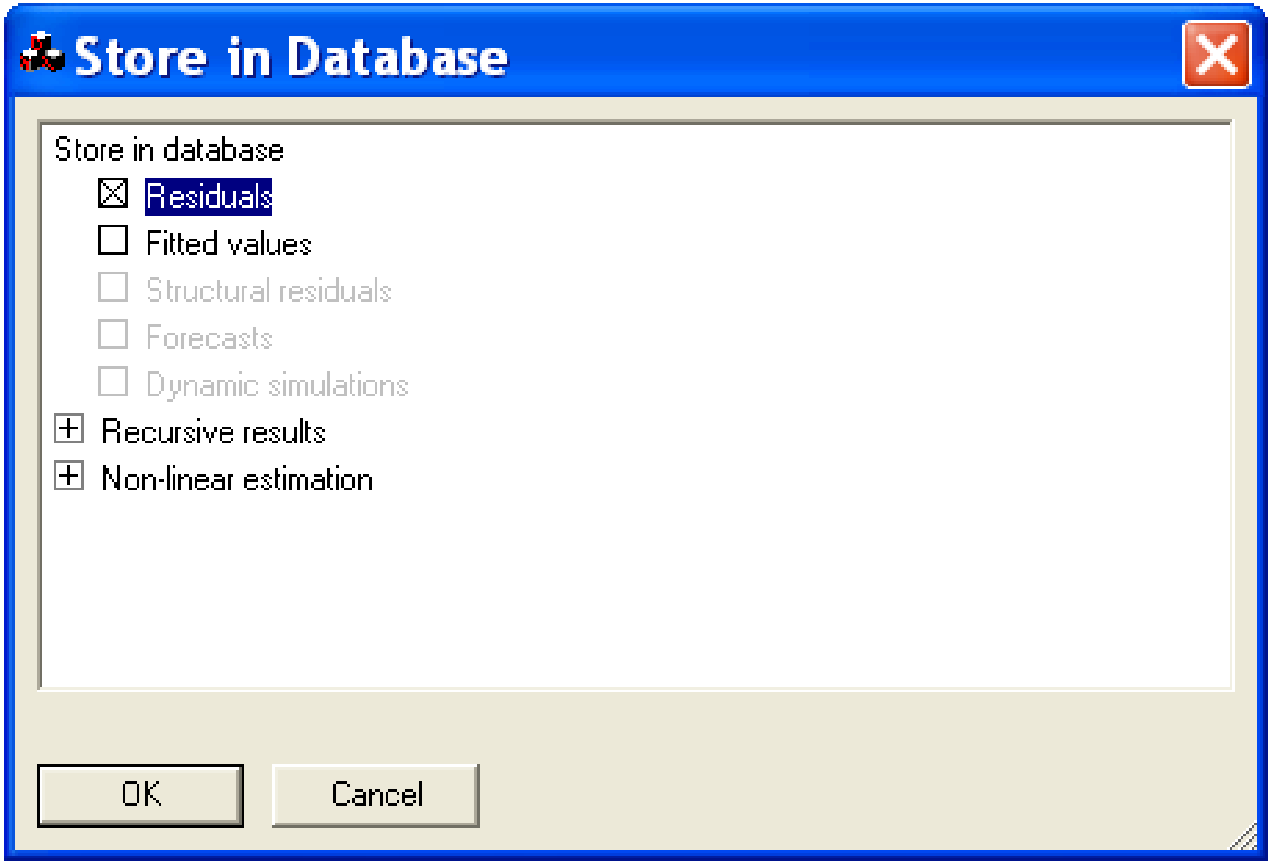
Exemple: Prix du logement

```
EQ(4) Modelling PRICE by OLS-CS (hprice1.in7)

       Coefficient Sd.Er t-prob
Constant -21.7703 29.480 0.462
LOTSIZE     0.0021 0.0006 0.002 (taille terr.)
SQRFT       0.1228 0.0132 0.000 (taille mais.)
BDRMS      13.8525 9.0100 0.128 (# salles bain)

sigma          59.8335 RSS              300723.806
R^2           0.672362 F(3,84) = 57.46 [0.000]
no. of obs.         88 no. of parameters      4
```

GiveWin: Sauvegarder $\hat{u}$

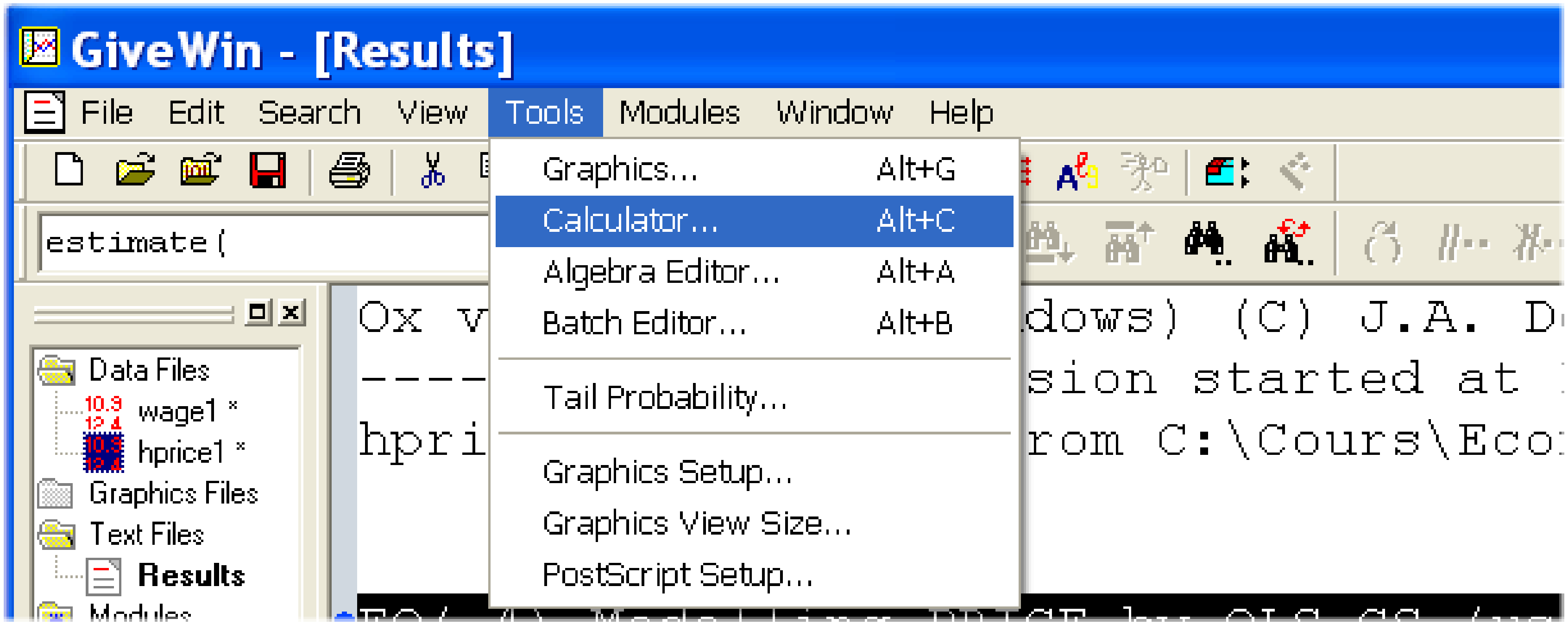


GiveWin: Sauvegarder $\hat{u}$

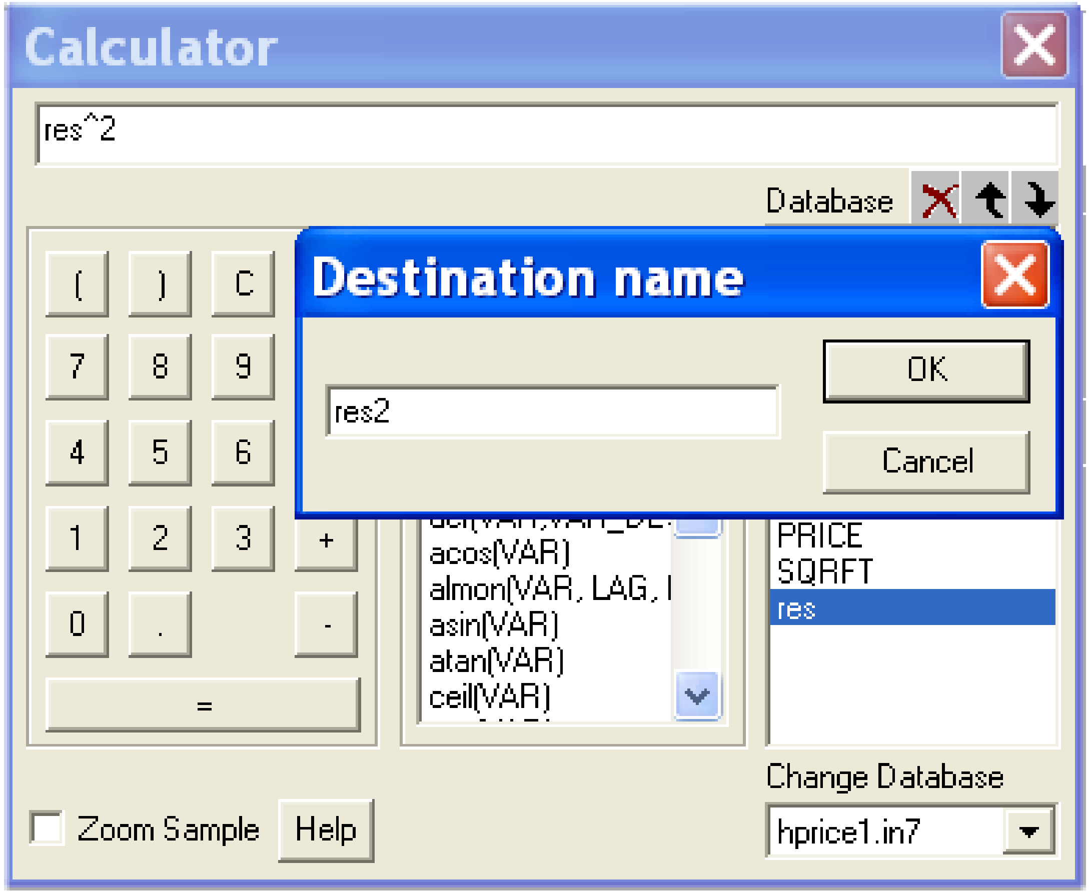


GiveWin: Sauvegarder $\hat{u}$

GiveWin: Créer $\hat{u}^2$



GiveWin: Créer $\hat{u}^2$



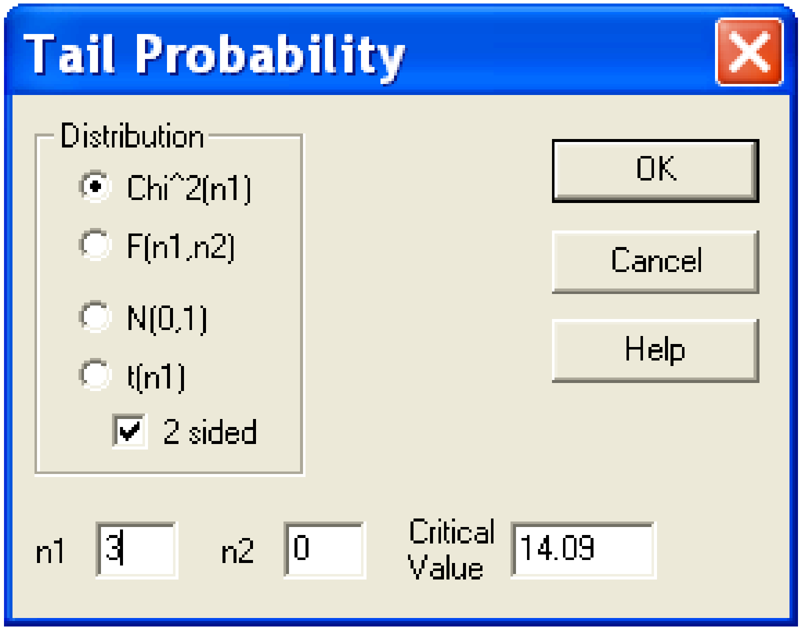
Test de Breusch-Pagan

```
EQ(5) Modelling res2 by OLS-CS

              Coefficient t-prob
Constant         -5522.79  0.094
LOTSIZE          0.201521  0.006
SQRFT             1.69104  0.251
BDRMS             1041.76  0.299

R^2 0.160141 F(3,84) = 5.339 [0.002]
no. of observations 88
```

$\rightarrow$ F stat de $5.339[0.002]$.

$\rightarrow$ LM stat de $88 \times 0.160141 \simeq 14.09$.

GiveWin: P-valeur du LM-test



→ P-valeur de 0.0028.

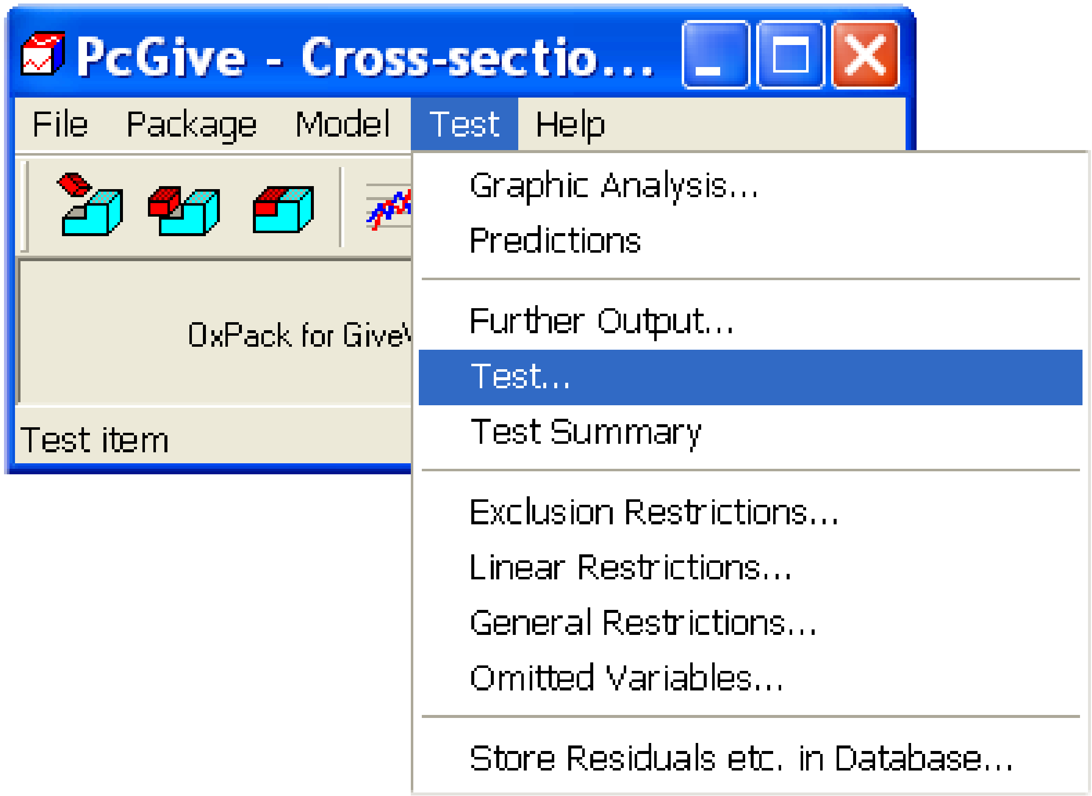
Exemple: prix du logement

- Modèle en log.
- $log(price) = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 bdrms + u$.
- $R^2_{\hat{u}^2} = .0480 \rightarrow F \approx 1.41 \rightarrow P - valeur = .245$. $LM \approx 4.22$ $\rightarrow p - valeur = .239$.$\rightarrow$
  Non rejet de l'hypothèse nulle d'homoscédasticité (vérifiez!!).
- Remarque: on peut ne pas inclure toutes les variables explicatives dans la régression des résidus au carré mais seulement celles dont on soupçonne avoir un lien avec la variance ($q < k$) $\rightarrow$ ajustement des DL dans F et χ^2.

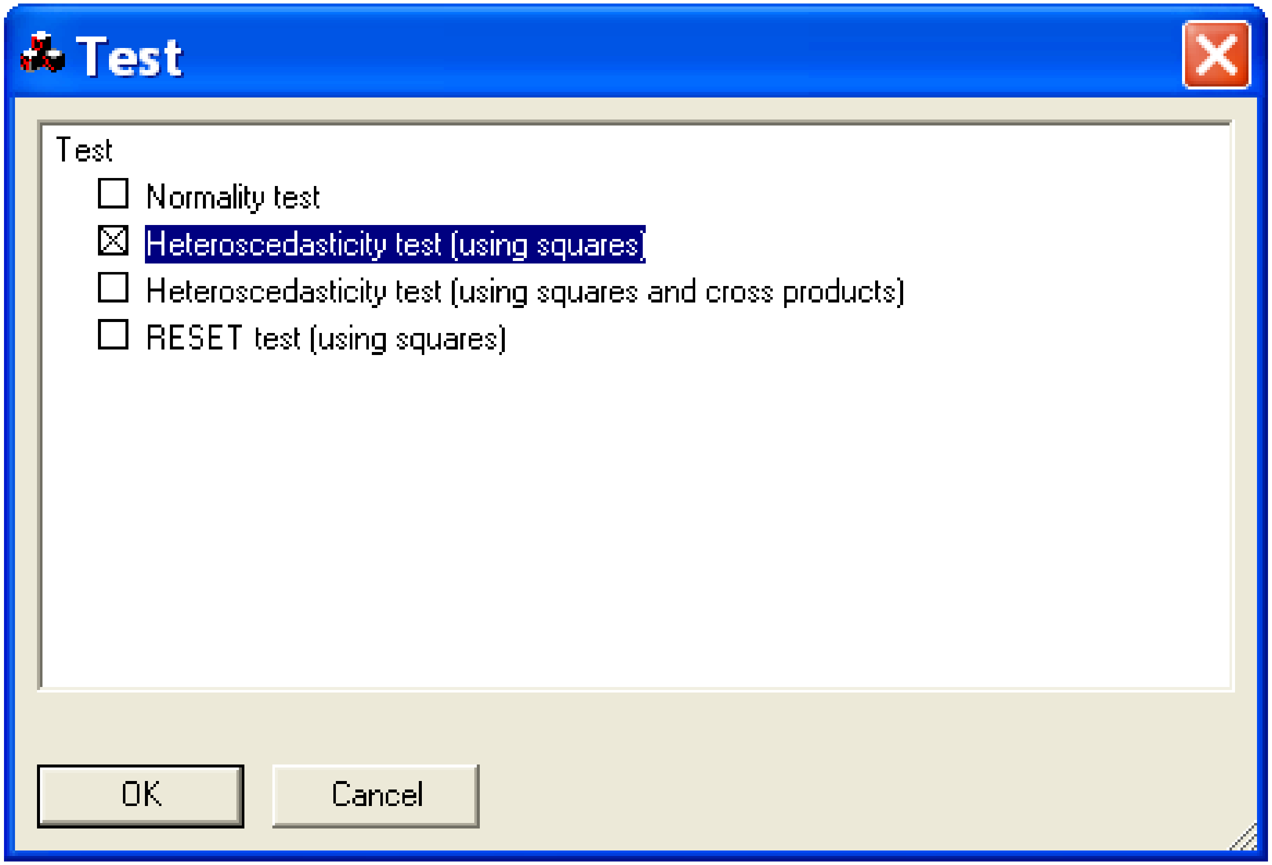
Test de White

- On étend le test vu précédemment en incluant, en plus des x_k, les carrés (x_k^2) et les produits croisés.
- Exemple: si $k = 3$, $\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 + \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + error \to$ Tests en F et en LM.
- Ceci permet de tester des formes alternatives d'hétéroscédasticité.
- Ce test est disponible dans la plupart des logiciels économétriques.

PcGive: Test de White



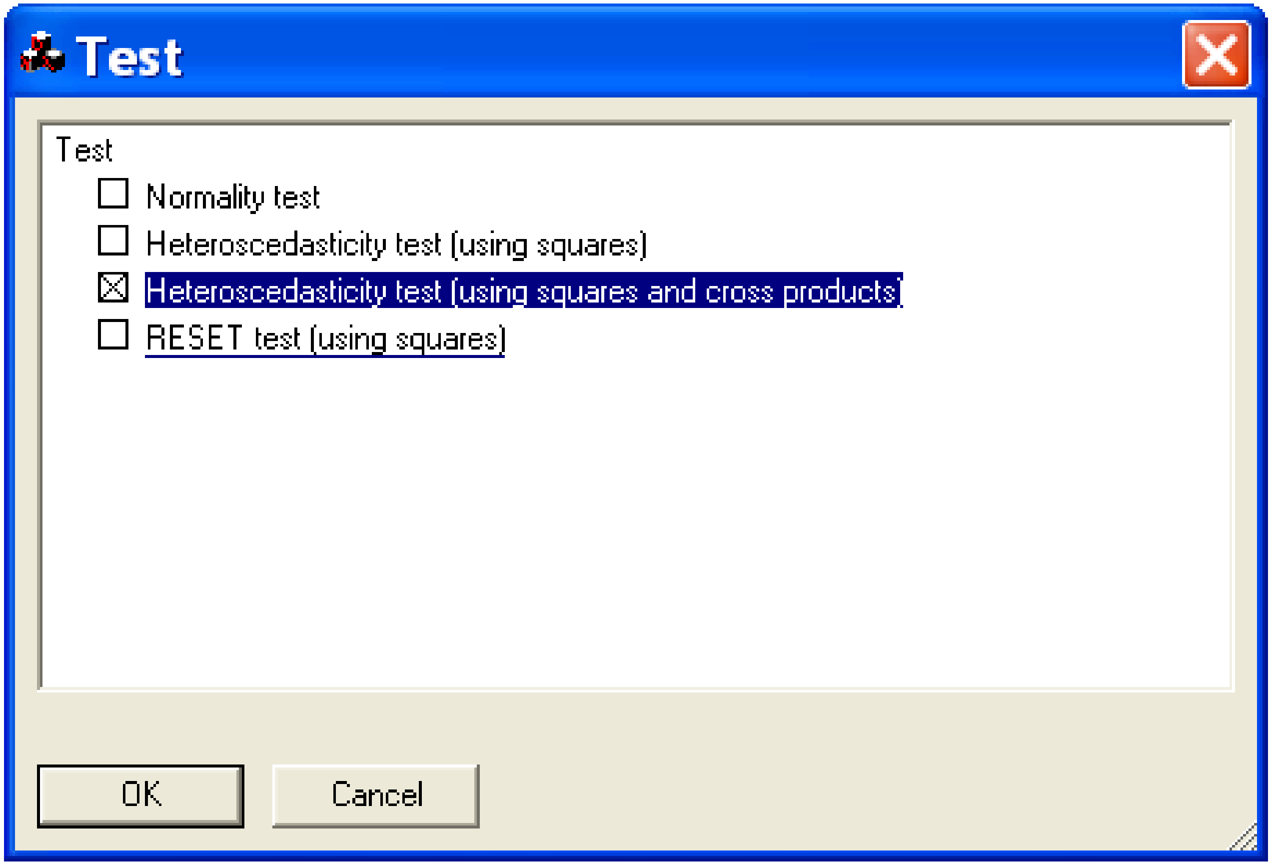
PcGive: Test de White



Test de White (modèle en niveau)

```
Heteroscedasticity coefficients:
              Coefficient t-value
LOTSIZE           0.62717  2.4527
SQRFT             -14.025 -1.6902
BDRMS              7454.8  1.5216
LOTSIZE^2 -5.0192e-006 -1.7990
SQRFT^2         0.0029752  1.7243
BDRMS^2           -770.65 -1.2873

RSS = 3.35361e+009 sigma = 6599.49
Regression in deviation from mean

Testing for heteroscedasticity using squares
Chi^2(6) = 20.602 [0.0022]**
F-form F(6,77) = 3.9229 [0.0018]**
```

PcGive: Test de White



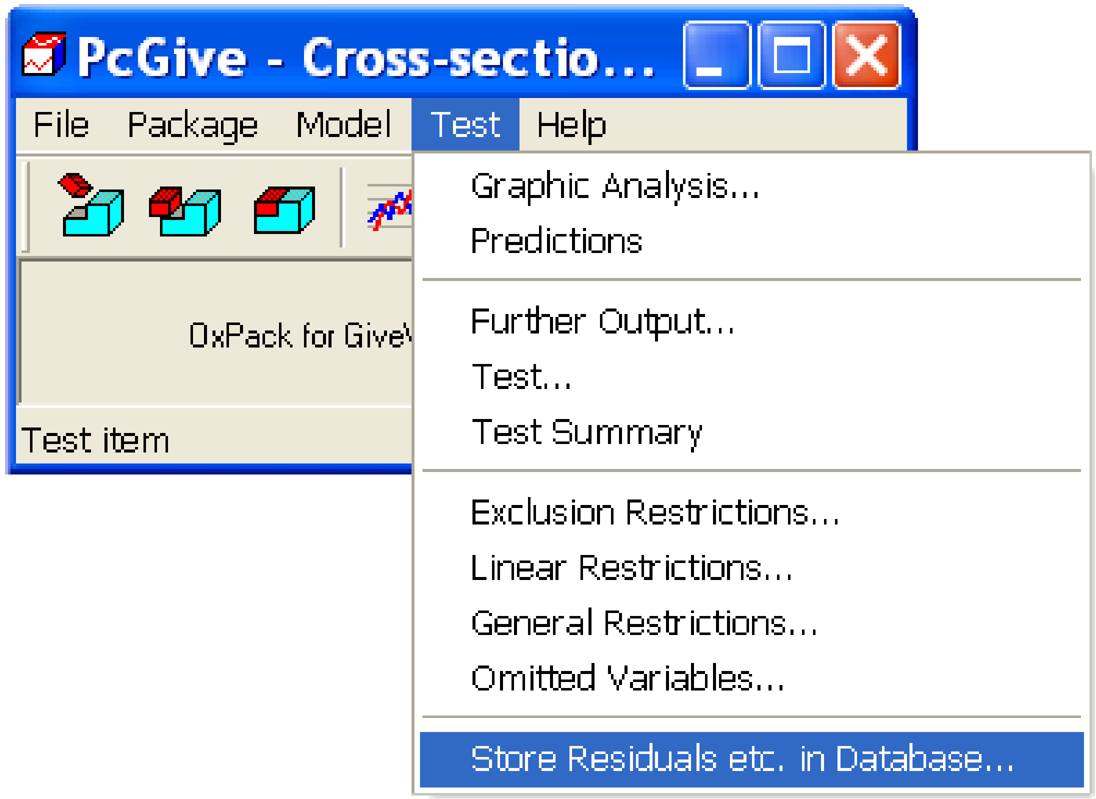
Test de White (modèle en niveau)

```
Heteroscedasticity coefficients:
                      Coefficient  t-value
LOTSIZE                   -1.8595  -2.8429
SQRFT                     -2.6739 -0.30067
BDRMS                     -1982.8 -0.35512
LOTSIZE^2            -4.9784e-007 -0.10471
SQRFT^2                0.00035226  0.18651
BDRMS^2                    289.75  0.37192
LOTSIZE*SQRFT          0.00045678   1.6068
SQRFT*BDRMS               -1.0209 -0.59643
LOTSIZE*BDRMS             0.31465   1.2157

Chi^2(9) = 33.732 [0.0001]**
F-form F(9,74) = 5.1107 [0.0000]**
```

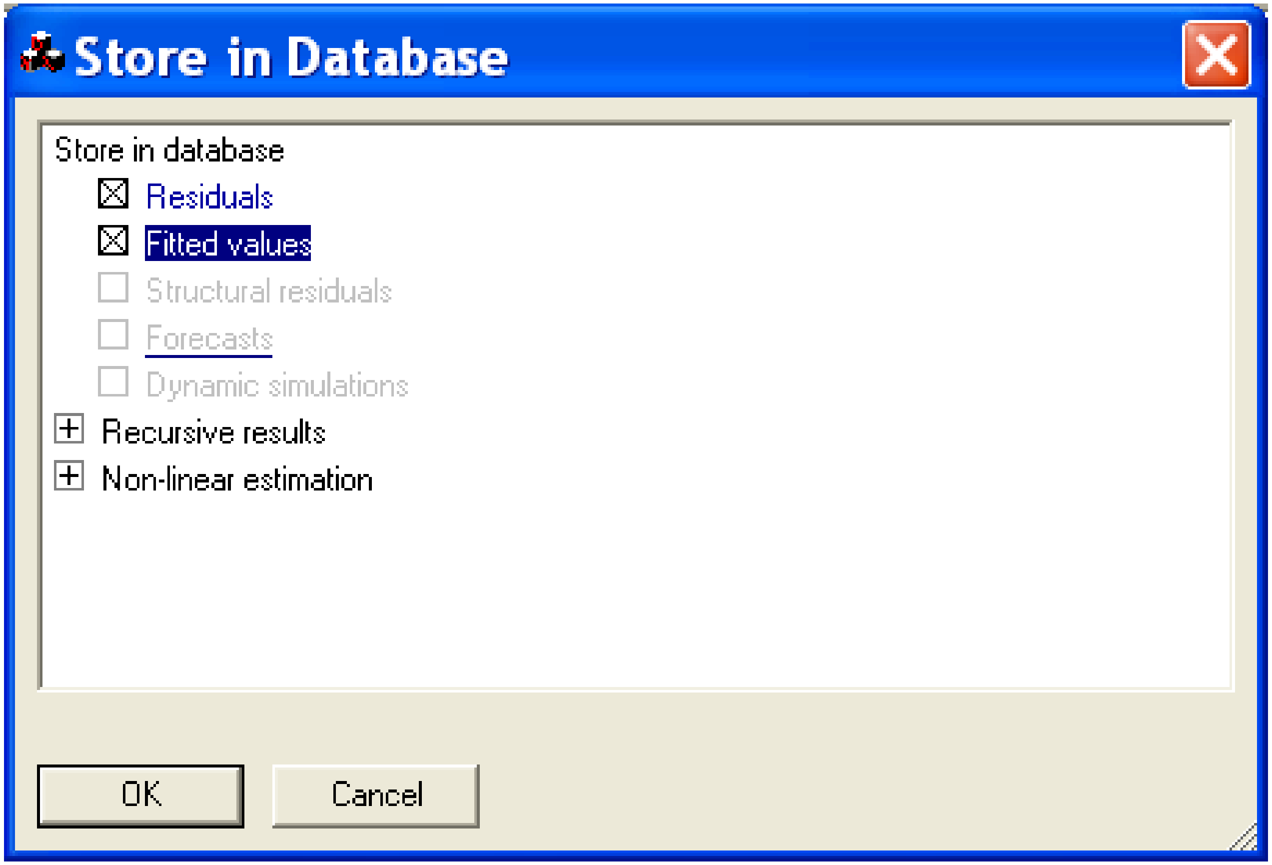
Test de White

- Une faiblesse du test de White est que le nombre de paramètres devient vite très grand.
- Exemple : 6 variables indépendantes $\rightarrow$ 27 paramètres.
- Idée : introduire les $\hat{y}$ qui sont en fait fonctions des x_k.
- Forme alternative : $\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + error \rightarrow$ Tests F et LM avec 2 restrictions.
- Avantage $=$ moins de paramètres à estimer. Inconvénient $=$ on ne sais plus ce qui cause l'hétéroscédasticité.

PcGive: Sauvegarder $\hat{y}$

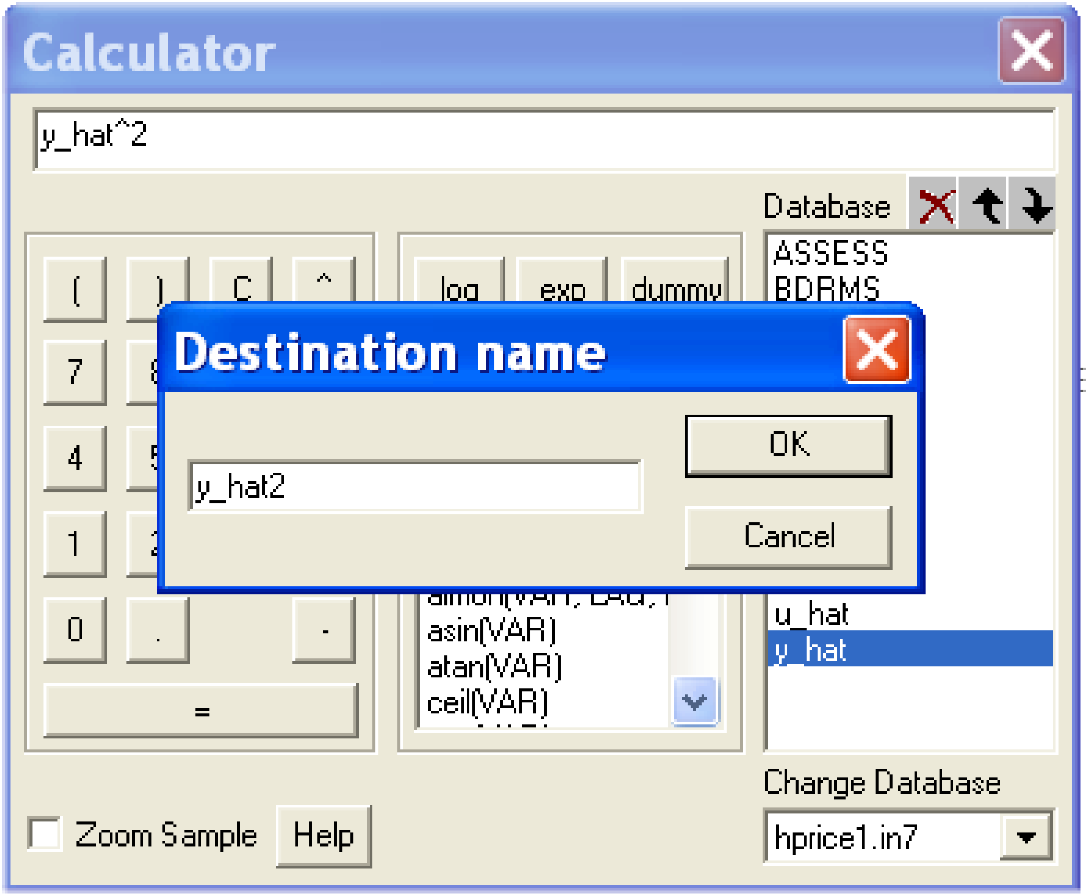


PcGive: Sauvegarder $\hat{y}$



PcGive: Sauvegarder $\hat{y}$

GiveWin: Créer $\hat{y}^2$



Test de White (modèle en niveau)

```
EQ(3) Modelling u_hat2 by OLS-CS (hprice1.in7)

             Coefficient t-prob
Constant         19071.6  0.035
y_hat           -119.655  0.027
y_hat2          0.208947  0.006

R^2    0.184868  F(2,85) = 9.639 [0.000]
no. of obs. 88  no. of parameters    3
```

$\rightarrow$ F-stat: $F(2,85) = 9.639[0.000]$.

$\rightarrow$ LM-stat : $88 \times 0.184868 = 16.268384[0.000]$.

Test de White (modèle en log)

```
EQ(5) Modelling u_hat2 by OLS-CS (hprice1.in7)

              Coefficient  t-prob
Constant          5.04683   0.135
y_hat            -1.70922   0.145
y_hat2           0.145135   0.154

R^2  0.0391735  F(2,85) = 1.733 [0.183]
no. of obs. 88  no. of parameters 3
```

→ F-stat: $F(2, 85) = 1.733[0.183]$.

→ LM-stat : $88 \times 0.0391735 \simeq 3.45[0.178]$.

Moindres carrés pondérés (MCP)

Forme d'hétéroscédasticité particulière

- Supposons que l'on connaisse la forme d'hétéroscédasticité $\rightarrow$ on va exploiter cela pour corriger les estimateurs MCO.
- $Var(u|x) = \sigma^2 h(x)$, avec $h(x) > 0$ pour garantir une variance positive.
- La variance n'est pas constante et est fonction des x: $\sigma_i^2 = Var(u_i|x_i) = \sigma^2 h(x_i) = \sigma^2 h_i$.
- Exemple. Fonction d'épargne: $saving_i = \beta_0 + \beta_1 inc_i + u_i$, et $Var(u_i|inc_i) = \sigma^2 inc_i$. $\rightarrow$ La variabilité de l'épargne augmente avec le niveau du revenu.

Estimation

- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i$.
- Si u_i est hétéroscédastique, comment estimer optimalement les β_j ?
- En fait : $E(u_i|x_i) = 0 \Rightarrow E(u_i/\sqrt{h_i}|x_i) = 0$ et $E(u_i^2|x_i) = \sigma^2 h_i \Rightarrow E(u_i^2/h_i|x_i) = E(u_i^2|x_i)/h_i = (\sigma_i^2 h_i)/h_i = \sigma^2$.
- $\rightarrow$ On retrouve les conditions de Gauss-Markov.
- $y_i/\sqrt{h_i} = \beta_0/\sqrt{h_i} + \beta_1(x_{i1}/\sqrt{h_i}) + \beta_2(x_{i2}/\sqrt{h_i}) + ... + \beta_k(x_{ik}/\sqrt{h_i}) + (u_i/\sqrt{h_i})$.
- $y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + ... + \beta_k x_{ik}^* + u_i^*$, avec $x_{i0}^* = 1/\sqrt{h_i}$, $x_{i1}^* = x_{i1}/\sqrt{h_i}$, etc.

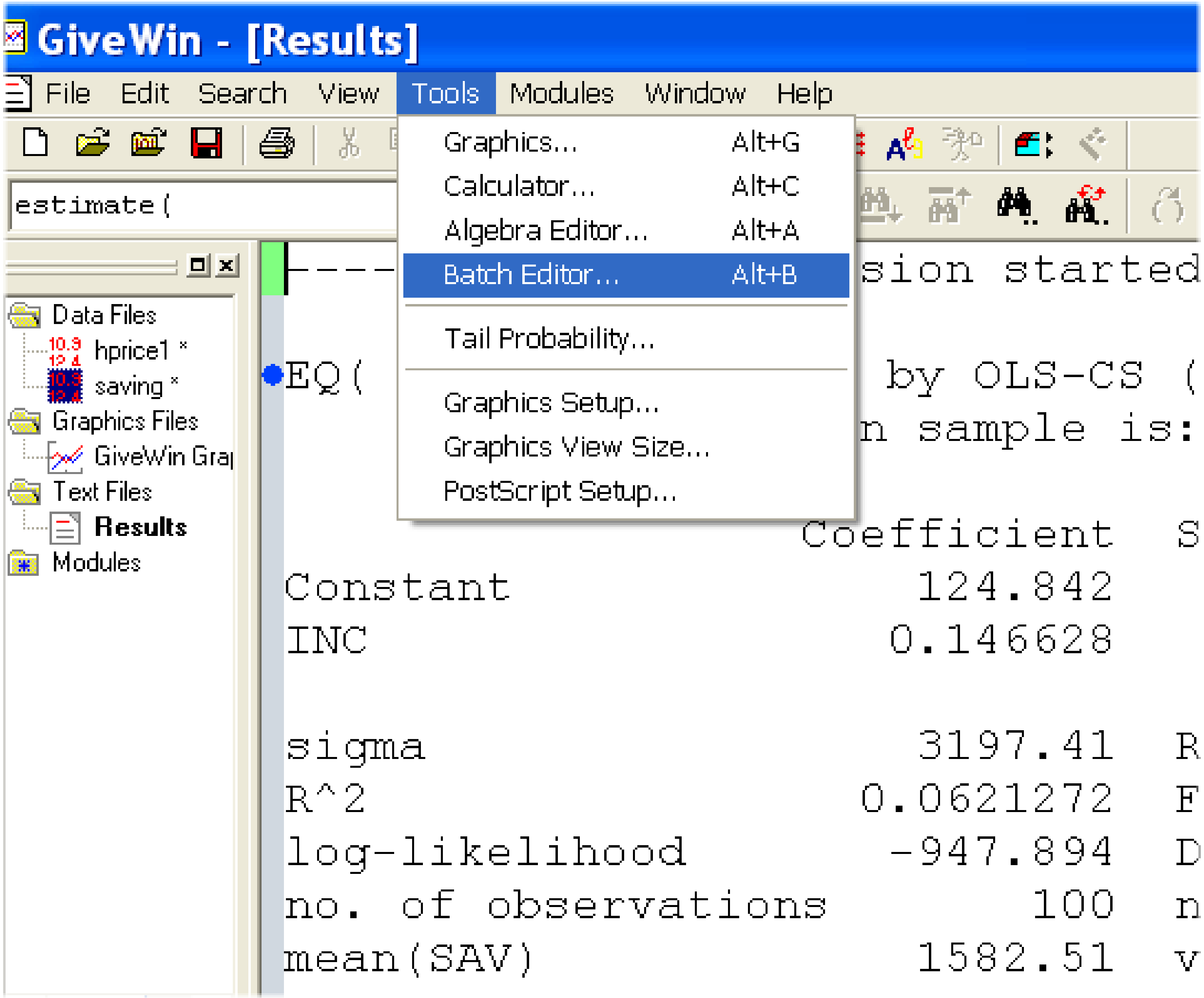
- Attention: dans cette nouvelle régression, l'ancien intercept est maintenant multiplié par x_{i0}^*.
- Exemple de la fonction d'épargne. On pose $Var(u_i|inc_i) = \sigma^2 h_i = \sigma^2 inc_i$.
  $\rightarrow$ Le modèle de régression estimé par la méthode des MCP devient :
  $saving_i/\sqrt{inc_i} = \beta_0(1/\sqrt{inc_i}) + \beta_1\sqrt{inc_i} + u_i^*$.
- Si le modèle de départ satisfait les 4 premières hypothèses de Gauss-Markov, alors le modèle transformé satisfait les cinq hypothèses de Gauss-Markov $\rightarrow$ on peut estimer ce modèle par MCO.
- Les estimateurs MCP sont en général différents des estimateurs MCO.

- Les estimateurs des MCG ont les bonnes propriétés : t, F, $\hat{\sigma}^2$ corrects pour l'inférence.
- Cette technique d'estimation s'appelle "moindres carrés généralisés" (MCG). Les MCP en sont un cas particulier (lorsque les MCG sont utilisés pour corriger des problèmes d'hétéroscédasticité).
- Idée : les estimateurs MCP sont obtenus en minimisant la somme des erreurs au carré pondérées (par $1/\sqrt{h_i}$).
- $\rightarrow \sum_{i=1}^{n} \left(y_i^* - \hat{\beta}_0 x_{i0}^* - \hat{\beta}_1 x_{i1}^* - \hat{\beta}_2 x_{i2}^* - ... - \hat{\beta}_k x_{ik}^* \right)^2$.

Exemple: Equation d'épargne

```
EQ(1) Modelling SAV by OLS-CS (saving.in7)

         Coefficient Std.Error t-prob
Constant     124.842     655.4  0.849
INC         0.146628   0.05755  0.012

sigma        3197.41 RSS        1.00189916e+009
R^2        0.0621272 F(1,98) = 6.492 [0.012]*
no. of obs       100 no. of parameters        2
```

GiveWin: MCG en Batch



Exemple: Equation d'épargne

```
module("PcGive");
package("PcGive");
usedata("saving.in7");
algebra
{
    SAV_W=SAV/sqrt(INC);
    CST_W=1/sqrt(INC);
    INC_W=INC/sqrt(INC);
}
system
{
    Y = SAV_W;
    Z = CST_W, INC_W;
}
estimate("OLS-CS", 1, 1, 100, 1);
```

MCG de l'équation d'épargne

```
EQ(2) Modelling SAV_W by OLS-CS (saving.in7)
      Coefficient Std.Error t-prob
CST_W    -124.953     480.9  0.796
INC_W    0.171756   0.05681  0.003

sigma   29.7118 RSS         86513.4819
no. of obs. 100 no. of parameters   2
```

Exemple: Equation d'épargne

Var dépendante	sav			
Var explicatives	MCO	MCP	MCO	MCP
inc	.147	.172	.109	.101
	(.058)	(.057)	(.071)	(.077)
$size$	-	-	67.66	-6.87
			(222.96)	(168.43)
$educ$	-		151.82	139.48
		-	(117.25)	(100.54)
age	-		.286	21.75
			(50.031)	(41.31)
$black$	-	-	518.39	137.28
			(1308.06)	(844.59)
$constante$	124.84	-124.95	-1605.42	-1854.81
	(655.39)	(480.86)	(2830.71)	(2351.80)
observations	100	100	100	100
R^2	.0621	.0853	.0828	.1042

Exemple: Equation d'épargne

- On a supposé que la variance est proportionnelle au revenu: $Var(u_i|inc_i) = \sigma^2 inc_i$.
- Les estimateurs MCP et MCO sont quelque peu différents mais les différences ne sont pas très grande $\rightarrow$ mêmes conclusions concernant la significativité.

Moindres carrés généralisés faisables

- En pratique, on connaît rarement la forme exact de l'hétéroscédasticité (comme on l'a supposé dans l'exemple précédent) $\rightarrow$ il faut estimer la relation entre les h_i et les x_i.
- $\rightarrow$ Utiliser des $\hat{h}_i$.
- $\rightarrow$ Moindres carrés généralisés faisables (MCGF).
- On va supposer une fonction $h(x)$ particulière (d'autres fonctions sont évidemment possibles) : $Var(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k)$.
- On va donc estimer les δ et non les supposer connus $u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k)\nu$.
- On va remplacer les u^2 par sa contrepartie observable $\hat{u}^2$.

Résumé des MCGF

- On va rendre le modèle linéaire :
  $log(\hat{u}^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + e$
  $\rightarrow \hat{h}_i = exp(\hat{\alpha}_0 + \hat{\delta}_1 x_1 + \hat{\delta}_2 x_2 + ... + \hat{\delta}_k x_k) = exp(\hat{g}_i)$.
- 1) Obtenir les résidus $\hat{u}$ par MCO.
- 2) Calculer $log(\hat{u}^2)$ et régresser $log(\hat{u}^2)$ sur les x (ou un sous-ensemble) $\rightarrow$ calculer $\hat{g}$.
- 3) Calculer $\hat{h} = exp(\hat{g})$.
- 4) Estimer l'équation suivante :
  $y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + ... + \beta_k x_{ik}^* + u_i^*$, en calculant les x_i^* grâce au poids $1/\sqrt{\hat{h}_i}$.
- Une alternative: régresser $log(\hat{u}^2)$ sur les $\hat{y}$ et les $\hat{y}^2$ pour calculer $\hat{g}$.

Exemple : Demande de cigarettes

- Estimer par MCO
  $cigs = \beta_0 + \beta_1 log(income) + \beta_2 log(cigprice) + \beta_3 educ + \beta_4 age + \beta_5 age^2 + \beta_6 restaurn + u_i$,
- où, $cigs$ est le nombre de cigarettes fumées par jour, ..., et $retaurn = 1$ si la personne habite dans un état où il y a des restrictions contre la cigarette dans les restaurants.
- Faire le test de Breusch Pagan (régresser les résidus au carré sur les variables explicatives. Calculer le test $LM = R^2_{\hat{u}^2}$ à comparer avec une χ^2_6.
- Réestimer $cigs = \beta_0 + \beta_1 log(income) + \beta_2 log(cigprice) + \beta_3 educ + \beta_4 age + \beta_5 age^2 + \beta_6 restaurn + u_i$ par MCGF et comparer.

Exemple : Demande de cigarettes

```
EQ(1) Modelling CIGS by OLS-CS (using smoke)

              Coefficient Std.Error   JHCSE t-prob
Constant          -3.639      24.08    25.84  0.888
LINCOME           0.8802     0.7278   0.6010  0.143
LCIGPRIC         -0.7508      5.773    6.086  0.902
EDUC             -0.5014     0.1671   0.1630  0.002
AGE               0.7706     0.1601   0.1394  0.000
AGESQ          -0.009022   0.001743 0.001476  0.000
RESTAURN          -2.825      1.112    1.011  0.005

Testing for heteroscedasticity using squares
Chi^2(10)=   36.146 [0.0001]**
and F-form F(10,789)=   3.6997 [0.0001]**
```

Exemple : Demande de cigarettes

```
EQ(2) Modelling log_u2 by OLS-CS (using smoke)

              Coefficient Std.Error t-prob
Constant         -1.92070     2.563  0.454
LINCOME          0.291541   0.07747  0.000
LCIGPRIC         0.195421    0.6145  0.751
EDUC           -0.0797036   0.01778  0.000
AGE              0.204005   0.01704  0.000
AGESQ         -0.00239214 0.0001855  0.000
RESTAURN        -0.627012    0.1183  0.000

R^2 0.247362  F(6,800) = 43.82 [0.000]**
```

Exemple : Demande de cigarettes

EQ(3) Modelling CIGS_W by OLS-CS (using smoke)

	MCPF		MCO	
	Coefficient	t-prob	Coefficient	t-prob
CST_W	5.635	0.752	-3.639	0.888
LINCOME_W	1.295	0.003	0.8802	0.143
LCIGPRIC_W	-2.940	0.510	-0.7508	0.902
EDUC_W	-0.4634	0.000	-0.5014	0.002
AGE_W	0.4819	0.000	0.7706	0.000
AGESQ_W	-0.005627	0.000	0.0090	0.000
RESTAURN_W	-3.461	0.000	-2.825	0.005

Exercices

- Exercice 8.1.
- Exercice 8.2.
- Exercice 8.3.
- Exercice 8.6
- Exercice 8.7

Problèmes supplémentaires de spécifications: chapitre 9

- Dans le chapitre 3, nous avons introduit 5 hypothèses dont l'hypothèse de non corrélation entre u et la (les) variable(s) explicative(s) x (MLR.3) .
- Nous allons reconsidérer ce problème dans plusieurs cas particuliers : erreurs de mesures et biais de variables omises $\rightarrow$ Comment traiter cela et peut-on solutionner tout en utilisant les MCO ?
- On va voir également les conséquences de formes fonctionnelles non appropriées

Mauvaise spécification due à la forme fonctionnelle

Problèmes de forme fonctionnelle

- De manière générale, problème de prise en compte de la relation entre y et les $x \rightarrow$ différents problèmes.
- Soit le vrai modèle:
  $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$.
- Si on omet $exper^2$ dans la régression, cela génère 2 problèmes.
- Biais dans l'estimation de β_0, β_1 et β_2. La taille du biais dépend de la taille de β_3 et de la corrélation entre les x.
- On appréhende très mal le rendement de l'expérience professionnelle qui est donné par $\beta_2 + 2\beta_3 exper$.

Problèmes de forme fonctionnelle

- Un autre problème : travailler sur $wage$ alors que la relation correcte est de travailler sur $log(wage) \rightarrow$ on obtient des estimateurs biaisés ou non consistents.
- Pour tester l'opportunité d'une relation non linéaire par l'inclusion d'un terme quadratique (exemple: $exper^2$), on peut utiliser un test d'exclusion en F (cfr chap.4).
- Néanmoins, on peut utiliser des tests de non linéarité plus généraux: Tests RESET et tests emboités.

Exemple: arrestations criminelles

Var dépendante	$narr86$			
Var explicatives	coeff	ec-t	coeff	ec-t
$pcnv$	-.133	.040	.533	.154
$pcnv^2$	-		-.730	.156
$avgsen$	-.011	.012	-.017	.012
$tottime$	.012	.009	.012	.009
$ptime86$	-.041	.009	.287	.004
$ptime86^2$	-		-.0296	.0039
$qemp86$	-.051	.014	-.014	.017
$inc86$	-.0015	.0003	-.0034	.0008
$inc86^2$	-		.000007	.000003
$black$	.327	.045	.292	.045
$hispan$	.194	.040	.164	.039
$Constant$	.596	.036	.505	.037
$n = 2724$	$R^2 = .0723$		$R^2 = .1035$	

Exemple: arrestations criminelles

- Dans la seconde équation, les termes quadratiques sont tous individuellement significatifs (voir les $t-stats$).
- Ils sont également significatifs collectivement: un test en F donne $F = 31.37$ avec $3dl \rightarrow p - valeur = .000$.
- Les effets de $pcnv$,$ptime86$ sont concaves et celui de $inc86$ est convexe .
- On voit que certains paramètres d'autres variables changent entre les 2 régressions $\rightarrow$ biais de variable(s) omise(s).

Test RESET

- Test pour détecter un problème général de mauvaise forme fonctionnelle.
- On estime le modèle général:
  $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$.
- L'inclusion de termes quadratiques pour tester leur significativité fait que le nombre de liberté peut diminuer fortement. En outre, cela ne teste qu'une forme de non linéarité particulière.
- On va estimer un nouveau modèle étendu:$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u$.
- les termes $\hat{y}^2$ et $\hat{y}^3$ dépendent de manière non linéaire des $x \rightarrow$ suggèrent des relations non linéaires.
- On teste leur significativité conjointe : test en F de $H_0 : \delta_1 = 0, \delta_2 = 0 \rightarrow$ suit une $F(2, n-k-3)$.

Exemple: Prix du logement

- $price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$.
- RESET: $F = 4.87$, $p - val = .012$.
- $lprice = \beta_0 + \beta_1 llotsize + \beta_2 lsqrft + \beta_3 bdrms + u$.
- RESET: $F = 2.56$, $p - val = .084$.
- $\rightarrow$ la specification log-log est préférée.

Tests emboîtés

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$.
- On évalue ce modèle contre le modèle alternatif : $y = \beta_0 + \beta_1 log(x_1) + \beta_2 log(x_2) + u$.
- On ne peut pas utiliser un test en F car les modèles sont non emboités : aucun modèle n'est un cas particulier de l'autre .
- Première approche : estimer $y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 log(x_1) + \gamma_4 log(x_2) + u$.
- On peut tester alors : $H_0 : \gamma_1 = \gamma_2 = 0 \longrightarrow$ rejet du modèle de base en faveur du second modèle
- On peut aussi tester : $H_0 : \gamma_3 = \gamma_4 = 0 \longrightarrow$ rejet du second modèle en faveur du modèle de base

Variables proxy à la place des variables omises

Variable omise

- Problème lorsqu'un modèle exclut une variable pertinente car celle-ci n'est pas disponible.
- Exemple : ommission de la compétence professionnelle:
  $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u$.
- Si $educ$ et $abil$ sont corrélés, omettre $abil$ et le mettre dans $u \rightarrow$ estimateurs biaisés.
- Idée : utiliser une variable proxy observable de la variable omise. Exemple : QI.

Variables proxy

- Modèle de base $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u$.
- Si x_3^* est non observé, remplacer par x_3 qui est reliée à la variable omise: $x_3^* = \delta_0 + \delta_3 x_3 + \nu_3$. S'il s'agit d'une vraie proxy, $\delta_3 > 0$, i.e. on a une relation positive.
- Idée : régresser y sur x_1, x_2 et x_3 $\rightarrow$ cela permettra peut- être d'éviter un biais d'estimation sur x_1 et x_2.
- 2 hypothèses nécessaires pour que β_1 et β_2 soient consistents.

Variables proxy

- Hypothèse 1: u est non corrélé avec x_1, x_2 et x_3^* ;u est non corrélé avec x_3.
- Hypothèse 2: ν_3 est non corrélé avec x_1, x_2 et $x_3 \rightarrow x_3$ est une bonne proxy car alors $E(x_3^*|x_3) = \delta_0 + \delta_3 x_3$.
- Exemple: Le niveau général de la compétence (non observée) change avec le niveau du QI, pas avec celui de l'éducation et de l'expérience professionnelle : $E(abil|educ, exper, IQ) = E(abil|IQ) = \delta_0 + \delta_3 IQ$.

- Si les hypothèses sont satisfaites, alors le modèle devient: $y = (\beta_0 + \beta_3\delta_3) + \beta_1 x_1 + \beta_2 x_2 + \beta_3\delta_3 x_3 + u + \beta_3\nu_3$.
- Sous les 2 hypothèses, le terme d'erreur composite $e = u + \beta_3\nu_3$ est non corrélé avec x_1, x_2 et x_3. $\rightarrow$ les estimateurs de β_1 et β_2 sont non biaisés;
- Parfois, l'estimateur de la variable proxy ($\alpha_3 = \beta_3\delta_3$) est très intéressant également .
- Si la variable proxy ne respecte pas toutes les hypothèses, cela peut générer des estimateurs biaisés. Exemple : $x_3^* = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \nu_3$.
- Le modèle devient : $y = (\beta_0 + \beta_3\delta_3) + (\beta_1 + \beta_3\delta_1)x_1 + (\beta_2 + \beta_3\delta_2)x_2 + \beta_3\delta_3 x_3 + u + \beta_3\nu_3$.
- On voit que $plim(\hat{\beta}_1) = \beta_1 + \beta_3\delta_1$ et $plim(\hat{\beta}_2) = \beta_2 + \beta_3\delta_2$ $\rightarrow$ estimateurs non consistents.

Exemple

Var dép.	$log(wage)$	$n = 935$	
Var expl.	(1)	(2)	(3)
$educ$	.065 (.006)	.054 (.007)	.018 (.041)
$exper$	.014 (.003)	.014 (.003)	.014 (.003)
$tenure$	.012 (.002)	.011 (.002)	.011 (.002)
$married$	.199 (.039)	.200 (.039)	.201 (.039)
$south$	-.091 (.026)	-.080 (.026)	-.080 (.026)
$urban$	.184 (.027)	.182 (.027)	.184 (.027)
$black$	-.188 (.038)	-.143 (.039)	-.147 (.040)
IQ	-	.0036 (.0010)	.0009 (.0052)
$educ.IQ$	-	-	-.00034 (.00038)
$Constant$	5.395 (.113)	5.176 (.128)	5.648 (.546)
R^2	.253	.263	.263

Exemple: éducation et QI

- Question importante : quel effet sur le rendement de l'éducation ? Si la compétence non observée est corrélée avec $educ$, alors le rendement estimé ($.065$) est biaisé
- On intègre le QI (IQ) : le QI est significatif et le rendement de $educ$ diminue : $.054 \rightarrow$ exemple de biais dû à l'omission d'une variable pertinente.
- On intègre aussi un terme d'interaction entre $educ$ et IQ (le rendement de l'éducation est peut-être plus élevé pour les travailleurs intelligents) $\rightarrow$ non significatif $\rightarrow$ on préfère les estimateurs de la colonne précédente.

Variables dépendente laggée

- Si on n'a pas de variable proxy, on peut parfois appréhender l'effet des variables omises par une variable dépendante retardée . $\rightarrow$ permet de prendre en compte des facteurs historiques qui interagissent avec les variables explicatives du modèle.
- Exemple : effet des dépenses de justice ($expend$) et du chômage ($unem$) sur le taux de criminalité ($crime$). $crime = \beta_0 + \beta_1 unem + \beta_2 expend + \beta_3 crime_{-1} + u$.
- On s'attend évidemment à $\beta_3 > 0$. Il est tout à fait possible que les villes avec un haut taux de criminalité dépensent plus en justice $\rightarrow$ facteurs non observés. $\rightarrow$ pris en compte par $crime_{-1}$ (taux de criminalité passé).

Exemple

Var dép.	$log(crmrate_{87})$	$n = 46$
Var expl.	(1)	(2)
$unem_{87}$	-.029 (.032)	.009 (.020)
$log(lawexp_{87})$	.203 (.173)	-.140 (.109)
$log(crmrate_{82})$	-	1.194 (.132)
$Constant$	3.34 (1.25)	.076 (.821)
R^2	.057	.680

MCO et erreurs de mesure

Erreurs de mesure

- Parfois, les données récoltées sont entachées d’erreurs de mesure.
- Exemple: le taux de taxation marginal d’une famille $\rightarrow$ appréhendé par le taux moyen $\rightarrow$ erreurs probables.
- Il y a des similitudes dans le problème d’erreurs de mesure et le problème des variables omises.2 différences néanmoins.
- Ici, on n’observe pas la variable pertinente mais on en a une information quantitative qui est entachée d’erreur de mesure.
- Ici, la variable d’intérêt est aussi la variable entachée d’erreur de mesure alors que dans le cas précédent, c’est surtout les conséquences sur les autres coefficients qui nous intéressent.

Erreur sur la variable dépendente

- Modèle de base :$y^* = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$.
- Problème : y^* est observé avec erreur par y : on a une erreur de mesure $e_0 = y - y^*$. Exemple : épargne de la famille : la famille reporte cela avec erreur (omission de certains postes, ...).
- On estime donc: $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u + e_0$. Quelles conséquences sur l'estimateur MCOs ?
- Si $E(e_0) = 0$, alors l'estimateur de β_0 est non biaisé $\rightarrow$ Sinon, biais dans l'intercept β_0(pas grave).
- Hypothèse cruciale: e_0 non corrélé avec les x_j $\rightarrow$ les estimateurs MCO sont corrects et inférence $(t, F, ...)$ valide. Si u et e_0 non corrélés, alors $Var(e_0 + u) = \sigma_u^2 + \sigma_0^2 > \sigma_u^2$.$\rightarrow$ les estimateurs sont estimés avec plus d'incertitude (écart-types plus élevés).

Erreur sur une var explicative

- Considérons le cas où l'erreur porte sur une var explicative: $e_1 = x_1 - x_1^*$. On suppose qu'en moyenne, l'erreur est égale à zéro: $E(e_1) = 0$.
- Modèle de base :$y = \beta_0 + \beta_1 x_1^* + u$. On suppose comme auparavant que u est non corrélé avec x_1 et x_1^*.
- Sous quelles hypothèses l'estimation par MCO permet -elle de générer des estimateurs consistents en utilisant x_1 à la place de x_1^* ? Cela dépend de la nature de e_1. $\rightarrow$ 2 hypothèses extrêmes opposées.
- Hypothèse 1: $Cov(e_1, x_1) = 0$. Dans ce cas, on a $y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1)$. $\rightarrow (u - \beta_1 e_1)$ est non corrélé avec x_1 et donc les estimateurs de β_0 et de β_1 sont consistents. Comme $Var(u - \beta_1 e_1) = \sigma_u^2 + \beta_1^2 \sigma_{e_1}^2 > \sigma_u^2$, la précision des estimateurs est moindre.

Erreur sur une var explicative

- Hypothèse 2: Erreurs classiques sur variables : $Cov(e_1, x_1^*) = 0$. Erreur de mesure est non corrélée avec la vraie variable non observée
- Dans ce cas, le problème se pose pour l'hypothèse précédente: $Cov(x_1, e_1) = E(x_1 e_1) = E(x_1^* e_1) + E(e_1^2) = 0 + \sigma_{e_1}^2 = \sigma_{e_1}^2$.
- Cette corrélation pose problème car $Cov(x_1, u - \beta_1 e_1) = -\beta_1 Cov(x_1, e_1) = - - \beta_1 \sigma_{e_1}^2$. $\rightarrow$ Les MCO vont donner des estimateurs biaisés et non consistents.

Erreur sur une var explicative

- En utilisant les propriétés asymptotiques, on peut caractériser le biais.
-
$$
\begin{aligned}
plim(\hat{\beta}_1) &= \beta_1 + \frac{Cov(x_1, u - \beta_1 e_1)}{Var(x_1)} \\
&= \beta_1 - \frac{\beta_1 \sigma_{e_1}^2}{(\sigma_{x_1^*}^2 + \sigma_{e_1}^2)} \\
&= \beta_1 (1 - \frac{\sigma_{e_1}^2}{(\sigma_{x_1^*}^2 + \sigma_{e_1}^2)}) \\
&= \beta_1 (\frac{\sigma_{x_1^*}^2}{(\sigma_{x_1^*}^2 + \sigma_{e_1}^2)}).
\end{aligned}
$$

.

Erreur sur une var explicative

- Le terme $(\frac{\sigma^2_{x_1^*}}{(\sigma^2_{x_1^*}+\sigma^2_{e_1})})$ est toujours < 1 et est appelé le biais d'atténuation. $\rightarrow$ si $\beta_1 > 0$, alors $\hat{\beta}_1$ sous-estimera β_1.
- Si erreur de mesure est petite relativement à la variable inobervée, alors $\frac{\sigma^2_{x_1^*}}{\sigma^2_{e_1}}$ sera grand et le biais d'estimation sera faible.

Cas de régression multiple

- $y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + \beta_3 x_3 + u$.
- L'hypothèse cruciale est une fois de plus de savoir si e_1 et x_1 sont corrélés.
- Si ce n'est pas le cas, la régression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u - \beta_1 e_1$ donnera des estimateurs consistents.
- Sous Erreurs classiques sur variables :

  $plim(\hat{\beta}_1) = \beta_1 (\frac{\sigma^2_{r_1^*}}{(\sigma^2_{r_1^*} + \sigma^2_{e_1})})$ dans laquelle r_1^* est erreur de l'équation $x_1^* = a_0 + a_1 x_2 + a_2 x_3 + r_1^*$.

Exemple: effet du revenu familial

- Supposons que l'on veuille voir l'impact du revenu familial ($faminc$) sur les résultats au collège ($colGPA$).
- Régression :
$colGPA = \beta_0 + \beta_1 faminc^* + \beta_2 hsGPA + \beta_3 SAT + u$.
- Les familles ont tendance à reporter leur revenu avec des erreurs. Si $faminc = faminc^* + e_1$ et si l'hypothèse d'erreurs classiques sur variables tient, alors biais de $\hat{\beta}_1$ vers zéro et on sous-estimera l'effet du revenu sur les résultats scolaires.

Autres problèmes

Données manquantes

- Souvent, variables manquantes sous la forme d'observations manquantes pour plusieurs ou une variable
- Exemple: poids du bébé à la naissance : 1388 observations mais pour 197, pas d'info sur l'éducation de la mère et/ou du père
- Si on enregistre les données manquantes, les logiciels appliquent l'approche la plus simple : ignorer les observations pour lesquelles données manquantes. Exemple précédent : régression expliquant le poids du bébé en fonction de l'éducation du père et de la mère : 1388-197 observations utilisées.

Echantillon non aléatoire

- Si les données sont manquantes de manière aléatoire cette approche produit des estimateurs moins précis mais non biaisés (hyp M.L.R.2 OK). $\rightarrow$ très souvent utilisée.
- Parfois l'échantillon obtenu est un échantillon non-aléatoire de la population. Exemple : niveau d'éducation manquant pour les peu qualifiés ; données de salaire pour les travailleurs avec un QI au-delà d'un certain seuil.$\rightarrow$ l'hypothèse M.L.R. 2 est violée. Quid pour les estimateurs MCO ?

Echantillon non aléatoire

- On peut procéder à une sélection d'échantillon exogène sur base d'une variable indépendante. Exemple:fonction d'épargne $\rightarrow$ utiliser l'épargne pour les individus de 35 ans et plus. Hypothèse cruciale: $E(saving|income, age, size)$ identique pour ce sous-échantillon. $\rightarrow$ l'estimateur MCO peut être alors utilisé.
- Si sélection est basée sur la variable dépendante $\rightarrow$ possibilité de biais de sélection.
- Exemple : $wealth = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 age + u$. Si seulement individus avec une richesse ($wealth$)supérieure à USD 75000 inclus dans l'échantillon, estimateurs biaisés et non consistents car $E(wealth|educ, age, exper) \neq E(wealth|educ, age, exper, wealth > 75000)$

Echantillon non aléatoire

- Echantillonnage stratifié : la population est divisée en strates qui ne se chevauchent pas. Exemple : strates basées sur le sexe. $\longrightarrow$ on peut faire de la surrépresentation de certains groupes minoritaires dans la population pour étudier des phénomènes propres à ces groupes. Exemple: sur-représenter les femmes militaires pour étudier les facteurs influençant le salaire dans l'armée dont la discrimination.
- Les échantillons stratifiés peuvent amener des biais de sélection.Cas par cas. Exemple : impact sur le salaire offert au travailleur :
$log(wage^0) = \beta_0 + \beta_1 educ + \beta_2 exper + U$.

Echantillon non aléatoire

- On n'a que le salaire offert pour les travailleurs, pas pour les sans-emploi; MCO biaisés ? pas nécessairement car l'absence de salaire offert peut résulter d'une décision individuelle (non travailler à ce salaire), non pas du niveau de salaire offert.

Valeurs aberrantes - outliers

- Dans les petits échantillons, l'occurrence de quelques observations peut inluencer la valeurs des MCO. $\rightarrow$ valeurs aberrantes ou outliers.
- Raison: les MCOs sont basés sur la minimisation de la somme des résidus $\rightarrow$ le résidu lié à un outlier reçoit un poids important.
- Les outliers peuvent provenir de mauvaise saisie des données $\rightarrow$ corriger les données après détection (statistiques descriptives).
- Un outlier peut provenir d'une observation issue d'un autre (petite) population (exemple : une firme avec un niveau de CA plus de 4 fois plus grand que les autres firmes) $\rightarrow$ Faire régression avec et sans cette observation et rapporter les résultats.

Exemple: Dépenses de R et D

- On étudie les déterminants de l'importance de la R et D sur 32 firmes: $rdintens = \beta_0 + \beta_1 sales + \beta_2 profmarg + u$.
- 1 firme présente un niveau de ventes $sales$ de USD 40 millions alors que toutes les autres ont moins de USD 20 millions.
- Le coefficient $\hat{\beta}_1$ est très influencé par l'inclusion ou non de cette firme : .000053(.000044) avec ; .000186(.000084) sans $\rightarrow$ la taille et la significativité de l'effet dépendent de l'échantillon.
- Certains tests permettent de détecter automatiquement ces outliers.