Licence AES 3ème année

Université de Lille II

Notes du cours d'introduction à l'économétrie

P. De Vreyer

Partie IV: Propriétés des estimateurs des Moindres Carrés Ordinaires.

Nous savons maintenant comment calculer les estimateurs des MCO. Mais une question demeure: étant donné que la vraie valeur des coefficients du modèle ne sera jamais connue, quelle confiance peut-on accorder aux valeurs estimées de ces coefficients ? Comme nous l'avons vu précédemment, deux critères sont généralement retenus pour évaluer la qualité d'un estimateur: l'absence de biais et l'efficacité. Dans cette partie nous allons examiner les conditions sous lesquelles les estimateurs des MCO satisfont ces deux critères, puis nous verrons la façon dont on peut évaluer le degré de confiance que l'on peut accorder aux estimations des paramètres.

1. La dimension aléatoire des estimateurs des MCO.

Nous allons de nouveau insister sur la dimension aléatoire des estimateurs et sur la notion de distribution d'échantillonnage. Souvenons-nous que le modèle s'écrit:

$$Y = \alpha + \beta X + \epsilon$$

Les estimateurs des MCO sont des fonctions du couple de vecteurs aléatoires (X,Y). A ce titre ce sont donc eux-mêmes des variables aléatoires, c'est à dire des variables dont la valeur ne peut être prédite *a priori* par l'économètre parce qu'avant que l'échantillon soit tiré et les observations réalisées, il n'est pas possible de savoir quelle sera la valeur prise par X et Y. C'est pour cette raison que les estimateurs ont une moyenne et une variance auxquelles il faut s'intéresser pour effectuer des comparaisons entre les différentes stratégies d'estimation. Nous allons voir dans ce qui suit que, sous certaines conditions, les estimateurs des MCO sont sans biais et efficaces.

2. Hypothèses concernant la perturbation.

Les propriétés des estimateurs des MCO dépendent de façon étroite de celles du terme d'erreur ϵ. Trois conditions doivent être vérifiées par ϵ pour garantir que les estimateurs des MCO aient les bonnes propriétés.

i. Première condition: $E(\epsilon_i | X_i) = 0$.

   Cette condition est la plus importante de toutes. Pour bien la comprendre reprenons l'écriture du modèle au niveau individuel: $Y_i = \alpha + \beta X_i + \epsilon_i$. Selon cette écriture, le caractère aléatoire de la variable dépendante, Y_i, a deux origines: une origine dite « déterministe », représentée par $\alpha + \beta X_i$, et une origine dite « stochastique » ou « non déterministe », représentée par ϵ_i. L'origine stochastique est par définition indéterminée. Cela signifie, d'une part, qu'un très grand nombre de variables peuvent intervenir dans le terme ϵ_i et, d'autre part, que ces variables ne sont pas observées par l'économètre. Or, l'objectif de l'estimation économétrique est d'identifier le rôle de X dans la détermination de Y. Pour ce faire, il est nécessaire que l'action de X passe *entièrement* par la composante déterministe. Or, ce n'est pas le cas si $E(\epsilon_i | X_i)$ est une fonction de

X_i ou, ce qui revient au même, si parmi les variables dont la réalisation détermine ϵ_i certaines sont corrélées avec X_i, car alors on ne peut pas prétendre avoir identifié totalement le rôle de X en estimant les coefficients α *et* β.

Supposons par exemple que l'on observe le salaire et l'éducation d'un échantillon d'individus. On cherche à identifier l'effet de l'éducation sur le salaire. Pour ce faire on régresse le salaire sur l'éducation (mesurée en années d'études), ce qui conduit à l'estimation d'un coefficient β qui représente l'effet d'une année d'études supplémentaire sur le salaire. Sous quelle hypothèse cette interprétation est-elle correcte ? Selon les termes employés précédemment, l'action des années d'études sur le salaire doit passer entièrement par la composante déterministe du modèle. La composante stochastique doit être indépendante de l'éducation. Or on peut facilement imaginer des situations où cette indépendance n'est pas garantie: par exemple si les individus ayant de grandes capacités à travailler acquièrent plus d'éducation que les autres en moyenne et si les entreprises rémunèrent cette capacité, ce qui semble probable, alors il est possible, lorsque l'on régresse le salaire sur l'éducation, que ce qui est mesuré n'est pas tant l'effet de l'éducation en lui-même que l'effet de la capacité à travailler. Dans cet exemple, la capacité à travailler est une variable non observée par l'économètre, qui contribue à la détermination du terme stochastique et qui est corrélée avec l'éducation, de sorte que la condition $E(\epsilon_i | X_i)=0$ n'est pas vérifiée. Nous verrons ci-dessous que ceci conduit à biaiser les estimateurs des MCO.

ii. Deuxième condition: $V(\epsilon_i | X_i)=\sigma^2 \textit{ pour tout } i.$

Cette condition, appelée homoscédasticité, est moins importante que la première. Lorsqu'elle n'est pas respectée, les estimateurs des MCO sont inefficaces (en d'autres termes, on peut en trouver d'autres qui sont plus efficaces). Elle signifie que la variance de la composante stochastique de Y doit être constante et ne pas dépendre de la variable qui apparaît dans la composante déterministe.

Là encore il est facile d'imaginer des situations où cela n'est pas le cas. Reprenons l'exemple précédent. De nouveau il s'agit de régresser le salaire sur l'éducation mesurée en années. La seconde condition signifie quel que soit le niveau d'éducation des individus, les variations du revenu autour d'une « moyenne » égale à la composante déterministe $\alpha+\beta X_i$, ont la même variance, autrement dit qu'elles ont le même ordre de grandeur. Or, il est raisonnable de penser que plus le niveau d'éducation est élevé et plus les variations de revenus entre personnes ayant une éducation identique sont importantes. Autrement dit, une personne sans qualification a fort peu de chances d'avoir un salaire élevé, alors qu'une proportion non négligeable de personnes ayant un niveau d'éducation élevé ont un salaire faible. En d'autres termes, plus le niveau d'éducation est élevé et moins la connaissance du niveau d'éducation d'une personne apporte d'information sur son salaire.

Lorsque la condition d'homoscédasticité est violée (on parle alors hétéroscédasticité), les estimateurs des MCO ne sont pas efficaces, car ils ne tiennent pas compte du fait que toutes les observations sur la variable explicative n'apportent pas la même quantité d'information sur la variable expliquée. D'autres estimateurs, qui en tiennent compte, sont plus efficaces.

iii. Troisième condition: $cov(\epsilon_i, \epsilon_j | X_i, X_j)=0 \textit{ pour tout } i \textit{ et } j.$

Cette hypothèse reçoit le nom « d'absence d'autocorrélation ». Comme l'hypothèse d'homoscédasticité, lorsqu'elle est violée les estimateurs des MCO sont non efficaces.

L'intuition pour le comprendre est du même ordre que pour la condition d'homoscédasticité. Reprenons de nouveau l'exemple précédent. Supposons que l'on dispose d'observations sur deux individus: chacun d'entre-eux a un niveau d'éducation et un niveau de salaire. Supposons que la condition d'absence d'autocorrélation ne soit pas vérifiée. Cela signifie que le salaire de ces individus est, bien sûr, déterminé par leur niveau d'éducation, mais aussi par une composante stochastique *dont une partie est commune* aux deux individus (par exemple le salaire de ces deux personnes peut être relativement faible, parce qu'elles sont toutes les deux employées dans une branche d'activité en crise). Dans ces conditions, les observations des variables aléatoires Y_i ne sont pas indépendantes. Comme l'estimateur des MCO ne tient pas compte de cette absence d'indépendance il n'est pas efficace. En fait, si l'on sait qu'une corrélation existe entre les composantes stochastiques des variables Y_i, il est préférable d'utiliser cette information pour mieux estimer les paramètres du modèle, ce que ne fait pas l'estimateur des MCO.

A ces trois conditions s'ajoute une quatrième qui concerne cette fois les observations des variables explicatives du modèle, regroupées dans la matrice *X*. Si *k* est le nombre de variables explicatives, le rang de la matrice *X* doit être égal à *k*, ce qui signifie qu'il ne doit pas exister une relation linéaire parfaite entre les différentes variables explicatives. En effet, si une telle relation existe, alors une variable peut s'écrire en fonction des autres et n'apporte donc rien à l'explication de la variable dépendante. Dans ce cas une variable doit être retirée de la liste.

3. Propriétés statistiques des estimateurs des MCO.

Les conditions présentées dans le paragraphe précédent sont appelées conditions de Gauss-Markov. Elles conduisent à l'énoncé du théorème du même nom, dont la démonstration est l'objet de ce paragraphe.

Théorème de Gauss-Markov.
Sous les conditions de Gauss-Markov, les estimateurs des MCO du modèle linéaire simple sont, dans la classe des estimateurs linéaires, sans biais et efficaces.

Avant d'en donner la preuve, précisons ce que signifie la linéarité des estimateurs des MCO. Les estimateurs sont linéaires, car ils s'expriment sous la forme de fonctions linéaires des observations de la variable expliquée Y_i. En effet:

$$\hat{\beta}=\frac{cov(X,Y)}{V(X)}=\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})Y_i}{V(X)}$$
$$\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X}$$

sont des fonctions linéaires des Y_i.

Preuve du théorème:

i. absence de biais.

Comme nous allons le voir cette propriété découle entièrement de la première condition de Gauss-Markov. On peut écrire:

$$\begin{aligned}\hat{\beta} &= \frac{cov(X,Y)}{V(X)}\\ &= \frac{cov(X,\alpha+\beta X+\epsilon)}{V(X)}\\ &= \frac{cov(X,\alpha)+cov(X,\beta X)+cov(X,\epsilon)}{V(X)}\\ &= 0+\beta+\frac{cov(X,\epsilon)}{V(X)}\\ &= \beta+\frac{cov(X,\epsilon)}{V(X)}\\ &= \beta+\frac{\sum_{i=1}^{n}(X_i-\bar{X})\epsilon_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\end{aligned}$$

La quatrième égalité résulte du fait que la covariance entre une variable aléatoire et une constante est nulle et du fait que la variance d'une variable n'est autre que la covariance entre cette variable et elle même.

Pour calculer l'espérance de β on peut procéder en deux temps: d'abord on considère l'espérance conditionnellement à X, puis on prend l'espérance de cette espérance conditionnelle (selon la loi des espérances itérées vue en début de cours). On obtient:

$$\begin{aligned}E(\hat{\beta}) &= E(E(\hat{\beta}|X))\\ &= E(E(\beta+\frac{\sum_{i=1}^{n}(X_i-\bar{X})\epsilon_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2}|X))\\ &= E(\beta+\frac{\sum_{i=1}^{n}(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}E(\epsilon_i|X))\\ &= E(\beta+0)\\ &= \beta\end{aligned}$$

où l'avant dernière égalité résulte du fait que l'espérance conditionnelle à X de toute fonction de X est égale à la fonction elle-même (ce qui explique que l'on peut « sortir » de l'espérance conditionnelle les termes qui dépendent de X) et la dernière égalité est une conséquence directe de la première condition de Gauss-Markov.

Pour $\hat{\alpha}$ la démonstration est plus simple:

$$\begin{aligned}E(\hat{\alpha}) &= E(E(\hat{\alpha}|X))\\ &= E(E(\bar{Y}-\hat{\beta}\bar{X}|X))\\ &= E(E(\alpha+\beta\bar{X}+\bar{\epsilon}-\hat{\beta}\bar{X}|X))\\ &= E(\alpha+\beta\bar{X}+E(\bar{\epsilon}|X)-\bar{X}E(\hat{\beta}|X))\\ &= E(\alpha)\\ &= \alpha\end{aligned}$$

ii. Efficacité.

Pour montrer l'efficacité, nous allons employer la loi de décomposition de la variance:

$$V(\hat{\beta})=V(E(\hat{\beta}|X))+E(V(\hat{\beta}|X))$$

Le premier terme du membre de droite de l'équation est nul, puisque $E(\hat{\beta}|X)=\beta$. Il reste donc le second terme à calculer. On sait que $\hat{\beta}=\beta+\frac{\sum_{i=1}^{n}(X_i-\bar{X})\epsilon_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$. La variance conditionnelle de $\hat{\beta}$ sachant X est donnée par:

$$\begin{aligned}V(\hat{\beta}|X)&=V(\beta+\frac{\sum_{i=1}^{n}(X_i-\bar{X})\epsilon_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2}|X)\\&=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2V(\epsilon_i|X)}{(\sum_{i=1}^{n}(X_i-\bar{X})^2)^2}\\&=\frac{\sigma^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\end{aligned}$$

La seconde égalité résulte de l'hypothèse d'absence d'autocorrélation et des propriétés de la variance. L'absence d'autocorrélation permet d'écrire que la variance de $\sum_{i=1}^{n}(X_i-\bar{X})\epsilon_i$ est égale à la somme des variances de $(X_i-\bar{X})\epsilon_i$, car les termes de covariance sont nuls. La dernière égalité résulte de l'hypothèse d'homoscédasticité.

Par conséquent la variance de $\hat{\beta}$ est donnée par: $V(\hat{\beta})=\sigma^2E((\sum_{i=1}^{n}(X_i-\bar{X})^2)^{-1})$ où l'espérance est prise par rapport à la loi de X.

De même $E(\hat{\alpha})=E(V(\hat{\alpha}|X))$. On a:

$$\begin{aligned}V(\hat{\alpha}|X)&=V(\bar{y}-\hat{\beta}\bar{X}|X)\\&=V(\alpha+\beta\bar{X}+\bar{\epsilon}-\hat{\beta}\bar{X}|X)\\&=V(\bar{\epsilon}-\hat{\beta}\bar{X}|X)\,\textit{étant donné que la variance de }\bar{X}\textit{ sachant X est nulle}\\&=V(\bar{\epsilon}|X)+\bar{X}^2V(\hat{\beta}|X)-2cov(\bar{\epsilon},\hat{\beta}\bar{X}|X)\\&=\sigma^2(\frac{1}{n}+\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2})\end{aligned}$$

La dernière égalité résulte du fait que $cov(\bar{\epsilon}, \hat{\beta}\bar{X} \mid X)=0$ (voir preuve ci-dessous). Par conséquent, la variance de l'estimateur de α est donnée par:

$$V(\hat{\alpha})=\sigma^2\left(\frac{1}{n}+E\left(\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)\right)$$

où, comme précédemment l'espérance est calculée suivant la loi de X.

$$\begin{aligned}cov(\bar{\epsilon}, \hat{\beta}\bar{X} \mid X) &= \bar{X}\, cov(\bar{\epsilon}, \hat{\beta} \mid X)\\ &= cov\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i, \frac{\sum_{i=1}^{n}(X_i-\bar{X})\epsilon_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2} \mid X\right)\\ &= \frac{1}{n\sum_{i=1}^{n}(X_i-\bar{X})^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\bar{X})cov(\epsilon_i, \epsilon_j \mid X)\end{aligned}$$

dans cette double somme seuls les termes de la forme $cov(\epsilon_i, \epsilon_i)$ *sont non nuls donc:*

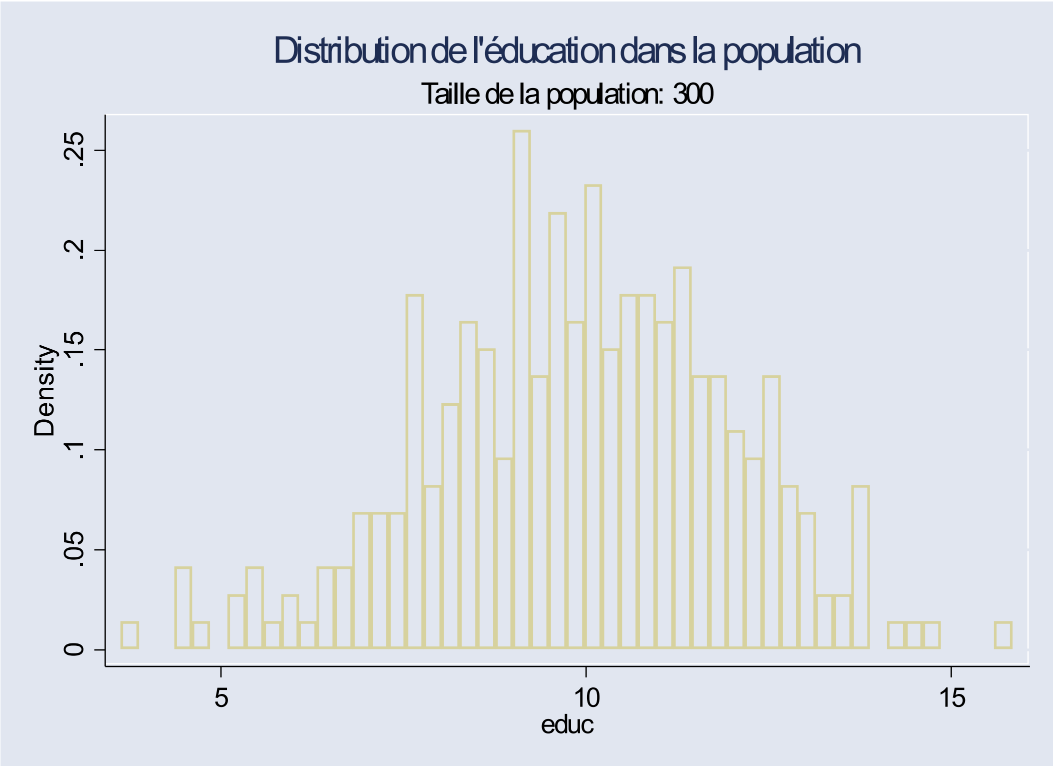
$$\begin{aligned}cov(\bar{\epsilon}, \hat{\beta}\bar{X} \mid X) &= \frac{1}{n\sum_{i=1}^{n}(X_i-\bar{X})^2}\sum_{i=1}^{n}(X_i-\bar{X})V(\epsilon_i \mid X)\\ &= \frac{\sigma^2}{n\sum_{i=1}^{n}(X_i-\bar{X})^2}\sum_{i=1}^{n}(X_i-\bar{X})\\ &= 0 \text{ puisque } \sum_{i=1}^{n}(X_i-\bar{X})=0.\end{aligned}$$

Ceci ne suffit cependant pas à démontrer l'efficacité des estimateurs des MCO. En effet, cette propriété signifie que la variance des estimateurs des MCO est la plus faible parmi l'ensemble des estimateurs linéaires des paramètres. C'est ce que nous allons maintenant montrer pour l'estimateur de β, la démonstration pour l'estimateur de la constante étant analogue. Nous avons vu que $\hat{\beta}$ est un estimateur linéaire, c'est à dire qu'il peut s'écrire sous la forme: $\hat{\beta}=\sum_{i=1}^{n}\frac{(X_i-\bar{X})}{nV(X)}Y_i=\sum_{i=1}^{n}c_i y_i$. Considérons maintenant un autre estimateur linéaire de $\beta: \tilde{\beta}=\sum_{i=1}^{n}d_i y_i$. En remplaçant y_i par son expression nous obtenons: $\tilde{\beta}=\alpha\sum_{i=1}^{n}d_i+\beta\sum_{i=1}^{n}d_i x_i+\sum_{i=1}^{n}d_i\epsilon_i$. Pour que cet estimateur soit sans biais il faut que les conditions suivantes soient respectées: $\sum_{i=1}^{n}d_i=0$ *et* $\sum_{i=1}^{n}d_i x_i=1$. De sorte que: $\tilde{\beta}=\beta+\sum_{i=1}^{n}d_i\epsilon_i$. La variance de $\tilde{\beta}$ *est égale à* $V(\tilde{\beta})=\sigma^2\sum_{i=1}^{n}d_i^2$. Soit maintenant $v_i=d_i-c_i$, *alors* $d_i=c_i+v_i$ *et* $V(\tilde{\beta})=\sigma^2\sum_{i=1}^{n}(c_i+v_i)^2=\sigma^2\sum_{i=1}^{n}(c_i^2+v_i^2+2c_i v_i)$ Pour achever la démonstration il suffit de montrer que $\sum_{i=1}^{n} c_i v_i = 0$:

$\sum_{i=1}^{n} c_i v_i = \sum_{i=1}^{n} (c_i d_i - c_i^2) = \frac{1}{V(X)} - \frac{1}{V(X)} = 0$ étant données les conditions que doivent respecter les termes d_i pour que l'estimateur $\tilde{\beta}$ soit sans biais. Par conséquent, il est clair que tout estimateur linéaire sans biais de β a une variance au moins égale à celle de l'estimateur des MCO, ce qui achève la démonstration du théorème de Gauss-Markov.

4. Illustration du caractère stochastique de l'estimateur des MCO.

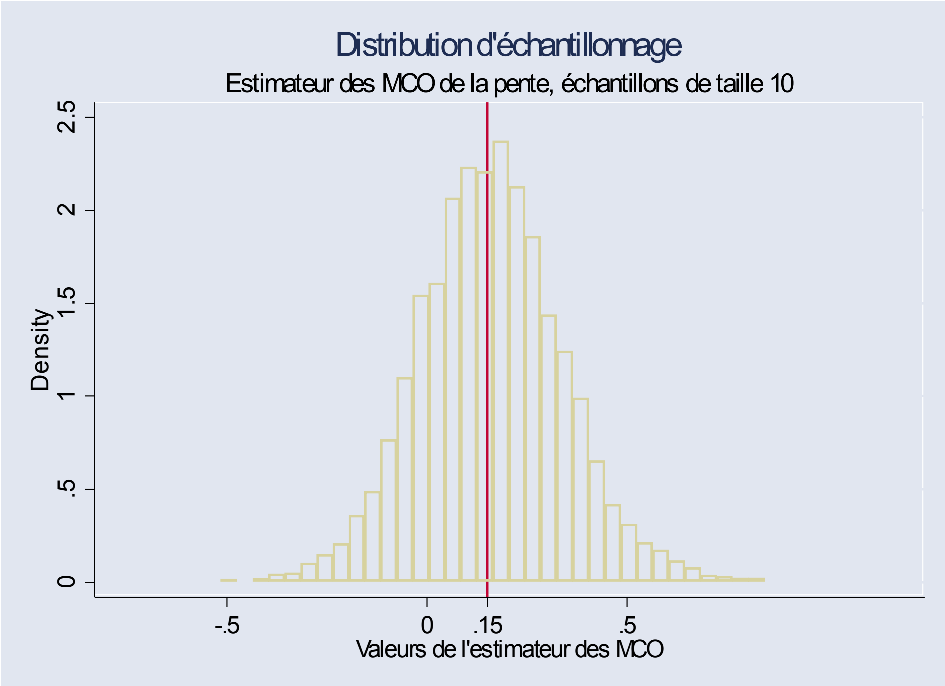
Pour illustrer la dimension aléatoire de l'estimateur des MCO, nous allons dans ce paragraphe nous livrer à un exercice de simulation de Monte Carlo. Considérons de nouveau notre population imaginaire de 300 étudiants. Ceux-ci ont des niveaux d'éducation variables dont la distribution est montrée ci-dessous:



Supposons que l'éducation soit reliée au revenu selon la relation suivante:

$$\ln(revenu_i) = 1 + 0{,}15 * education_i + \epsilon_i$$

où ϵ_i est distribuée selon une loi dite normale de moyenne nulle et de variance égale à l'unité. Supposons que l'on tire 10000 échantillons de taille 10 sur lesquels ont calcule la valeur de l'estimateur des MCO du coefficient de l'éducation. Cette expérience conduit à la distribution d'échantillonnage de l'estimateur de la pente, ici 0.15, qui est représentée ci-dessous:



Cette expérience illustre le caractère aléatoire de l'estimateur des MCO. Celui-ci peut prendre une valeur proche ou, au contraire, très différente de la « vraie » valeur, selon l'échantillon qui a été tiré. Comme il est impossible de connaître cette vraie valeur, il faut adopter la stratégie qui « marche » le mieux en moyenne: celle qui conduit à un estimateur sans biais et qui minimise la probabilité que la valeur estimée soit très éloignée de la vraie valeur.

Les paragraphes précédents ont montré qu'il est possible de calculer l'expression de la variance de l'estimateur des MCO. Celle-ci dépend de la variance du terme de perturbation σ^2. Si l'on peut obtenir une estimation de ce paramètre, on pourra alors estimer la variance de l'estimateur des Mco et évaluer le risque que la valeur estimée soit éloignée de la vraie valeur.

5. Estimation de la variance des estimateurs des MCO.

Souvenons nous que $\sigma^2=V(\epsilon_i \mid X_i)$. Etant donné que, par hypothèse, $E(\epsilon_i|X_i)=0$, on peut écrire: $\sigma^2=E(\epsilon_i^2|X_i)$. Puisque $\hat{\epsilon}_i=y_i-\hat{y}_i$ est un estimateur de ϵ_i il semble naturel d'utiliser $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^{\,2}$ pour estimer σ^2. En fait, même si l'idée est bonne, on peut montrer que cette expression ne conduit pas à une estimation non biaisée de σ^2. L'estimateur sans biais de σ^2 est donné par:

$$\hat{\sigma}^2=\frac{1}{n-2}\sum_{i=1}^{n}\hat{\epsilon}_i^{\,2}=\frac{SCR}{n-2}$$

La raison de ce résultat est la suivante: pour calculer les termes $\hat{\epsilon}_i$ il faut au préalable avoir estimé les coefficients $\alpha \; et \; \beta$. On dit que l'on « perd » deux « degrés de liberté » (un pour chaque coefficient estimé). Cette « perte » se traduit par le fait que pour avoir un estimateur sans biais de σ^2 il faut diviser la SCR par n-2 au lieu de n.

Nous pouvons maintenant donner une estimation de la variance des estimateurs des MCO:

$$\hat{V}(\hat{\beta})=\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$V(\hat{\alpha})=\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)$$

Notons que dans les expressions précédentes, les opérateurs « espérance », E, ont disparu, car pour estimer l'espérance des termes qui dépendent des observations de X, nous n'avons d'autre choix que de prendre la valeur de ces termes calculée avec les observations de X dont nous disposons.

6. Distribution des estimateurs des MCO.

Nous avons calculé la moyenne et la variance des estimateurs des MCO. Pour aller plus loin, nous devons maintenant faire des hypothèses supplémentaires sur la distribution du terme d'erreur. Nous allons supposer que, conditionnellement à la variable indépendante X, le terme d'erreur ϵ_i est distribué selon une loi dite normale, centrée et de moyenne inconnue égale à σ^2.

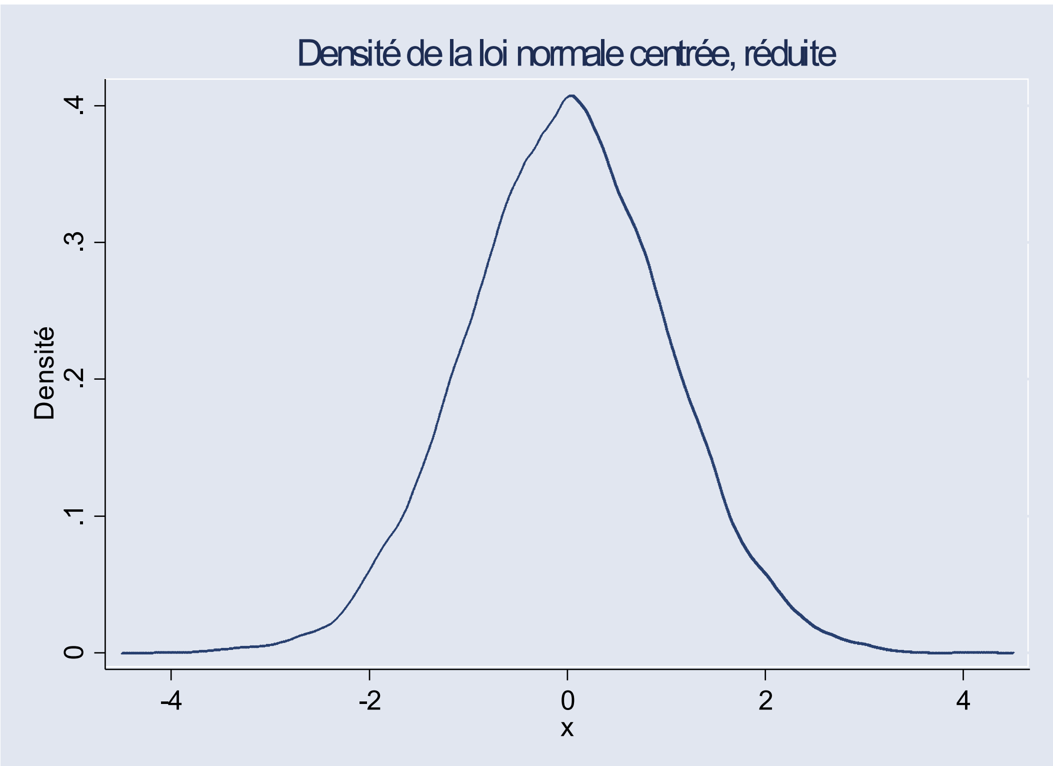
Définition 1: loi de distribution normale univariée.

Une variable aléatoire X est dite suivre une loi normale de moyenne μ et de variance σ^2 lorsque sa densité admet l'expression suivante:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

Il est habituelle de noter que X suit une loi normale de moyenne μ et de variance σ^2 sous la forme suivante: $X\rightarrow N(\mu,\sigma^2)$.

Le graphique ci-dessous montre la densité de la loi normale centrée (ie. de moyenne nulle) et réduite (ie. de variance égale à l'unité):



En théorie la densité est symétrique par rapport à sa moyenne (ici 0). Dans le cas présent les petites déformations que l'on peut observer sur la figure proviennent du fait que ce qui est représenté est la densité estimée d'un échantillon de 10000 observations tirées dans la loi normale centrée réduite.

On peut aussi définir la loi de distribution normale multivariée.

Définition 2: Loi normale multivariée.

Soit X un vecteur aléatoire de dimension n. On dit que X suit une loi normale de moyenne μ et de variance Σ lorsque sa densité a pour expression:

$$f(x)=f(x_1,\dots,x_n)=\frac{1}{(2\pi)^{n/2}det(\Sigma)^{1/2}}\exp(\frac{-1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))$$

Notez bien que dans cette expression la moyenne, μ, est un vecteur de dimension n et la variance, Σ, une matrice carrée de dimension (n,n). L'expression $det(\Sigma)$ note le déterminant de la matrice Σ.

Un vecteur aléatoire normal est également appelé vecteur aléatoire gaussien. Ces vecteurs ont une propriété remarquable:

Propriété: Combinaison linéaire d'éléments d'un vecteur aléatoire normale (ou gaussien).

Soit X un vecteur aléatoire normal de dimension n. Alors toute combinaison linéaire des éléments de X est également normale.

Ainsi si $X\rightarrow N(\mu,\Sigma),\ alors\ AX+b\rightarrow N(A\mu+b, A\Sigma A')$.

Nous pouvons maintenant déterminer la distribution des estimateurs des MCO. Sous l'hypothèse que, conditionnellement à X_i, $\epsilon_i\rightarrow N(0,\sigma^2)$ et sous les hypothèses du théorème de Gauss, Markov, la variable dépendante Y_i, conditionnellement à X_i, est distribuée selon la loi $N(\alpha+\beta X_i,\sigma^2)$. Le vecteur aléatoire $Y=(Y_1,\dots,Y_n)'$ est donc distribué selon la loi normale multivariée $N(\alpha+\beta X,\sigma^2 I_n)$ *où* I_n *est la matrice identité de dimension (n,n).*

La distribution des estimateurs des MCO s'en déduit en remarquant que ces estimateurs sont des fonctions linéaires des éléments de Y et sont donc distribués selon une loi normale, puisque Y est gaussien. Par conséquent:

$$\hat{\beta}\rightarrow N(\beta,\sigma^2.E(\sum_{i=1}^{n}(X_i-\bar{X})^2)^{-1})$$
$$\hat{\alpha}\rightarrow N(\alpha,\sigma^2(\frac{1}{n}+E(\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2})))$$

7. Tests d'hypothèses et intervalles de confiance.

La connaissance de la distribution des estimateurs des MCO est importante si l'on désire pouvoir évaluer le degré de confiance que l'on peut accorder aux estimations calculées sur un échantillon particulier. Supposons que l'on désire construire un intervalle, centré autour de valeur estimée du paramètre β, $\hat{\beta}$, et tel que la probabilité que cet intervalle contient la vraie valeur du paramètre est égale à 0,95. Soit $[a,b]$ cet interval. Il s'agit donc de trouver les valeurs de a et de b telles que $P(\beta \in [a,b])=0{,}95$. Ceci peut s'écrire, successivement:

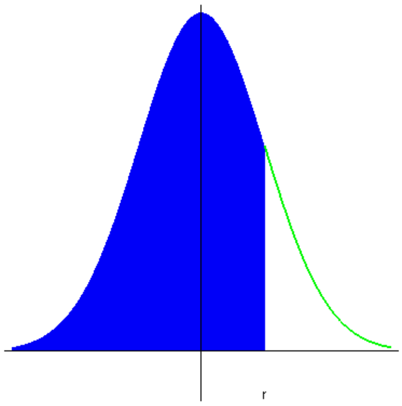
$$P(\beta \in [a,b])=0{,}95$$
$$P(a \leqslant \beta \leqslant b)=0{,}95$$
$$P(a-\hat{\beta} \leqslant \beta-\hat{\beta} \leqslant b-\hat{\beta})=0{,}95$$
où $\sigma_{\hat{\beta}}$ est l'écart-type de $\hat{\beta}$.
$$P\left(\frac{(a-\hat{\beta})}{\sigma_{\hat{\beta}}} \leqslant \frac{(\beta-\hat{\beta})}{\sigma_{\hat{\beta}}} \leqslant \frac{(b-\hat{\beta})}{\sigma_{\hat{\beta}}}\right)=0{,}95$$

Puisque $\hat{\beta} \rightarrow N(\beta, \sigma_{\hat{\beta}}^2)$ il est clair que $\frac{(\beta-\hat{\beta})}{\sigma_{\hat{\beta}}} \rightarrow N(0,1)$. Supposons donc que l'on cherche la valeur de λ telle que: $P(-\lambda \leqslant N(0,1) \leqslant \lambda)=0{,}95$. Autrement dit on cherche la limite de l'intervalle symétrique et centré autour de 0, tel qu'une variable aléatoire suivant la loi normale centrée et réduite ait une probabilité égale à 0,95 de prendre une valeur comprise dans cet intervalle.

La valeur de λ peut facilement être trouvée en utilisant une table telle que celle-ci:

Table de la loi normale (fonction de répartition)

$P(X \leqslant r)$ *avec* $X \to N(0,1)$.



	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

La table donne la valeur de $P(X \leqslant r)$ pour r compris entre 0 et 2,99. La table se lit de la façon suivante: dans la première colonne se trouve la valeur de r avec une précision de 1 décimale. Chacune des colonnes suivantes donne la valeur de $P(X \leqslant r)$ lorsque l'on ajoute une décimale supplémentaire à la précision de r (la première ligne du tableau indique la valeur de cette décimale). Par exemple, la deuxième ligne de la deuxième colonne donne la valeur de $P(X \leqslant 0{,}00) = 0{,}5$. La quatrième ligne de la sixième colonne donne la valeur de $P(X \leqslant 0{,}24) = 0{,}5948$. Comment utiliser cette table pour trouver la valeur de λ ? On cherche la valeur de λ telle que $P(-\lambda \leqslant N(0,1) \leqslant \lambda) = 0{,}95$. Comme la loi normale est symétrique, il suffit juste de trouver la valeur de λ telle que: $P(N(0,1)\leqslant\lambda)=0{,}975$. La table doit ici être utilisée à l'envers. Il faut rechercher la valeur 0,975 dans le tableau des probabilités et retrouver la valeur correspondante de r. On trouve: 1,96.

Nous avons donc établi que: $\frac{(a-\hat{\beta})}{\sigma_{\hat{\beta}}}=-\lambda$ *et que* $\frac{(b-\hat{\beta})}{\sigma_{\hat{\beta}}}=\lambda$. Ce qui conduit à:

$a=\hat{\beta}-\lambda\sigma_{\hat{\beta}}$ *et* $b=\hat{\beta}+\lambda\sigma_{\hat{\beta}}$. Etant donnée la valeur de λ , on obtient l'intervalle recherché: $P(\beta\in[\hat{\beta}-1{,}96\sigma_{\hat{\beta}},\hat{\beta}+1{,}96\sigma_{\hat{\beta}}])=0{,}95$. Cet intervalle est appelé intervalle de confiance à 95%. Cet intervalle n'est pas connu de l'économètre, mais peut facilement être estimé en remplaçant l'écart-type de $\hat{\beta}$ par son estimation.

L'estimation est donc parvenue à donner un intervalle qui a une forte probabilité de contenir la véritable valeur du paramètre recherché. Il faut bien noter qu'il n'est pas possible de trouver un intervalle fini qui a une probabilité certaine de contenir cette vraie valeur. Si l'on juge que la probabilité que l'intervalle de confiance ne contient pas la vraie valeur du paramètre, ici 5%, est trop grande, on peut élargir l'intervalle. Par exemple si 1% est une probabilité d'erreur jugée acceptable au lieu de 5%, la valeur de λ à retenir est 2,575.

Maintenant que nous savons construire des intervalles de confiance pour nos estimateurs, nous pouvons tester des hypothèses sur les coefficients. Supposons par exemple que l'on estime la relation: $consommation=\alpha+\beta*revenu+\epsilon$ On veut savoir si le revenu a un impact sur la consommation. Pour le savoir nous allons tester l'hypothèse que l'impact du revenu sur la consommation est nul, ce qui est noté:

$$H_0:\beta=0$$

Cette hypothèse est appelée hypothèse nulle. L'hypothèse dite alternative est notée:

$$H_1:\beta\neq0.$$

Lorsque l'on procède à un test de ce type on dit que l'on teste H_0 contre H_1. Supposons que l'on dispose de n observations sur le couple (revenu,consommation) et que l'on utilise ces observations pour estimer la valeur des paramètres α *et* β. Si l'hypothèse nulle est vérifiée, on sait qu'il y a une probabilité égale à 0,95 que 0 appartient à l'intervalle $[\hat{\beta}-1{,}96\sigma_{\hat{\beta}},\hat{\beta}+1{,}96\sigma_{\hat{\beta}}]$. Deux cas peuvent maintenant être envisagés:

- 0 appartient à l'intervalle de confiance ;
- 0 n'appartient pas à cet intervalle.

Supposons que l'on soit dans le second cas. On sait que si l'hypothèse nulle est vérifiée, il n'y a que 5% de chances pour que 0 n'appartienne pas à cet intervalle. On peut alors en conclure que, *probablement*, l'hypothèse nulle n'est pas vérifiée. Naturellement il est possible que l'on se trompe, mais on sait que la probabilité d'avoir tort, et de conclure que l'hypothèse nulle doit être rejetée, n'est que de 5%.

Cette probabilité est appelée niveau de significativité ou seuil du test. Plus exactement c'est la probabilité de ce que l'on appelle le risque de première espèce (ou erreur de type I) en théorie des tests. Selon cette théorie deux types de risques sont encourus lorsque l'on teste une hypothèse:

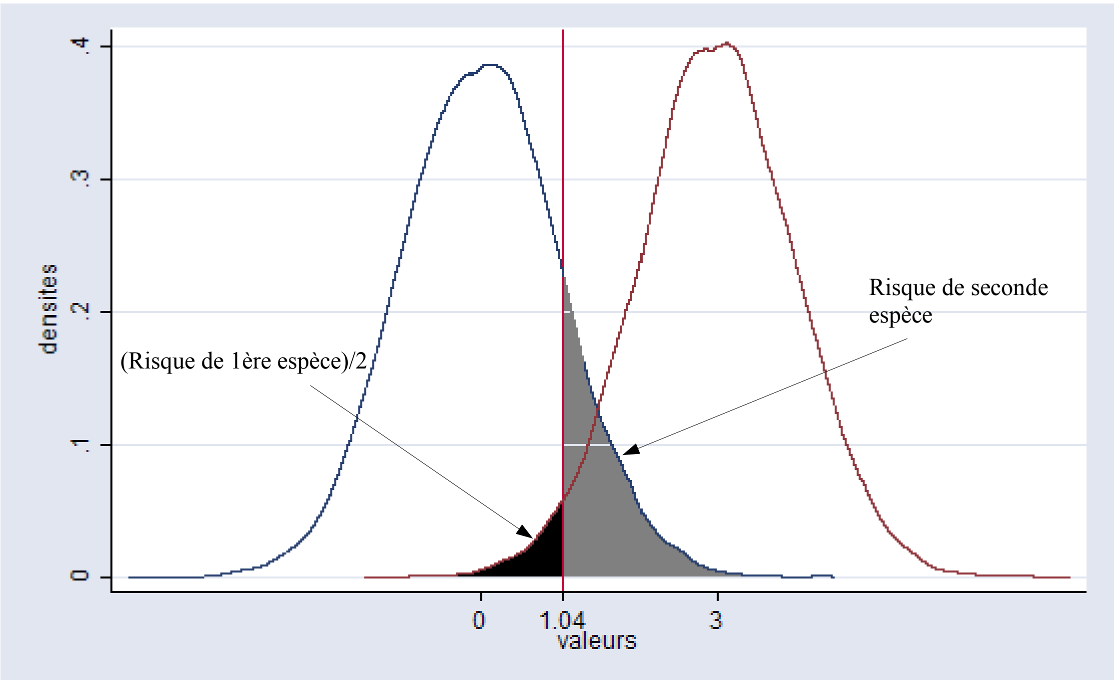
- Rejeter l'hypothèse nulle alors qu'elle est vraie est appelé le risque de première espèece (ou erreur de type I);

- Ne pas rejeter l'hypothèse nulle alors qu'elle est fausse est appelé le risque de seconde espère (ou erreur de type II).

Un aspect particulièrement important de la théorie des tests est qu'elle enseigne qu'il n'est pas possible de minimiser simultanément les deux types d'erreurs. En fait, plus on cherche à minimiser le risque de première espèce et plus le risque de seconde espèce augmente. Pour le voir supposons que l'on teste l'hypothèse: $H_0 : \beta = \beta_0$. Sous l'hypothèse nulle, c'est à dire si cette hypothèse est vérifiée, alors on sait qu'il y a une probabilité égale à $1-\gamma$ que β_0 appartient à l'intervalle $[\hat{\beta} - \lambda_\gamma \sigma_{\hat{\beta}}, \hat{\beta} + \lambda_\gamma \sigma_{\hat{\beta}}]$ où γ est le seuil du test (par exemple 0,05) et λ_γ est la valeur critique correspondant à ce seuil (par exemple 1,96). Il est alors facile de montrer que, toujours sous l'hypothèse nulle, l'estimateur de β, $\hat{\beta}$, a une probabilité égale à $1-\gamma$ d'appartenir à l'intervalle $[\beta_0 - \lambda_\gamma \sigma_{\hat{\beta}}, \beta_0 + \lambda_\gamma \sigma_{\hat{\beta}}]$:

$$P(\hat{\beta} - \lambda_\gamma \sigma_{\hat{\beta}} \leqslant \beta_0 \leqslant \hat{\beta} + \lambda_\gamma \sigma_{\hat{\beta}}) = 1-\gamma$$
$$P(\hat{\beta} - \beta_0 - \lambda_\gamma \sigma_{\hat{\beta}} \leqslant 0 \leqslant \hat{\beta} - \beta_0 + \lambda_\gamma \sigma_{\hat{\beta}}) = 1-\gamma$$
$$P(-\beta_0 - \lambda_\gamma \sigma_{\hat{\beta}} \leqslant -\hat{\beta} \leqslant -\beta_0 + \lambda_\gamma \sigma_{\hat{\beta}}) = 1-\gamma$$
$$P(\beta_0 - \lambda_\gamma \sigma_{\hat{\beta}} \leqslant \hat{\beta} \leqslant \beta_0 + \lambda_\gamma \sigma_{\hat{\beta}}) = 1-\gamma$$

Supposons maintenant que l'hypothèse nulle n'est pas vérifiée et que $\beta = \beta_1$ *inférieur à* β_0 . Les probabilités des risques de première et seconde espèces peuvent être lus dans le graphique ci-dessous:



Supposons que l'hypothèse nulle soit: $H_0 : \beta = 3$ et que l'on sache que la valeur de σ est égale à l'unité. Sous cette hypothèse la valeur du risque de première espèce est égale à deux fois l'aire située sous la courbe de densité à la droite du graphique et à gauche de 3 – 1,96 = 1,04. Pour calculer le risque de seconde espèce il faut connaître la valeur de β sous l'hypothèse nulle. J'ai ici supposé qu'elle était égale à 0. Le risque de seconde espèce est égal à la probabilité de ne pas rejeter l'hypothèse nulle alors que celle-ci est fausse. Dans le graphique ci-dessus elle est égale à l'aire située sous la courbe à la gauche du graphique et à droite de la valeur 1,04. En effet, on accepte l'hypothèse nulle lorsque la valeur estimée du coefficient est supérieure à 1,04. La probabilité que ceci arrive lorsque c'est l'hypothèse alternative qui est vérifiée est donnée par cette aire. On voit donc que plus on cherche à diminuer le risque de première espèce et plus le risque de seconde espèce augmente.

Puisqu'il n'est pas possible de contrôler les deux risques en même temps, il est habituel de fixer la valeur acceptable du risque de première espèce et de rechercher les tests d'hypothèse qui minimisent le risque de seconde espèce (on dit que l'on recherche les tests les plus puissants. La puissance d'un test étant égale à 1-P(risque de seconde espèce)). Les tests utilisés en économétrie répondent à cet contrainte.

8. Test de student.

Dans le paragraphe précédent nous avons obtenus nos résultats en supposant implicitement que la variance de l'estimateur des MCO était connue. Or ce n'est pas le cas et celle-ci ne peut être qu'estimée. Ceci complique un peu les chose, car s'il est vrai que le rapport $\frac{(\beta-\hat{\beta})}{\sigma_{\hat{\beta}}}$ est distribué selon la loi normale centrée réduite, cela n'est pas le cas du rapport $\frac{(\beta-\hat{\beta})}{\hat{\sigma}_{\hat{\beta}}}$. Il est facile de montrer, mais sort du cadre de ce cours, que ce rapport suit une loi dite de Student à n-2 degrés de liberté. La loi de Student est symétrique par rapport à 0. La table ci-dessous donne les valeurs prises par la fonction de répartition de la loi de Student en fonction du nombre de degrés de liberté:

La fonction de répartition de la loi de Student

$n \mid P$	0,60	0,70	0,80	0,90	0,95	0,975	0,990	0,995	0,999	0,9995
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	0,256	0,530	0,853	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	0,255	0,529	0,852	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	0,255	0,529	0,852	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	0,255	0,529	0,851	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,261	3,496
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
70	0,454	0,527	0,847	1,294	1,667	1,994	2,381	2,648	3,211	3,435
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
90	0,254	0,526	0,846	1,291	1,662	1,987	2,368	2,632	3,183	3,402
100	0,254	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Le nombre de degrès de liberté apparaît dans la première colonne. La première ligne donne les valeur des r, telles que $P(X \leqslant r)$. On lit la table de la façon suivante: pour un nombre de degrés de liberté égal à 1, la probabilité qu'une loi de student avec ce nombre de degrés de liberté prenne une valeur inférieure à 12,71 est égale à 0,975. Comme la loi de student est symétrique par rapport à 0, cela signifie qu'il y a une probabilité égale à 0,95 qu'une valeur tirée dans la loi de student à 1 degré de liberté appartient à l'intervalle [-12,71,+12,71]. L'intervalle équivalent, lorsque le nombre de degrés de liberté est égal à 100, est donné par [-1,984,+1,984]. Lorsque le nombre de degrés de liberté tend vers l'infini, l'intervalle dont la probabilité est égale à 0,95 est donné par [-1,96,+1,96]. Ce n'est pas un hasard si l'on retrouve la valeur critique de la loi normale: ceci résulte du fait que lorsque le nombre de degrés de libertés tend vers l'infini, la loi de student tend vers la loi normale. La procédure de test est donc exactement semblable à celle décrite dans le paragraphe précédent, à ceci près que, sauf si le nombre d'observations est très grand, les valeurs critiques doivent être lues dans la table de la loi de Student.

$\hat{y}_1$