TD Econométrie 1

Exercice 1. Rappels sur les estimateurs des Moindres Carrés. Propriétés à distance finie et propriétés asymptotiques

1.1. Estimateurs des MCO et propriétés à distance finie

Nous considérons ici l'inférence sur le modèle linéaire ($\mathbf{x}_i$ contient un terme constant) :

(1) $y_i = \mathbf{a}_0'\mathbf{x}_i + u_i$ et $E[u_i] = 0$

avec :

(2) $E[u_i/\mathbf{x}_i] = E[u_i] = 0 \quad \Rightarrow \quad E[\mathbf{x}_i u_i] = \mathbf{0}_{P\times 1}$

Dans la suite, les notations suivantes seront utilisées :

$$\mathbf{y} \equiv \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}, \quad \mathbf{X} \equiv \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix}_{N\times P} \quad \text{et} \quad \mathbf{u} \equiv \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}_{N\times 1}.$$

1.1.1. Commenter rapidement les hypothèses posées ici

1.1.2. Rappeler la forme des estimateurs des MCO de $\mathbf{a}_0$ ***et donner leurs principales propriétés à distance finie (avec*** N ***fixe)***

1.2. Propriétés asymptotiques de l'estimateur des MCO

1.2.1. Propriétés asymptotiques de l'estimateur des MCO à partir de l'exogénéité des $\mathbf{x}_i$.

Nous considérons dans un premier temps le modèle linéaire à variables explicatives exogènes.

(1) $y_i = \mathbf{a}_0'\mathbf{x}_i + u_i$ et $E[u_i] = 0$

avec :

(2) $E[u_i/\mathbf{x}_i] = E[u_i] = 0 \quad \Rightarrow \quad E[\mathbf{x}_i u_i] = \mathbf{0}_{P\times 1}$

et les (y_i ,$\mathbf{x}_i$) sont i.i.d..

1.2.1a. Montrer qu'il est possible de construire l'estimateur des MCO à partir de la contre-partie empirique de l'hypothèse d'exogénéité des $\mathbf{x}_i$: $E[\mathbf{x}_i u_i] = \mathbf{0}_{P\times 1}$.

1.2.1b. Montrer à l'aide des propriétés données dans l'annexe que $\hat{\mathbf{a}}_N^{MCO}$ ***est un estimateur convergent en probabilité de*** $\mathbf{a}_0$

1.2.1c. Montrer à l'aide des propriétés données dans l'annexe que $\hat{\mathbf{a}}_N^{MCO}$ ***est asymptotiquement normal***

Exercice 2 : Identification *versus* prédiction, le cas d'une fonction de production

On cherche ici à mesurer l'effet des engrais (de la fertilisation) sur les rendements agricoles de manière à voir si la fertilisation est importante au niveau de la production agricole et donc au niveau du revenu des agriculteurs. L'idée est ici que l'utilisation d'engrais par l'agriculture

est génératrice de pollutions (nitrates, …) ce qui fait qu'il est sérieusement envisagé de contraindre les agriculteurs à réduire leurs utilisations d'engrais.
Si on connaît l'impact de la fertilisation sur les rendements, il est alors possible de calculer l'impact de la fertilisation sur le revenu des agriculteurs. Ceci permet de calculer *in fine* les compensations à verser aux agriculteurs dans le cas où il leur serait, par exemple, interdit d'utiliser des engrais. Ceci permettrait, par là même, de calculer les coûts de l'interdiction des engrais (ou de tout autre réduction de l'utilisation des engrais) pour l'agriculture.

L'objectif est alors de mesurer l'effet des engrais sur les rendements, i.e. de mesurer comment les engrais « causent » les rendements. Nous supposerons que nous disposons de deux jeux de données.
Le premier est constitué de données expérimentales de l'INRA. L'INRA dispose de N sites d'expérimentations (i = 1,…,N) répartis en France. Les données correspondantes seront notées $(\tilde{y}_i, \tilde{x}_i)$, i = 1, …,N où $\tilde{x}_i$ représente la quantité d'engrais utilisée (en kilogrammes par hectare) et $\tilde{y}_i$ représente le rendement en blé (la culture de référence) obtenu (en quintaux par hectare) sur la parcelle i.
Le second est constitué de données issues de l'observation du comportement d'agriculteurs. Ce jeu de données est lié au premier dans la mesure où nous avons N agriculteurs dont les exploitations sont voisines des sites d'expérimentation de l'INRA. Aussi, les parcelles des agriculteurs sont comparables à celles de l'INRA : la qualité des sols y est identique et les parcelles des agriculteurs sont soumises aux mêmes conditions climatiques que celles de l'INRA. Les données correspondantes seront notées (y_i, x_i), i = 1, …,N. Les variables x_i et y_i ont la même signification que $\tilde{x}_i$ et $\tilde{y}_i$.
Si deux observations (y_i, x_i) et $(\tilde{y}_i, \tilde{x}_i)$ ont le même indice, alors elles sont issues de sites voisins. Elles sont donc comparables.
Le principal (en dehors de la présentation de quelques calculs « typiques ») objectif de cet exercice est de montrer que même si ces deux jeux de données sont *a priori* comparables, ils ne peuvent pas être analysés avec les mêmes méthodes d'inférence.

2.1. Utilisation des données expérimentales

Nous supposons pour simplifier que l'effet des engrais sur les rendements peut être représenté par une fonction de production agricole de la forme suivante (Cobb-Douglas) :

$$\ln \tilde{y}_i = \alpha + \beta \ln(1+\tilde{x}_i) + u_i \text{ et } E[u_i] = 0 \qquad (0 < \beta < 1)$$

et que les $(\tilde{y}_i, \tilde{x}_i)$ sont i.i.d. pour i = 1, …,N. Ce modèle est linéaire en $\ln \tilde{y}_i$ et $\ln(1+\tilde{x}_i)$. Nous supposons que le chercheur de l'INRA en charge de l'expérimentation a choisi les niveaux de quantité d'engrais de telle sorte que $\tilde{x}_i$ et u_i puissent être considérés comme indépendants l'un de l'autre.
Par souci de simplification, nous supposerons également que :

$$E\left[(u_i)^2 / \tilde{x}_i\right] = \sigma^2.$$

Remarque : Exemple tiré de Zellner, A., J. Kmenta and J. Drèze (1966). « Specification and estimation of Cobb-Douglas production models » Econometrica 34(4), pp784-795.
Remarque : Historiquement, l'estimation de fonctions de productions agricoles est un des premiers problèmes traité en micro-économétrie.

2.1.1. Quelle est la signification de l'hypothèse $(\tilde{y}_i, \tilde{x}_i)$ ***sont i.i.d. pour*** $i = 1, \ldots, N$ ***?***

2.1.2. Quelle est sont les significations « naturelles » des termes α ***et*** $\beta\tilde{x}_i$***, du terme d'erreur*** u_i ***?***

2.1.3. Quelle est la signification de « $\tilde{x}_i$ ***et*** u_i ***indépendants l'un de l'autre », quel est l'intérêt de cette indépendance d'un point de vue intuitif ?***

2.1.4. Montrer qu'on a : $E\left[\ln \tilde{y}_i / \tilde{x}_i\right] = \alpha + \beta \ln(1+\tilde{x}_i)$***, quelle est la signification de cette équation ?***

2.1.5. Quel est l'estimateur le plus approprié pour les paramètres α ***et*** β***, quels sont les principales propriétés de cet estimateur ?***

2.1.6. Représenter graphiquement la logique de la construction des estimateurs des MCO de α ***et*** β ***dans le plan avec*** $\ln(1+\tilde{x}_i)$ ***en abscisse et*** $\ln \tilde{y}_i$ ***en ordonnée.***

<u>2.2. Utilisation des données issues de l'observation du comportement des agriculteurs</u>

2.2.1. Montrer que le modèle correspondant aux données issues du comportement des agriculteurs s'écrit sous la forme : $\ln y_i = \alpha + \beta \ln(1+x_i) + u_i$ ***et*** $E\left[u_i\right] = 0$***.***

2.2.2. Il est très vraisemblable que x_i ***et*** u_i ***ne soient pas indépendants dans le sens où ils sont vraisemblablement positivement corrélés. Donner l'intuition de cette remarque.***

2.2.3. Confirmer cette intuition à l'aide du modèle micro-économique du comportement de l'agriculteur :

$$\underset{x_i, y_i}{Max}\left[p_i y_i - w_i x_i\right] \text{ s.c. } \ln y_i = \alpha + \beta \ln(1+x_i) + u_i$$

où : p_i ***est le prix du blé et*** w_i ***est le coût de d'utilisation d'un kilo d'engrais par hectare (prix de l'engrais plus prix de l'épandage). Il sera supposé que*** u_i ***et*** w_i / p_i ***sont indépendants.***

2.2.4. Montrer que l'estimateur des MCO n'est pas approprié pour l'échantillon des comportements observés des agriculteurs. Calculer le biais associé à l'utilisation de cet estimateur et l'interpréter.

2.2.5. Représentation graphique :
2.2.5a. Représenter graphiquement l'effet causal de la fertilisation sur les rendements dans le plan avec $\ln(1+x_i)$ ***en abscisse et*** $\ln y_i$ ***en ordonnée***
2.2.5b. Représenter graphiquement les u_i ***dans le plan avec*** $\ln(1+x_i)$ ***en abscisse et*** $\ln y_i$ ***en ordonnée***
2.2.5c. Représenter graphiquement les $\ln y_i$ ***dans le plan avec*** $\ln(1+x_i)$ ***en abscisse et*** $\ln y_i$ ***en ordonnée***

2.2.5d. Représenter graphiquement la droite de la régression de $\ln y_i$ ***sur le terme constant et*** $\ln(1+x_i)$ ***dans le plan avec*** $\ln(1+x_i)$ ***en abscisse et*** $\ln y_i$ ***en ordonnée***

<u>2.3. Utilisation correcte de l'échantillon des données issues de l'observation du comportement des agriculteurs</u>

2.3.1. Montrer qu'il existe au moins une variable instrumentale potentielle pour x_i ***dans celles qui ont été utilisées dans les développements précédents.***

2.3.2 Calculer un estimateur utilisant cette variable instrumentale et en donner les propriétés.
Notation :

$$\mathbf{Z}_i \equiv \begin{bmatrix} 1 \\ \ln(w_i / p_i) \end{bmatrix}.$$

2.3.3. De quelle autre approche dispose-t-on pour construire un estimateur de β ***et*** α ***?***