Toulouse School of Economics

UNIVERSITÉ TOULOUSE 1 CAPITOLE

Année universitaire 2017-2018

Session 1 - Semestre 6

Licence 3 mention Économie
Licence 3 mention Économie et MIASHS

ÉPREUVE : INTRODUCTORY ECONOMETRICS

Date de l'épreuve : **Mercredi 02 Mai 2018**

Durée de l'épreuve : 1h30

Liste des documents autorisés : aucun

Liste des matériels autorisés : Calculatrice FX-92, dictionnaire bilingue français/anglais

Nombre de pages : 4

Rappels: Le quantile d'une loi normale centrée et réduite à 95% (respectivement 97,5%) est égal à 1,63 (respectivement 1,96).

Problème I: On dispose d'un échantillon de 2324 hommes de 25 à 55 ans qui travaillent et reçoivent un taux de salaire horaire dont le logarithme est noté, y_i. On connait leur niveau d'éducation et leur âge mais de façon succincte.

- Si leur niveau d'éducation est de niveau d'un bac général ou d'un diplôme du supérieur, on note que $Bac_i = 1$. Sinon, pour les autres, on note $Bac_i = 0$.
- Si leur âge est supérieur à 40 ans, on note que $Senior_i = 1$. Sinon, pour les autres, on note $Senior_i = 0$.

On écrit le modèle de régression suivant en fonction de $x_i = (Bac_i, Senior_i)$:

$$y_i = \beta_0 + \beta_E Bac_i + \beta_A Senior_i + u_i, \tag{1}$$

où $E(u_i \mid x_i) = 0$.

Question 1: (0,5 pt) On admettra que l'échantillon est aléatoire. Qu'est-ce que cette hypothèse implique en termes de (y_i, x_i)?

Question 2: (0,5 pt) On admettra aussi l'hypothèse d'absence de multicolinéarité. Qu'est-ce que la multicolinéarité impliquerait en termes d'éducation et d'âge?

On obtient les résultats d'estimation suivants par Moindres Carrés Ordinaires:

$$\hat{\beta}_E = 0,258 \ \ (0,013), \hat{\beta}_A = 0,132 \ \ (0,012) \tag{2}$$

où entre parenthèses figurent les écarts-types estimés de cette estimation.

Question 3: (1,5 pts) On veut tester $H_0 : \beta_E = 0,20$ contre $H_a : \beta_E > 0,20$ à un niveau de 5%. Construire la statistique de test à partir des résultats précédents. Donner la loi de cette statistique sous H_0. Donner le résultat de la procédure de test. Quelle est la conclusion économique que vous en tirez?

Question 4: (1,5 pts) On veut tester $H_0 : \beta_E = \beta_A$ contre $H_a : \beta_E \neq \beta_A$. Expliciter une procédure de test en explicitant la formule de la statistique et la loi que celle-ci suit sous H_0. Que nous manque-t-il dans les résultats précédents pour calculer cette statistique de test?

Question 5: (1 pt) Expliciter les conditions sous lesquelles l'omission de l'interaction entre Bac_i et $Senior_i$ dans la régression, biaise l'estimation.

On agrége maintenant ces données en utilisant les quatre sous-groupes qui croisent les valeurs des variables Bac_i et $Senior_i$. Par exemple, un de ces sous-groupes rassemble les hommes n'ayant pas le baccalauréat et plus de 40 ans. Pour chacun de ces sous-groupes, on note pour les variables $z_i \in \{y_i, x_i, u_i\}$:

$$\overline{z}_g = \frac{1}{n_g} \sum_{i \in g} z_i$$

où n_g est la taille du groupe g.

Question 6: (0,5 pt) Quelles sont les valeurs de $\overline{Bac}_g$ et $\overline{Senior}_g$?

Question 7: (1,5 pts) Montrer que la régression de $\bar{y}_g$ sur $\overline{Bac}_g$ et $\overline{Senior}_g$ par Moindres Carrés Ordinaires donnent des estimateurs sans biais et convergents des paramètres β_0, β_E et β_A.

On obtient pour résultat de cette estimation, $\hat{\beta}_E^{(2)} = 0,247\ (0,037), \hat{\beta}_A^{(2)} = 0,120\ (0,037)$.

Question 8: (0,5 pt) Calculer $V(\bar{u}_g \mid \bar{x}_g)$ si on fait l'hypothèse d'homoscédasticité dans le modèle (1).

Question 9: (1 pt) Pourquoi $\hat{\beta}_E^{(2)} \neq \hat{\beta}_E$ où $\hat{\beta}_E$ est le résultat reporté en (2)?

Question 10: (1,5 pts) Expliquer comment construire l'estimateur des Moindres Carrés Généralisés (ou des Moindres carrés Pondérés) qui, dans la régression de $\bar{y}_g$ sur $\overline{Bac}_g$ et $\overline{Senior}_g$, donnent numériquement les mêmes résultats que ceux reportés en (2).

Problème II : Nous considérons le modèle linéaire simple satisfait par un jeu de n observations $(Y_j, X_j)_j = 1, ..., n$ composées d'une variable dépendante Y_j et d'une variable explicative $X_{j,1}$. Un terme constant est présent dans le modèle et correspond à la variable explicative $X_{j,0}$. Nous notons X_j le vecteur ligne $\begin{pmatrix} X_{j,0} & X_{j,1} \end{pmatrix}$, si bien que le modèle s'écrit :

$$Y_j = X_j\beta + \epsilon_j$$

où β est le vecteur colonne des paramètres et ϵ_j est le terme d'erreur. Nous supposons que $V(\epsilon_j \mid X_j) = \sigma^2$

Question 1: (0,5 pt) Rappeler les conditions sous lesquelles le vecteur de paramètres β peut être estimé et est interprétable dans le cadre des moindres carrés ordinaires. Nous supposerons ces conditions satisfaites dans la suite du problème.

Question 2: (0,5 pt) Donner l'espérance et la variance de l'estimateur des Moindres carrés ordinaires $\widehat{\beta}_{MCO}$. Est-ce un estimateur efficace ?

Question 3: (1 pt) Nous sommes intéressés par le paramètre de pente, c'est-à-dire le coefficient β_1 de la variable $X_{1,j}$. Appliquer le théorème de Frisch-Waugh pour dériver l'estimateur de β_1 dans ce cadre, il sera noté $\widehat{\beta}_{1,MCO}$. Cet estimateur est-il différent de l'estimateur usuel des Moindres carrés ordinaires du paramètre de pente sans appliquer le théorème de Frisch-Waugh ?

Question 4: (8 pts) On vous propose d'utiliser un estimateur alternatif de la pente défini comme suit:

$$\widehat{\beta}_{1,L}(\lambda) = \frac{\sum_{j=1}^{n}\left(X_{1,j} - \overline{X}_1\right)\left(Y_j - \overline{Y}\right) + \lambda \frac{\sum_{j=1}^{n}\left(X_{1,j} - \overline{X}_1\right)\left(Y_j - \overline{Y}\right)}{\sum_{j=1}^{n}\left(X_{1,j} - \overline{X}_1\right)^2}}{1 + \sum_{j=1}^{n}\left(X_{1,j} - \overline{X}_1\right)^2}$$

où $\lambda \in [0,1]$, $\overline{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$ et $\overline{X}_1 = \frac{1}{n}\sum_{j=1}^{n} X_{1,j}$. On posera

$$W(X,\lambda) = \frac{\lambda + \sum_{j=1}^{n}\left(X_{1,j} - \overline{X}_1\right)^2}{1 + \sum_{j=1}^{n}\left(X_{1,j} - \overline{X}_1\right)^2}$$

Question 4.a: (0,5 pt) Montrer que $\widehat{\beta}_{1,L}(\lambda)$ est un estimateur linéaire.

Question 4.b: (0,5 pt) Calculer son espérance. Montrer que cet estimateur est en général biaisé.

Question 4.c: (1 pt) Calculer sa variance.

Question 4.d: (0,5 pt) Comparer sa variance à celle de $\widehat{\beta}_{1,MCO}$ l'estimateur des Moindres carrés ordinaires du paramètre de pente. L'usage de l'estimateur $\widehat{\beta}_{1,L}(\lambda)$ est-il en général préférable ?

Question 4.e: (1 pt) Pour analyser les performances des différents estimateurs du paramètre de pente, nous proposons de comparer l'erreur quadratique moyenne de chaque estimateur définie par $E\left(\left(\widehat{\beta}_1 - \beta_1\right)^2 |X\right)$ où $\widehat{\beta}_1 = \widehat{\beta}_{1,MCO}$ ou $\widehat{\beta}_1 = \widehat{\beta}_{1,L}(\lambda)$. Montrer que cette erreur quadratique moyenne est telle que $E\left(\left(\widehat{\beta}_1 - \beta_1\right)^2 |X\right) = \left(E\left(\widehat{\beta}_1 |X\right) - \beta_1\right)^2 + V\left(\widehat{\beta}_1 |X\right)$. Interpréter cette équation.

Question 4.f: (0,5 pt) Calculer l'erreur quadratique moyenne de $\widehat{\beta}_{1,MCO}$ et $\widehat{\beta}_{1,L}(\lambda)$.

Question 4.g: (1,5 pts) Calculer la valeur de λ qui minimise l'erreur quadratique moyenne de $\widehat{\beta}_{1,L}(\lambda)$.

Question 4.h: (1 pt) Comparer la valeur minimale de l'erreur quadratique moyenne de $\widehat{\beta}_{1,L}(\lambda)$ à celle de $\widehat{\beta}_{1,MCO}$. Conclure.

Question 4.i: (1 pt) Lorsque la variable explicative X_1 varie très peu dans l'échantillon disponible, quel estimateur recommanderiez-vous ? Pourquoi ?

Question 4.j: (0,5 pt) Comment peut-on utiliser en pratique cette approche ?