Université Cergy-Pontoise

L3 Économétrie Approfondie Sc session 1

Examen. 7 mai 2014 2h00

Il vous est conseillé de lire une première fois l'ensemble du sujet avant de commencer.
Les documents ne sont pas autorisés.
L'ensemble des réponses devra être justifié.

Exercice 1 : 5 points

On considère le modèle :

$$\underset{(T,1)}{y} = \underset{(T,K+1)}{X}\underset{(K+1,1)}{\beta} + \underset{(T,1)}{\epsilon} \quad (1)$$

Ainsi que les hypothèses :

- $H_1 : E(\underset{(T,1)}{\epsilon}) = \underset{(T,1)}{0}$
- H_2 : La matrice $\underset{(T,K+1)}{X}$ est certaine.
- H_3 : la matrice $\underset{(T,K+1)}{X}$ est de plein rang colonne.
- $H_4 : E(\underset{(T,1)}{\epsilon} . \underset{(1,T)}{\epsilon'}) = \sigma^2 \underset{(T,T)}{I_T}$

1. Donner l'expression de l'estimateur $\hat{\beta}$ des MCO de β associé au modèle 1.
2. Calculer, l'espérance de $\hat{\beta}$, en précisant, parmi les hypothèses H_1 à H_4, celles que vous utilisez pour calculer l'espérance.
3. Calculer, la variance de $\hat{\beta}$, en précisant, parmi les hypothèses H_1 à H_4, celles que vous utilisez pour calculer la variance.

Exercice 2 : 3 points

On considère le modèle :

$$\underset{(T,1)}{y} = \underset{(T,K+1)}{X}\underset{(K+1,1)}{\beta} + \underset{(T,1)}{\epsilon} \quad (2)$$

Ainsi que les hypothèses :

- $H_1 : E(\underset{(T,1)}{\epsilon}) = \underset{(T,1)}{0}$
- H_2 : La matrice $\underset{(T,K+1)}{X}$ est certaine.
- H_3 : la matrice $\underset{(T,K+1)}{X}$ est de plein rang colonne.
- $H_4 : E(\underset{(T,1)}{\epsilon} . \underset{(1,T)}{\epsilon'}) = \sigma^2 \underset{(T,T)}{I_T}$

- $H_6 : \epsilon \rightsquigarrow \mathcal{N}(0, \sigma^2 I_T)$

Rappel : si la v.a $z \rightsquigarrow \mathcal{N}(m, \sigma^2)$ alors la densité de z est $f(z) = \frac{1}{\sqrt{2\Pi}\sigma} exp(\frac{-1}{2\sigma^2}(z-m)^2)$.

On va maintenant construire un estimateur de β par la méthode du maximum de vraisemblance.

1. Étant donnée la spécification 2, écrire la valeur de y_t, $\forall t$.
2. Donner la loi de y_t
3. Donner la densité de y_t
4. Écrire la log-vraisemblance de l'échantillon $(y_1, y_2, \dots, y_T)$
5. En déduire que $\hat{\beta}_{MV}$, l'estimateur de β par la méthode du maximum de vraisemblance est le même que $\hat{\beta}_{MCO}$, l'estimateur de β par la méthode des MCO.

Exercice 3 : 2 points

On considère le modèle :

$$\underset{(T,1)}{y} = \underset{(T,K+1)}{X} \underset{(K+1,1)}{\beta} + \underset{(T,1)}{\epsilon} \quad (3)$$

Ainsi que les hypothèses :

- $H_1 : E(\underset{(T,1)}{\epsilon}) = \underset{(T,1)}{0}$
- H_2 : La matrice $\underset{(T,K+1)}{X}$ est certaine.
- H_3 : la matrice $\underset{(T,K+1)}{X}$ est de plein rang colonne.
- $H_4 : E(\underset{(T,1)}{\epsilon} . \underset{(1,T)}{\epsilon'}) = \sigma^2 \underset{(T,T)}{I_T}$
- $H_6 : \epsilon \rightsquigarrow \mathcal{N}(0, \sigma^2 I_T)$

1. Donner la loi de $\hat{\beta}$
2. Donner un estimateur sans biais de σ^2 et sa loi.

Exercice 4 : 6 points

On veut estimer une fonction de production de type Cobb-Douglas

$$Y = A_0 K_t^{\alpha} L^{\beta} \quad (4)$$

- Y : le niveau de production
- K : le niveau de capital
- L : l'emploi

1-En transformant le modèle 4, faire apparaitre le modèle linéaire suivant :

$$log(Y_t) \quad = \quad \gamma + \alpha log(K_t) + \beta log(L_t) + \epsilon_t \tag{5}$$

2-Que représente la variable ϵ_t ?

On effectue la régression correspondant au modèle 5 sur un échantillon de $T = 360$ mois de 1970 à 1999. On obtient les résultats suivants :

$$\widehat{log(Y_t)} \quad = \quad \underset{(0.022)}{-0.112} + \underset{(0.017)}{0.317} log(K_t) + \underset{(0.034)}{0.545} log(L_t) \quad t = 1970:1, 1999:12 \tag{6}$$
$$R^2 = 0.901 \qquad SCR = 3.421$$

3-Expliquer pourquoi les valeurs entre parenthèses, sous les coefficients estimés, sont les écarts-types estimés des coefficients estimés.

4-Effectuer, en détail, le test de Fisher de significativité globale des coefficients.

On souhaite maintenant effectuer un test de stabilité des coefficients du modèle 5 sur la période 1970-1999. On dispose pour cela de deux régressions effectuées pour les sous périodes suivantes :

$$\widehat{log(Y_t)} \quad = \quad \underset{(0.012)}{-0.219} + \underset{(0.035)}{0.279} log(K_t) + \underset{(0.123)}{0.609} log(L_t) \quad t = 1970:1, 1985:12 \tag{7}$$
$$R^2 = 0.890 \qquad SCR = 2.102$$

$$\widehat{log(Y_t)} \quad = \quad \underset{(0.051)}{-0.139} + \underset{(0.100)}{0.320} log(K_t) + \underset{(0.210)}{0.560} log(L_t) \quad t = 1986:1, 1999:12 \tag{8}$$
$$R^2 = 0.921 \qquad SCR = 1.280$$

5-Présenter le test de stabilité des coefficients comme un test de contraintes linéaires sur un modèle particulier.

6-Donner la statistique du test et sa loi sous H_0.

7-Finir le test.

Exercice 5 : 4 points

Nous allons utiliser un extrait du Panel Study on Income and Dynamics (PSID) pour estimer une équation de salaire. On dispose des variables suivantes.

- `lwage` : le logarithme du salaire de i à la date t.
- `ed` : le nombre d'années d'éducation i.
- `fem` : le sexe i (1=femme, 0=homme).
- `black` : la couleur i (1=noir, 0=autre).
- `tdum`t : une indicatrice pour l'année t (1=1976,...,7=1982).

- `ms` : le statut marital de i à la date t (1=marié, 0=non marié).
- `union` : le mode de fixation du salaire de i à la date t (1=syndical, 0=autre).
- `south` : le lieu de résidence de i à la date t (1=sud USA, 0=autre).

Nous allons regardons l'influence du sexe sur le salaire :

$$lwage_{it} = \alpha + \beta FEM_i + \epsilon_{it} \tag{9}$$

1-Montrer que le coefficient β dans la modèle 9 représente la différence de salaire (en logarithme) moyen entre les femmes et les hommes.

2-Les résultats de la régression 9 sont donnés dans la figure 1. Tester au seuil de 5%, l'hypothèse selon laquelle il y a une différence de salaire entre les femmes et les hommes. **Il n'est pas nécessaire ici de construire l'ensemble du test mais il est nécessaire de donner les éléments de la figure 1 permettant de répondre à la question et de conclure.**

```
. reg lwage fem
```

Source	SS	df	MS
Model	93.6914807	1	93.6914807
Residual	793.213421	4163	.190538895
Total	886.904902	4164	.212993492

Number of obs	= 4165
F(1, 4163)	= 491.72
Prob > F	= 0.0000
R-squared	= 0.1056
Adj R-squared	= 0.1054
Root MSE	= .43651

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
fem	-.4744661	.0213967	-22.17	0.000	-.516415	-.4325171
_cons	6.729774	.00718	937.29	0.000	6.715697	6.74385

Figure 1: Source : PSID

3-On va maintenant (figure 2), introduire dans la régression des variables de contrôle. Comment interpréter maintenant le coefficient β associé à la variable `FEM` ?

4-Quel est le pourcentage d'augmentation du salaire lorsque les individus ont une année d'éducation supplémentaire ?

. reg lwage fem ed blk exp union ms

Source	SS	df	MS
Model	311.601708	6	51.933618
Residual	575.303194	4158	.138360557
Total	886.904902	4164	.212993492

Number of obs	=	4165
F(6, 4158)	=	375.35
Prob > F	=	0.0000
R-squared	=	0.3513
Adj R-squared	=	0.3504
Root MSE	=	.37197

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
fem	-.3452944	.0265123	-13.02	0.000	-.3972726	-.2933162
ed	.0776277	.0022099	35.13	0.000	.073295	.0819604
blk	-.1602561	.0230949	-6.94	0.000	-.2055345	-.1149777
exp	.011817	.000547	21.60	0.000	.0107446	.0128893
union	.0869909	.0125559	6.93	0.000	.0623747	.1116072
ms	.0541826	.0218042	2.48	0.013	.0114346	.0969305
_cons	5.419251	.0388458	139.51	0.000	5.343093	5.49541

Figure 2: Source : PSID

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

Table 1: Fonction de répartition de la loi normale centrée réduite

$$X \rightsquigarrow \mathcal{N}(0,1) \; F(x) = Prob(X \leq x) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

$n\backslash F$	**0,75**	**0,90**	**0,95**	**0,975**	**0,99**	**0,995**	**0,9995**
1	1,000	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	31,599
3	0,765	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	1,296	1,671	2,000	2,390	2,660	3,460
80	0,678	1,292	1,664	1,990	2,374	2,639	3,416
100	0,677	1,290	1,660	1,984	2,364	2,626	3,390
120	0,677	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,675	1,282	1,645	1,960	2,327	2,576	3,291

Table 2: Quantiles d'une loi de Student à n degrés de liberté. Valeurs de x telles que $Prob(T_n \leq x) = F$