Econométrie

Répétition 8

1. Matière : F-test, colinéarité, forme fonctionnelle et variables omises

- Régression polynomiale.
- Cas particuliers du F-test.
- t-tests et intervalles de confiance de combinaisons linéaires scalaires de paramètres.
- Colinéarité.
- Test de la forme fonctionnelle et variables omises.

2. Exercices

1- On considère les données reprises dans le fichier R7_manufact2.txt, où Y, L et K mesurent respectivement la valeur ajoutée (en milliers de FF), l'effectif (en nombre de travailleurs) et la valeur (en milliers de FF) du capital d'un échantillon de firmes de l'industrie manufacturière française.

a- Estimez les paramètres du modèle de fonction de production Cobb-Douglas :

$$\ln Y_i = \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + e_i$$

b- Testez au seuil de 5% la significativité de la régression dans son ensemble. Quelle est la P-valeur de ce test ? Répondez en utilisant directement les résultats d'estimation fournis par GRETL, et en effectuant un F-test calculé sur base de la SCR et du R^2 de la régression.

c- Testez au seuil de 5%, à l'aide d'un t-test, $\mathrm{H}_0 : \beta_3 = 0$ contre $\mathrm{H}_1 : \beta_3 \neq 0$. Quelle est la P-valeur de ce test ? Comparez votre résultat à celui d'un F-test de la même hypothèse nulle. Commentez.

d- Testez au seuil de 5%, à l'aide d'un t-test, $\mathrm{H}_0 : \beta_2 = 0,75$ contre $\mathrm{H}_1 : \beta_2 \neq 0,75$. Quelle est la P-valeur de ce test ? Comparez votre résultat à celui d'un F-test de la même hypothèse nulle. Commentez.

e- Testez au seuil de 5%, à l'aide d'un t-test, $\mathrm{H}_0 : \beta_2 + \beta_3 = 1$ contre $\mathrm{H}_1 :$

$\beta_2 + \beta_3 \neq 1$. Quelle est la P-valeur de ce test ? Comparez votre résultat à celui d'un F-test de la même hypothèse nulle. Commentez.

f- Testez au seuil de 5% $\mathrm{H}_0: \beta_2 + \beta_3 \leq 1$ contre $\mathrm{H}_1: \beta_2 + \beta_3 > 1$. Quelle est la P-valeur de ce test ?

g- Testez au seuil de 5% $\mathrm{H}_0: \beta_2 + \beta_3 \geq 1$ contre $\mathrm{H}_1: \beta_2 + \beta_3 < 1$. Quelle est la P-valeur de ce test ?

h- Testez au seuil de 5%, à l'aide d'un t-test, $\mathrm{H}_0: \beta_2 = 3\beta_3$ contre $\mathrm{H}_1: \beta_2 \neq 3\beta_3$. Quelle est la P-valeur de ce test ? Comparez votre résultat à celui d'un F-test de la même hypothèse nulle. Commentez.

i- Testez au seuil de 5% $\mathrm{H}_0: \beta_2 \geq 3\beta_3$ contre $\mathrm{H}_1: \beta_2 < 3\beta_3$. Quelle est la P-valeur de ce test ?

j- Testez au seuil de 5% $\mathrm{H}_0: \beta_2 \leq 3\beta_3$ contre $\mathrm{H}_1: \beta_2 > 3\beta_3$. Quelle est la P-valeur de ce test ?

k- Donnez un intervalle de confiance à 99% pour les rendements d'échelle $(\beta_2 + \beta_3)$.

2- On considère le modèle :

$$Sal_i = \beta_1 + \beta_2 Educ_i + \beta_3 Expe_i + e_i \tag{1}$$

où : Sal_i = le salaire horaire de l'individu i
$Educ_i$ = le nombre d'années d'étude de l'individu i
$Expe_i$ = le nombre d'années d'expérience professionnelle de l'individu i

pour lequel un ensemble d'observations américaines (la variable Sal est le salaire horaire mesuré en $) est contenu dans le fichier R8_salaire.txt.

a- Quelle est l'interprétation des paramètres β_1, β_2 et β_3 du modèle ?

b- Estimez le modèle par MCO et interprétez les résultats obtenus à la lumière de votre réponse au point a.

c- Réestimez le modèle en omettant la variable $Expe_i$, et comparez la valeur estimée de β_2 à celle obtenue au point b. Expliquez pourquoi on pouvait s'attendre à ce que, lorsque la variable $Expe_i$ est omise, la valeur estimée de β_2 soit plus faible (conseil : examinez la corrélation qui existe entre $Educ_i$ et $Expe_i$).

d- Testez la forme fonctionnelle du modèle (1) en considérant comme modèle étendu un modèle incluant les carrés et le produit croisé des variables $Educ_i$ et $Expe_i$. Effectuez ce test :

i- en estimant le modèle étendu, puis en réalisant un F-test de la nullité des paramètres des variables ajoutées $Educ_i^2$, $Expe_i^2$ et $(Educ_i \times Expe_i)$ à l'aide la fonction 'Tests - Omit Variables - Wald test' de GRETL.

ii- sur base de l'estimation du modèle (1), en utilisant la fonction 'Tests - Add Variables' de GRETL.

e- Testez la forme fonctionnelle du modèle (1) en considérant comme modèle étendu un modèle incluant les puissances $\hat{y}_i^2$ et $\hat{y}_i^3$ de la valeur prédite $\hat{y}_i = X_i\hat{\beta}$ du modèle (test RESET). Effectuez ce test :

i- en estimant le modèle étendu, puis en réalisant un F-test de la nullité des paramètres des variables ajoutées $\hat{y}_i^2$ et $\hat{y}_i^3$ à l'aide la fonction 'Tests - Omit Variables - Wald test' de GRETL.

ii- sur base du l'estimation du modèle (1), en utilisant la fonction 'Tests - Add Variables' de GRETL.

iii- sur base du l'estimation du modèle (1), en utilisant la fonction 'Tests - Ramsey's RESET - Squares and Cubes' de GRETL.

3- On considère maintenant le modèle :

$$\begin{aligned} Sal_i &= \beta_1 + \beta_2(Educ_i - 12) + \beta_3(Expe_i - 15) + \beta_4(Educ_i - 12)^2 \\ &\quad + \beta_5(Expe_i - 15)^2 + \beta_6((Educ_i - 12)(Expe_i - 15)) + e_i \end{aligned}$$

où : Sal_i = le salaire horaire de l'individu i
$Educ_i$ = le nombre d'années d'étude de l'individu i
$Expe_i$ = le nombre d'années d'expérience professionnelle de l'individu i

pour le même ensemble d'observations américaines (la variable Sal est le salaire horaire mesuré en \$) contenu dans le fichier R8_salaire.txt.

a- Interprétez les paramètres du modèle en vous référant aux effets marginaux théoriques des variables $Educ$ et $Expe$ sur la variable Sal. Comparez en particulier l'interprétation des paramètres β_2 et β_3 à votre réponse au point a. de l'exercice 2 ci-dessus.

b- Estimez le modèle par MCO et interprétez les résultats obtenus à la lumière de votre réponse au point a.

c- Testez la significativité de la régression dans son ensemble.

d- Testez H_0 : $\beta_2 = \beta_4 = \beta_6 = 0$ contre H_1 : $\beta_2 \neq 0$ et/ou $\beta_4 \neq 0$ et/ou $\beta_6 \neq 0$. Quelle est la signification de ce test ? Quelle est sa P-valeur ? Qu'en concluez-vous ?

e- Testez H_0 : $\beta_3 = \beta_5 = \beta_6 = 0$ contre H_1 : $\beta_3 \neq 0$ et/ou $\beta_5 \neq 0$ et/ou $\beta_6 \neq 0$. Quelle est la signification de ce test ? Quelle est sa P-valeur ? Qu'en concluez-vous ?

f- Testez H_0 : $\beta_4 = \beta_5 = \beta_6 = 0$ contre H_1 : $\beta_4 \neq 0$ et/ou $\beta_5 \neq 0$ et/ou $\beta_6 \neq 0$. Quelle est la signification de ce test ? Quelle est sa P-valeur ? Qu'en concluez-vous ? Comparez votre réponse à celle obtenue au point d. de l'exercice 2 ci-dessus.

g- On considère un individu ayant accomplis 14 années d'étude et ayant 10 années d'expérience professionnelle.

i- Donnez un intervalle de prévision à 95% pour l'espérance du salaire de cet individu.

ii- Donnez un intervalle de prévision à 95% pour le salaire de cet individu. Sur quelle hypothèse implicite la validité de votre calcul repose-t-elle ?

4- On considère le modèle :

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + e_i$$

pour lequel un échantillon d'observations, générées par simulation pour un ensemble choisi de valeurs (x_{i2}, x_{i3}, x_{i4}), des valeurs de paramètres $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$ et $\sigma^2 = 0,2$, et $e_i \sim N(0, \sigma^2)$, est donné dans le fichier R8_colli.txt.

a- Examinez les corrélations entre les régresseurs du modèle.
b- Estimez par MCO les paramètres du modèle.
c- Testez au seuil de 2,5% $\mathrm{H}_0: \ \beta_2 = 0$ contre $\mathrm{H}_1: \ \beta_2 \neq 0$. Quelle est la P-valeur de ce test ?
d- Testez au seuil de 2,5% $\mathrm{H}_0: \ \beta_3 = 0$ contre $\mathrm{H}_1: \ \beta_3 \neq 0$. Quelle est la P-valeur de ce test ?
e- Testez au seuil de 5% $\mathrm{H}_0: \ \beta_2 = \beta_3 = 0$ contre $\mathrm{H}_1: \ \beta_2 \neq 0$ et/ou $\beta_3 \neq 0$. Quelle est la P-valeur de ce test ?
f- Calculez un intervalle de confiance à 95% pour β_2 et pour β_3.
g- Calculez un intervalle de confiance à 95% pour $(\beta_2 + \beta_3)$.
h- Commentez, à la lumière de ce que vous savez des 'vraies valeurs' des paramètres et de votre réponse au point a., les résultats obtenus aux points c. à g.

5- On considère les données du fichier R8_production.xlsx, où Q, L et K mesurent respectivement la valeur ajoutée (en millions de \$), l'effectif (en milliers de travailleurs) et la valeur (en millions de \$)) du capital d'un échantillon d'entreprises sidérurgiques américaines.

a- Estimez par MCO le modèle de fonction de production Cobb-Douglas :

$$\ln Q_i = \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + e_i$$

b- Calculez un intervalle de confiance à 95% pour β_2 et pour β_3.
c- Testez $\mathrm{H}_0: \ \beta_2 = 0$ contre $\mathrm{H}_1: \ \beta_2 \neq 0$. Quelle est la P-valeur de ce test ?
d- Testez $\mathrm{H}_0: \ \beta_3 = 0$ contre $\mathrm{H}_1: \ \beta_3 \neq 0$. Quelle est la P-valeur de ce test ?
e- Testez $\mathrm{H}_0: \ \beta_2 = \beta_3 = 0$ contre $\mathrm{H}_1: \ \beta_2 \neq 0$ et/ou $\beta_3 \neq 0$. Quelle est la P-valeur de ce test ?
f- Comment expliquez-vous la forte différence entre les résultats obtenus d'une part aux points c. et d., et d'autre part au point e. ?
g- Calculez un intervalle de prévision à 95% pour la valeur ajoutée d'une firme dont l'effectif est de 14 et la valeur du capital de 21.

6- On considère le modèle :

$$\begin{aligned} \ln y_i \ = \ & \beta_1 + \beta_2 (\ln l_i - \ln l^*) + \beta_3 (\ln k_i - \ln k^*) + \beta_4 (\ln l_i - \ln l^*)^2 \\ & + \beta_5 (\ln k_i - \ln k^*)^2 + \beta_6((\ln l_i - \ln l^*) (\ln k_i - \ln k^*)) + e_i \end{aligned}$$

où : $y_i =$ la production de la firme i
$l_i =$ le nombre de travailleurs occupés par la firme i
$k_i =$ le stock de capital de la firme i
l^* et $k^* =$ les valeurs moyennes dans l'échantillon de respectivement l_i et k_i

L'estimation de ce modèle par MCO a donné les résultats suivants :

Model 1:
OLS, using observations 1-150
Dependent variable: LnY

	coefficient	std. error	t-ratio	p-value
const	11.777	0.0605	194.687	2.77e-176 ***
$(LnL\text{-}LnL^*)$	0.713	0.1008	7.078	5.92e-011 ***
$(LnK\text{-}LnK^*)$	0.307	0.0849	3.625	0.0004 ***
$(LnL\text{-}LnL^*)^2$	0.094	0.0425	2.216	0.0283 **
$(LnK\text{-}LnK^*)^2$	0.076	0.0302	2.541	0.0121 **
$(LnL\text{-}LnL^*)(LnK\text{-}LnK^*)$	-0.142	0.0649	-2.190	0.0301 **

a- Quelle est la signification précise des paramètres β_2 et β_3 ?
b- Calculez un intervalle de confiance à 95% pour β_2 et pour β_3.
c- Sachant que $C\hat{o}v(\hat{\beta}_2, \hat{\beta}_3) = -0,0068$, testez au seuil de 5% $H_0 : \beta_2 + \beta_3 \leq 1$ contre $H_1 : \beta_2 + \beta_3 > 1$. Quelle est l'interprétation économique de ce test ?
d- Donnez un intervalle de confiance à 95% pour $(\beta_2 + \beta_3)$. Quelle est l'interprétation économique de ce intervalle ?