

ANNEE UNIVERSITAIRE 2018-2019

SESSION 1

Examen « Econométrie »

Durée 3h

Indiquez :

NOM : TIET

PRENOM : Christophe

Consignes:

Vous utilisez ce document pour donner vos réponses. Vous l'enregistrez et vous le déposez sous la forme d'un **Document PDF**. **Vous présentez tous les programmes (SAS ou R) ainsi que les tableaux des résultats des estimations en annexe.**

Dans le fichier des données « **data_exam** » on dispose des variables suivantes (4165 observations et 12 variables).

Fichier : **data_exam**

Les variables sont les suivantes :

- LWAGE=le logarithme népérien du salaire
- ED =nombre d'années d'éducation
- EXP =nombre d'années d'expérience
- WKS =nombre de semaines travaillées
- UNION =1 si le salarié est syndiqué
- IND =1 si l'individu travaille dans l'industrie manufacturière
- SMSA =1 si l'individu réside dans une ville
- OCC =1 si le salarié est ouvrier
- MS =1 si le salarié est marié
- SOUTH =1 si le salarié habite dans le sud
- FEM =1 si le salarié est une femme
- BLK =1 si le salarié est noir

1) On considère l'estimation de la fonction de gains suivante :

$$LWAGE_i = \alpha_1 + \beta_1 ED_i + \beta_2 EXP_i + \beta_3 UNION_i + \beta_4 MS_i + \varepsilon_{1i} \quad (1)$$

- a) Estimez le modèle (1) par la méthode des MCO.
- b) Interprétez la valeur estimée du terme constant α_1 . (**4 lignes au max**).

Réponse : α_1 est le terme constant de notre modèle. Celui-ci vaut 5.16. Cela signifie que pour toutes variables endogènes valant 0, LWAGE est égale à 5.16. Une personne non mariée et non syndiquée n'ayant aucune éducation et aucune expérience touchera quand même un salaire de $e(5.16)=174,16$.

2) A partir du fichier des données créez les deux variables suivantes : $UNIONC_i = 1 - UNION_i$ et $MSC_i = 1 - MS_i$. On considère le modèle (2) suivant :

$$LWAGE_i = \alpha_2 + \delta_1 ED_i + \delta_2 EXP_i + \delta_3 UNIONC_i + \delta_4 MSC_i + \varepsilon_{2i} \quad (2)$$

- a) Estimez le modèle (2) par la méthode des MCO.
- b) Interprétez la valeur estimée du terme constant α_2 . (**4 lignes au max**).

Réponse : α_2 est le terme constant de notre modèle. Celui-ci vaut 5.54. Cela signifie que pour toutes variables endogènes valant 0, LWAGE est égale à 5.54. Une personne mariée et syndiquée n'ayant aucune éducation et aucune expérience touchera quand même un salaire de $e(5.54)=254,67$.

- c) Comparez les estimations des deux coefficients α_1 et α_2 . Commentez (**8 lignes au max**).

Réponse : On remarque que le coefficient α_1 est supérieur à celui de α_2 . Cela équivaut à dire qu'une personne mariée et syndiquée touchera 80.51(254.67-174.16) unités de salaire en plus qu'une non mariée et non syndiquée si ces deux personnes ont une même éducation et une même expérience nulles.

- d) Exprimez α_1 en fonction des paramètres du modèle (2).

Réponse : En comparant nos paramètres sur les deux modèles, on se rend compte que

$\beta_1=\delta_1$, $\beta_2=\delta_2$, $\beta_3=-\delta_3$ et $\beta_4=-\delta_4$. On sait que $UNIONC=1-UNION$ et $MSC=1-MS$.

Donc $\alpha_1 + \beta_1 ED_i + \beta_2 EXP_i + \beta_3 (1 - UNIONC_i) + \beta_4 (1 - MSC_i) = \alpha_2 + \delta_1 ED_i + \delta_2 EXP_i + \delta_3 UNIONC_i + \delta_4 MSC_i$

On peut donc en déduire que $\alpha_1 = \alpha_2 + \delta_3 + \delta_4$.

- e) Comparez les estimations des deux coefficients β_1 et β_2 avec δ_1 et δ_2 respectivement. Commentez. Que peut-on conclure ? (**10 lignes au max**)

Réponse : En comparant nos deux modèles, on se rend compte que $\beta_1 = \delta_1$ et que $\beta_2 = \delta_2$. On peut donc en déduire que la situation maritale et le fait qu'une personne soit syndiquée ou non n'a pas de corrélation avec l'éducation et l'expérience.

Peu importe la situation maritale ou syndicale, Le rendement d'une année d'étude sur le salaire sera de 7.95% et le rendement d'une année d'expérience sur le salaire sera de 1.14%.

On notera que le fait d'être marié et syndiqué permet d'avoir un salaire légèrement plus élevée.

3) Supposons maintenant que :

$$LWAGE_i = \alpha + \lambda_1 ED_i + \lambda_2 WKS_i + \lambda_3 EXP_i + \lambda_4 EXP_i^2 + \lambda_5 UNION_i + \lambda_6 MS_i + \lambda_7 SMSA_i + \lambda_8 OCC_i + \varepsilon_i \quad (3)$$

a) Estimez cette fonction de gains par la méthode des MCO. Commentez les résultats de cette estimation. **(12 lignes au max)**.

Réponse : On remarque tout d'abord que notre modèle est très significatif avec une p-value de 2.2×10^{-16} et que toutes nos variables sont significatives au seuil de 1%. Le R^2 ajusté signifie que 37.82% de la dispersion de LWAGE est expliquée par le modèle.

Les coefficients semblent cohérents. Il semble logique que le salaire augmente si le niveau d'éducation et d'expérience est plus élevé et si le temps de travail est plus long. Il semble également normal qu'une personne habitant en ville gagne plus qu'une personne habitant en milieu rural.

On observe aussi deux coefficients négatifs, OCC et EXP2, on traitera le cas d'EXP2 dans la prochaine question. Concernant la variable OCC, un travail d'ouvrier aura tendance à demander des capacités plus restreintes qu'un travail de non ouvrier, d'où un salaire moins élevé.

b) Quelle est l'interprétation du signe négatif de λ_4 ? **(4 lignes au max)**.

Réponse : Le signe négatif de λ_4 signifie qu'à un certain point, l'augmentation de l'expérience n'a plus un effet positif sur le salaire mais négative. Cela peut s'expliquer par l'âge de l'individu qui vieillit au fur et à mesure que son expérience grandit.

c) Testez l'hypothèse : $\lambda_7 + \lambda_8 = 0$. Commentez. **(4 lignes au max)**.

Réponse : On obtient une p-value de 0.68, on ne peut donc pas rejeter l'hypothèse nulle : $\lambda_7 + \lambda_8 = 0$.

4) Dans l'estimation des différents paramètres du modèle (3) par la méthode des MCO, nous avons supposé que toutes les variables sont exogènes.

$$LWAGE_i = \alpha + \lambda_1 ED_i + \lambda_2 WKS_i + \lambda_3 EXP_i + \lambda_4 EXP_i^2 + \lambda_5 UNION_i$$

$$+\lambda_6 MS_i + \lambda_7 SMSA_i + \lambda_8 OCC_i + \varepsilon_i \quad (4)$$

Traitez la notion d'exogénéité de la variable ED_i dans le modèle (4).

Nous devons tout d'abord commencer par vérifier que nos trois variables instrumentales SOUTH, FEM et BLK sont corrélées avec ED. Pour cela, nous allons faire une régression OLS sur la forme réduite de ED à savoir sur l'équation :

$ED \sim WKS + EXP + EXP2 + UNION + MS + SMSA + OCC + SOUTH + FEM + BLK$.

Dans notre cas, la pvalue est très faible ce qui signifie que les variables instrumentales SOUTH, FEM et BLK sont bien corrélées avec ED.

Une fois cela vérifiée, nous pouvons tester l'exogénéité de la variable ED en procédant en deux étapes :

1) On calcule le résidu de la régression :

$ED \sim WKS + EXP + EXP2 + UNION + MS + SMSA + OCC + SOUTH + FEM + BLK$.

2) On estime l'équation structurelle par OLS en ajoutant ce résidu comme variable explicative.

On remarque ainsi que la variable ED est endogène étant donné que ce résidu est très significatif dans notre modèle.

Nous pouvons donc utiliser la méthode des DMC et faire un test de restrictions afin de tester l'exogénéité de nos VI.

Nous sommes, ici, en conditions de suridentifiants : nous disposons de trois variables instrumentales pour une variable endogène, autrement dit le nombre de restrictions suridentifiantes est égal à 2. Nous pouvons donc appliquer le test de Sargan.

Nous nous retrouvons finalement avec une pvalue de 3.55×10^{-14} . Nous rejettons donc H_0 : les variables instrumentales sont exogènes.

En conclusion, nous avons rejeté l'hypothèse selon laquelle ED était exogène, cependant nos variables instrumentales ne sont pas exogènes. Il nous semble donc que ED est bien une variable endogène mais qu'il faudrait reconsidérer le choix des variables instrumentales.

ANNEXE : Insérez vos tableaux des résultats ci-dessous.

Question 1

```
call:
lm(formula = LWAGE ~ ED + EXP + UNION + MS, data = don.sas)

Residuals:
    Min       1Q   Median       3Q      Max
-2.16962 -0.24535 -0.00035  0.24894  1.97853

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.166223   0.036005  143.486 < 2e-16 ***
ED           0.079537   0.002260   35.189 < 2e-16 ***
EXP          0.011382   0.000561   20.290 < 2e-16 ***
UNION        0.091105   0.012871    7.078 1.71e-12 ***
MS           0.281544   0.015547   18.109 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3822 on 4160 degrees of freedom
Multiple R-squared:  0.3149, Adjusted R-squared:  0.3143
F-statistic: 478.1 on 4 and 4160 DF, p-value: < 2.2e-16
```

Question 2

```
call:
lm(formula = LWAGE ~ ED + EXP + UNIONC + MSC, data = don.sas)

Residuals:
    Min       1Q   Median       3Q      Max
-2.16962 -0.24535 -0.00035  0.24894  1.97853

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.538872   0.033188  166.893 < 2e-16 ***
ED           0.079537   0.002260   35.189 < 2e-16 ***
EXP          0.011382   0.000561   20.290 < 2e-16 ***
UNIONC       -0.091105   0.012871   -7.078 1.71e-12 ***
MSC          -0.281544   0.015547  -18.109 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3822 on 4160 degrees of freedom
Multiple R-squared:  0.3149, Adjusted R-squared:  0.3143
F-statistic: 478.1 on 4 and 4160 DF, p-value: < 2.2e-16
```

Question 3

```
call:
lm(formula = LWAGE ~ ED + WKS + EXP + EXP2 + UNION + MS + SMSA +
    OCC, data = don.sas)

Residuals:
    Min       1Q   Median       3Q      Max
-2.18732 -0.24706 -0.01177  0.24110  2.08426
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.886e+00	7.055e-02	69.256	< 2e-16 ***
ED	6.072e-02	2.669e-03	22.755	< 2e-16 ***
WKS	5.611e-03	1.123e-03	4.996	6.09e-07 ***
EXP	3.911e-02	2.243e-03	17.436	< 2e-16 ***
EXP2	-6.496e-04	4.935e-05	-13.164	< 2e-16 ***
UNION	1.133e-01	1.307e-02	8.665	< 2e-16 ***
MS	2.947e-01	1.505e-02	19.585	< 2e-16 ***
SMSA	1.377e-01	1.240e-02	11.105	< 2e-16 ***
OCC	-1.293e-01	1.519e-02	-8.515	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3639 on 4156 degrees of freedom

Multiple R-squared: 0.3794, Adjusted R-squared: 0.3782

F-statistic: 317.6 on 8 and 4156 DF, p-value: < 2.2e-16

Question 4

Analysis of Variance Table

Model 1: ED ~ WKS + EXP + EXP2 + UNION + MS + SMSA + OCC

Model 2: ED ~ WKS + EXP + EXP2 + UNION + MS + SMSA + OCC + SOUTH + FEM + BLK

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4157	18598				
2	4154	18149	3	449.47	34.292	< 2.2e-16 ***

Call:

lm(formula = LWAGE ~ ED + WKS + EXP + EXP2 + UNION + MS + SMSA + OCC + residuals(etude1), data = don.sas)

Residuals:

	Min	1Q	Median	3Q	Max
	-2.21283	-0.24233	-0.01236	0.23765	2.16670

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.582e+00	2.667e-01	5.930	3.27e-09 ***
ED	2.741e-01	1.684e-02	16.278	< 2e-16 ***
WKS	9.229e-03	1.137e-03	8.116	6.30e-16 ***
EXP	4.847e-02	2.318e-03	20.910	< 2e-16 ***
EXP2	-6.324e-04	4.842e-05	-13.060	< 2e-16 ***
UNION	1.871e-01	1.406e-02	13.312	< 2e-16 ***
MS	1.724e-01	1.757e-02	9.811	< 2e-16 ***
SMSA	2.514e-02	1.500e-02	1.676	0.0937 .
OCC	5.477e-01	5.484e-02	9.987	< 2e-16 ***
residuals(etude1)	-2.186e-01	1.704e-02	-12.828	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.357 on 4155 degrees of freedom

Multiple R-squared: 0.403, Adjusted R-squared: 0.4017

F-statistic: 311.7 on 9 and 4155 DF, p-value: < 2.2e-16

Call:

```
ivreg(formula = LWAGE ~ ED + WKS + EXP + EXP2 + UNION + MS +
      SMSA + OCC | SOUTH + FEM + BLK + WKS + EXP + EXP2 + UNION +
      MS + SMSA + OCC, data = don.sas)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.79051	-0.41060	-0.03349	0.37673	2.03705

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.582e+00	4.332e-01	3.651	0.000264	***
ED	2.741e-01	2.735e-02	10.022	< 2e-16	***
WKS	9.229e-03	1.847e-03	4.997	6.07e-07	***
EXP	4.847e-02	3.765e-03	12.874	< 2e-16	***
EXP2	-6.324e-04	7.865e-05	-8.041	1.15e-15	***
UNION	1.871e-01	2.283e-02	8.196	3.28e-16	***
MS	1.724e-01	2.854e-02	6.040	1.67e-09	***
SMSA	2.514e-02	2.436e-02	1.032	0.302072	
OCC	5.477e-01	8.907e-02	6.149	8.54e-10	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5798 on 4156 degrees of freedom

Multiple R-Squared: -0.5752, Adjusted R-squared: -0.5783

Wald test: 112.2 on 8 and 4156 DF, p-value: < 2.2e-16

ANNEXE : Insérez vos programmes SAS ou R ci-dessous.

```
#chargement du fichier
```

```
load("C:/Users/Chris/Desktop/Prepa_M2/Econometrie/data_exam.RData")
```

```
#Aperçu du jeu de données
```

```
summary(don.sas)
```

Question 1

```
mod1=lm(LWAGE~ED+EXP+UNION+MS,data=don.sas)
```

```
summary(mod1)
```

Question 2

```
don.sas$UNIONC=1-don.sas$UNION
```

```
don.sas$MSC=1-don.sas$MS
```

```
mod2=lm(LWAGE~ED+EXP+UNIONC+MSC,data=don.sas)
```

```
summary(mod2)
```

Question 3

```
#création de la variable EXP2
```

```
don.sas$EXP2=don.sas$EXP^2
```

```
mod3=lm(LWAGE~ED+WKS+EXP+EXP2+UNION+MS+SMSA+OCC,data=don.sas)
```

```
summary(mod3)
```

```
R=matrix(c(0,0,0,0,0,0,0,1,1),nrow=1,ncol=9,byrow=TRUE)
```

```
c=matrix(c(0),nrow=1,ncol=1,byrow=TRUE)
```

```
fishernc(mod3,R,c)
```

Question 4

```
#Correlation IV avec ED
```

```
etude2=lm(ED~WKS+EXP+EXP2+UNION+MS+SMSA+OCC,data=don.sas)
```

```
etude1=lm(ED~WKS+EXP+EXP2+UNION+MS+SMSA+OCC+SOUTH+FEM+BLK,data=don.sas)
```

```
anova(etude2,etude1)
```



```
#test exogénéité de ED
```

```
endo=lm(LWAGE~ED+WKS+EXP+EXP2+UNION+MS+SMSA+OCC+residuals(etude1),data=don.sas)
```

```
summary(endo)
```

```
#2SLS
```

```
iv=ivreg(LWAGE~ED+WKS+EXP+EXP2+UNION+MS+SMSA+OCC | WKS+EXP+EXP2+UNION+MS+SMSA+OCC+SOUTH+FEM+BLK,data=don.sas)
```

```
#test de restrictions suridentifiant
```

```
sargan=lm(residuals(iv)~SOUTH+FEM+BLK+WKS+EXP+EXP2+UNION+MS+SMSA+OCC,data=don.sas)
```

```
summary(sargan)
```

```
jour_sargan=summary(sargan)
```

```
n=nrow(don.sas)
```

```
r2iv=jour_sargan$r.square
```

```
r2iv
```

```
obs2=jour_sargan$r.square*n
```

```
ddl2=2
```

```
pvalue2=1-pchisq(obs2,ddl2)
```

```
print(pvalue2)
```