

**TD n°4 d'économétrie : corrigé**

## Exercice 1

```

/*transformation des variables*/
data tdmco;set centrale.tdeco;
lva=log(vabcf);
leffec=log(efec);
libd=log(ibd);
run;

/*modélisation cobb douglas*/
proc reg data=tdmco;
model lva=leffec libd ;
run;

```

Sortie sas :

## Analyse de variance

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	2	2542.04597	1271.02298	14187.3	<.0001
Error	1655	148.26920	0.08959		
Corrected Total	1657	2690.31517			

Root MSE	0.29931	R-Square	0.9449
Dependent Mean	10.29668	Adj R-Sq	0.9448
Coeff Var	2.90689		

## Résultats estimés des paramètres

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr >  t
Intercept	1	3.13556	0.05079	61.74	<.0001
leffec	1	0.73829	0.01183	62.40	<.0001
libd	1	0.28160	0.00888	31.70	<.0001

/\*Lecture des résultats (cf cours pour les notations) :

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	K	Variance expliquée			
Error	N-K-1	Variance résiduelle (=SCR)			
Corrected Total	N-1	Variance totale			

**Remarque : DF = nombre de degrés de liberté**

**A droite : Stat Fischer : nullité jointe de tous les paramètres du modèle**  
**Pr > F : probabilité d'obtenir cette valeur de la statistique de Fischer si l'hypothèse nulle (tous les paramètres du modèle sauf la constante sont nuls) est vraie. Ici la probabilité est très faible, donc on rejette l'hypothèse nulle\*/**

**Le R2 (R-Square) vaut 0.9449, donc l'ajustement est de bonne qualité : la production est bien expliquée par les variables travail et capital.**

**L'effet de leffec (ln de l'effectif) sur la production est estimé à : 0.73829. L'écart type de ce paramètre estimé est de 0.01183.**

Rappel (voir cours) : on a  $T = \frac{\hat{b} - b}{\hat{\sigma}_b}$  suit une loi de Student à (N-K-1) degrés de liberté.

Intervalles de confiance à 5% :  $\left[ \hat{b} \pm \hat{\sigma}_b t_{97.5}(N-K-1) \right]$

Avec  $t_{97.5}(N-K-1)$  quantile de la loi de Student à 97.5% (ie : une variable aléatoire qui suit une loi de student à N-K-1 degré de liberté a une proba de 97.5% d'être inférieure à cette valeur).

Ici : N=1358, K=2 donc  $t_{97.5}(N-K-1)=1.96$

Donc on déduit les intervalles de confiance à 5% :

Pour a : [3,04 ; 3,24]

Pour b : [0,72 ; 0,76]

Pour c : [0,08 ; 0,48]

Les deux dernières colonnes indiquent la statistique de student liée à un test

de l'hypothèse nulle H0 : Le coefficient est égal à 0 :  $T = \frac{\hat{b}}{\hat{\sigma}_b}$

Pour leffet, le T de student est de 61,74. La probabilité d'obtenir cette valeur alors que l'hypothèse H0 (le coefficient est nul) est inférieure à <.0001 (dernière colonne).

Attention, la probabilité est associée à un test bilatéral.

Test unilatéral : Test de H0 : c=0 contre Ha : c>0

Pour un test unilatéral, la région critique du test à 5% est :

$$\frac{\hat{c}}{\hat{\sigma}_c} > t_{95\%}(N-K-1) = 1,64$$

Ici  $\frac{\hat{c}}{\hat{\sigma}_c} = 28$  donc on REJETTE l'hypothèse nulle au seuil de 5% (on a moins de 5%

de chance de se tromper en refusant H0 alors qu'elle est vraie).

/\*test de constance des rendements d'échelle\*/  
/\*cela se fait automatiquement sous sas par l'instruction test\*/

```
proc reg data=tdmco;
model lva=leffec libd ;
test leffec+libd=1;run;
```

/\*on a en plus des résultats précédents : \*/

Test 1 Results for Dependent Variable lva

Source	DF	Carré moyen	Valeur F	Pr > F
Numerator	1	0.84817	9.47	0.0021
Denominator	1655	0.08959		

Valeur F représente la statistique de Fischer de l'hypothèse nulle de rendements d'échelle constants : ici 9.47

Elle suit une loi de Fischer de paramètre (N-K-1, p) donc ici (1655,1)\*/

La dernière colonne donne la probabilité que F prenne cette valeur alors que l'hypothèse nulle est vérifiée : ici 0.0021

Si on considère par exemple un test de niveau 5%, on REFUSE l'hypothèse de rendements constants(ie. On a moins de 5% de chance de se tromper en refusant H0 alors qu'elle est vraie). C'est vrai aussi si l'on considère un niveau de seuil plus faible (1%).

/\*test égalité du coefficient à 0.75\*/

```
proc reg data=tdmco;
model lva=leffec libd ;
test leffec=0.75;run;
```

Test 1 Results for Dependent Variable lva

Source	DF	Carré moyen	Valeur F	Pr > F
Numerator	1	0.08774	0.98	0.3225
Denominator	1655	0.08959		

Ici la statistique de Fischer vaut 0.98. On a 0.3225% chances d'obtenir cette valeur de F si on a effectivement leffec=0.75.

Donc on ACCEPTE l'hypothèse de rendements constants à 5%.

/\*régression par sous secteurs\*/

/\*démarche : on construit une table qui contient les trois sous secteurs considérés, et on fait une régression par sous secteur\*/

```
data b;set tdmco;
if n40b='T07' or n40b='T15B' or n40b="T18";
proc sort data=b;
by n40b;
proc reg data=b;
model lva=leffec libd ;
by n40b;run;
```

N40B=T07					
Model	2	78.54437	39.27219	566.20	<.0001
Error	37	2.56635	0.06936		
Corrected Total	39	81.11072			
Root MSE		0.26336	R-Square	0.9684	
Dependent Mean		11.13117	Adj R-Sq	0.9666	
Coeff Var		2.36601			

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr >  t
Intercept	1	3.30457	0.34710	9.52	<.0001
leffec	1	0.80992	0.08788	9.22	<.0001
libd	1	0.22008	0.06738	3.27	0.0024

N40B=T15B					
Model	2	40.11706	20.05853	361.42	<.0001
Error	21	1.16549	0.05550		
Corrected Total	23	41.28255			
Root MSE		0.23558	R-Square	0.9718	
Dependent Mean		11.08856	Adj R-Sq	0.9691	
Coeff Var		2.12456			

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr >  t
Intercept	1	4.08448	0.47279	8.64	<.0001
leffec	1	1.01736	0.10728	9.48	<.0001
libd	1	0.02390	0.09626	0.25	0.8063

N40B=T18					
Model	2	246.97984	123.48992	1630.71	<.0001
Error	235	17.79596	0.07573		
Corrected Total	237	264.77580			
Root MSE		0.27519	R-Square	0.9328	
Dependent Mean		9.83838	Adj R-Sq	0.9322	
Coeff Var		2.79707			

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr >  t
Intercept	1	3.59167	0.12142	29.58	<.0001
leffec	1	0.81978	0.02849	28.78	<.0001
libd	1	0.17365	0.01972	8.80	<.0001

Dans tous les cas, le R2 est assez élevé, donc la régression est de bonne qualité.

Le secteur T15B est plus intensif en main d'œuvre qu'en investissement : on ne peut pas refuser l'hypothèse de nullité de libd pour ce secteur à 5%.

Tous les autres paramètres sont significatifs.

**/\*Pour tester que les trois secteurs sont régis par le même modèle : test de Chow (cf. chap 2 du cours)\*/**

**/\* modèle contraint : les trois secteurs suivent le même modèle : on fait une régression unique sur l'ensemble de ces trois secteurs\*/**

```
proc reg data=b;
model lva=leffec libd ;
run;
```

The REG Procedure

Model: MODEL1  
Dependent Variable: lva

Number of Observations Read 302  
Number of Observations Used 302

Analyse de variance

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	2	446.43254	223.21627	2903.30	<.0001
Error	299	22.98825	0.07688		
Corrected Total	301	469.42079			

Root MSE 0.27728 R-Square 0.9510  
Dependent Mean 10.10896 Adj R-Sq 0.9507  
Coeff Var 2.74290

Résultats estimés des paramètres

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr >  t
Intercept	1	3.34893	0.09658	34.67	<.0001
leffec	1	0.81868	0.02504	32.69	<.0001
libd	1	0.20245	0.01673	12.10	<.0001

On a besoin de :

Somme des carrés des résidus du modèle contraint :  $SCR_0 = 22.98825$

Somme des carrés des résidus du modèle noncontraint = somme des carrés des résidus des trois régressions séparées par secteur :

$SCR_a = 17.79596 + 2.56635 + 1.16549 = 21.5278$

La statistique de Fischer vaut :

$$F = \frac{SCR_0 - SCR_a}{SCR_a} \frac{N_1 + N_2 + N_3 - 3K}{2K} = \frac{22.98825 - 21.5278}{21.5278} \frac{293}{6} = 3,32$$

Si  $H_0$  (stabilité) est vraie, F suit une loi de Fischer de paramètres (6,293)

Le quantile à 5% de la loi de Fischer à (6,293) vaut 2,09.

Donc on rejette l'hypothèse de stabilité à 5% : les trois secteurs ne semblent pas avoir la même fonction de production.

```
/*modélisation CES */
```

```
data c;set tdmco;  
ll=log(vabcf/effec);  
lk=log(ibd/effec);  
lk2=lk**2;  
run;  
proc reg ;  
model ll=leffec lk lk2;run;
```

Analyse de variance					
Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	3	102.26553	34.08851	395.51	<.0001
Error	1654	142.55792	0.08619		
Corrected Total	1657	244.82344			
	Root MSE	0.29358	R-Square	0.4177	
	Dependent Mean	4.53650	Adj R-Sq	0.4167	
	Coef Var	6.47153			

Résultats estimés des paramètres

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr >  t
Intercept	1	3.94938	0.11170	35.36	<.0001
leffec	1	0.01980	0.00634	3.12	0.0018
lk	1	-0.09640	0.04725	-2.04	0.0415
lk2	1	0.04235	0.00520	8.14	<.0001

**Le coefficient de Lk2 est non nul : on rejette l'hypothèse d'une fonction de type Cobb Douglas, au profit d'une modélisation CES.**

## Exercice 2 : régression logistique

```
data td;
set centrale.tdeco2;
/*creation de variables dichotomiques liées à la taille de la commune*/
array com com1-com5;
do i=1 to 5;
com[i]=(tur5=i);
end;
/*creation de variables dichotomiques liées au statut*/
array cs cs1-cs7;
do i=1 to 7;
cs[i]=(cstotr=i);
end;
run;

/*regression logistique */

proc logistic data=td descending;
model pro=com2-com5 ag2-ag5 seul mono cou1 cou2 cou3 cs1-cs6 ;
run;

/*Remarque : on a pris comme modalité de référence :
pour l'âge : avoir plus de 60 ans
pour la cs : être retraité
pour la taille de la commune : habiter en commune rurale
pour type de famille : être en couple sans enfant
tous les résultats pour une variable doivent se lire par rapport à cette
modalité de référence*/
```

### Informations sur le modèle

Data Set	WORK.TD
Response Variable	pro
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read 34996

Valeur ordonnée	pro	Fréquence totale
1	1	20182
2	0	14814

Probability modeled is pro=1.

État de convergence du modèle  
Convergence criterion (GCONV=1E-8) satisfied.

### Statistiques d'ajustement du modèle

Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
	AIC	47690.106
SC	47698.569	37936.257
-2 Log L	47688.106	37726.997

Paramètre	DF	Estimation	Erreur std	Khi 2 de Wald	Pr > Khi 2
Intercept	1	2.5144	0.0415	3673.9962	<.0001
com2	1	-0.7751	0.0444	305.2141	<.0001
com3	1	-1.1716	0.0403	843.9582	<.0001
com4	1	-1.2050	0.0410	864.1314	<.0001
com5	1	-1.5553	0.0463	1129.3334	<.0001
ag2	1	-3.2222	0.0797	1633.8447	<.0001
ag3	1	-1.9258	0.0613	987.9886	<.0001
ag4	1	-1.0988	0.0591	345.5138	<.0001
ag5	1	-0.6183	0.0520	141.4717	<.0001
seul	1	-1.1885	0.0317	1406.3753	<.0001
mono	1	-1.1975	0.0509	553.9137	<.0001
cou1	1	-0.0675	0.0474	2.0282	0.1544
cou2	1	0.2562	0.0501	26.1401	<.0001
cou3	1	-0.0662	0.0647	1.0451	0.3066
cs1	1	0.5130	0.1352	14.3960	0.0001
cs2	1	0.5801	0.0753	59.3462	<.0001
cs3	1	0.9865	0.0621	252.6286	<.0001
cs4	1	0.5970	0.0581	105.6944	<.0001
cs5	1	-0.2390	0.0606	15.5505	<.0001
cs6	1	-0.3065	0.0560	29.9792	<.0001

**/\*lecture des résultats :**

Les cadres (cs=3) ont une proba plus grande d'être propriétaire de leur logement que toutes les autres catégories. Les ouvriers (cs=6) ont une proba plus faible que les employés (cs=5), ainsi que des retraités (cs de référence). Les couples avec deux enfants ont la proba max d'être propriétaire de leur logement.

Cette proba augmente également avec l'âge : toutes les catégories d'âge ont moins de chance d'être propriétaires que les plus de 60 ans (catégorie de référence). \*/

/\*la taille de la commune est corrélée avec le fait d'être propriétaire, mais il est difficile d'établir le sens de la causalité : il est probable que le fait de devenir propriétaire a aussi un impact sur la localisation géographique...\*/

/\*remarque 1 : voir support de cours : dans une telle modélisation il est plus difficile que dans le cas linéaire de donner une interprétation quantifiable de l'effet d'une variable explicative sur la variable dépendante. \*/

remarque : on explique le fait d'être propriétaire car on a ajouté l'option descending : sinon l'option par défaut sous sas est la valeur nulle de variable dépendante (ici ce serait ne pas être propriétaire de son logement\*/

```

/*variante : representation probit*/
proc logistic data=td descending ;
model pro=com2-com5 ag2-ag5 seul mono cou1 cou2 cou3 cs1-cs6/link=normit;
run;

```

Intercept	1	1.4885	0.0232	4114.7427	<.0001
com2	1	-0.4463	0.0257	301.7956	<.0001
com3	1	-0.6848	0.0233	864.6550	<.0001
com4	1	-0.7037	0.0237	880.7322	<.0001
com5	1	-0.9081	0.0270	1135.3124	<.0001
ag2	1	-1.8727	0.0446	1763.9772	<.0001
ag3	1	-1.1428	0.0361	1002.6687	<.0001
ag4	1	-0.6466	0.0350	340.7573	<.0001
ag5	1	-0.3581	0.0308	135.1518	<.0001
seul	1	-0.7068	0.0186	1440.0859	<.0001
mono	1	-0.7243	0.0303	569.9805	<.0001
cou1	1	-0.0428	0.0281	2.3167	0.1280
cou2	1	0.1525	0.0298	26.2654	<.0001
cou3	1	-0.0405	0.0388	1.0877	0.2970
cs1	1	0.2909	0.0760	14.6572	0.0001
cs2	1	0.3405	0.0443	59.0770	<.0001
cs3	1	0.5750	0.0364	249.1577	<.0001
cs4	1	0.3468	0.0343	102.3062	<.0001
cs5	1	-0.1494	0.0359	17.2916	<.0001
cs6	1	-0.1907	0.0332	32.9671	<.0001

**/\*on voit que le sens des résultats (seuls interprétables dans un modèle qualitatif) sont très proches du modèle logit : en pratique on peut donc faire l'un ou l'autre sans modifier les conclusions \*/**

```

/*test de l'hypothèse indep=agric */
proc logistic data=td descending;
model pro=com2-com5 ag2-ag5 seul mono cou1 cou2 cou3 cs1-cs6 ;
test cs1=cs2;
run;

```

Résultats des tests des hypothèses linéaires

Libellé	Chi 2 de Wald	DF	Pr > Chi 2
Test 1	0.2269	1	0.6339

**La proc logistic donne directement la valeur du test de Wald. Sous l'hypothèse nulle, elle suit un Khi2. Ici on a une proba de 0.6339 % d'obtenir cette valeur si l'hypothèse nulle est vraie : donc on ne peut pas rejeter l'hypothèse nulle : les indépendants ont le même comportement que les agriculteurs**