Économétrie Appliquée : Manuel des cas pratiques sur EViews et Stata

Jonas Kibala Kuma

Kinshasa, Mars 2018

Manuel d'Econométrie
(Inspiré de Kintambu Mafuku E.G. 2013, 4è édition)

Économétrie Appliquée : Recueil des cas pratiques sur EViews et Stata

Par

Jonas KIBALA KUMA

(DEA-PTC Economie/Unikin en cours)

--

Centre de Recherches Economiques et Quantitatives
(CREQ)

***

« Editions Ecodata »

Kinshasa, Mars 2018

Manuel d'Econométrie
(Inspiré de Kintambu Mafuku E.G. 2013, 4è édition)

Économétrie Appliquée : Recueil des cas pratiques sur EViews et Stata

Par

Jonas KIBALA KUMA

(DEA-PTC Economie/Unikin en cours)

--

Centre de Recherches Economiques et Quantitatives
(CREQ)

***

– 1ère édition –

« Editions Ecodata »

– Note –

Ce manuel d'Econométrie, qui s'inscrit dans le cadre de nos travaux/recherches sur « l'Econométrie appliquée » que nous nous efforçons de mettre à la disposition des chercheurs africains/congolais surtout et de tout bord, est un recueil des travaux pratiques qui s'inspire largement de l'ouvrage du Professeur KINTAMBU MAFUKU Emmanuel Gustave intitulé « Principes d'Econométrie » (4è édition, 206 pages). Cet ouvrage, riche en illustrations et à compter parmi les rares des éditions congolaises qui combinent la théorie et la pratique sur logiciels, est à mon avis indispensable lorsque l'on se propose d'apprendre l'Econométrie en théorie et en pratique à l'aide des logiciels « Eviews et Stata ». Toutefois, il restait une préoccupation sur cet ouvrage quant à la manière d'obtenir ou reproduire les outputs produits (estimations, tests, etc.). Le néophyte souhaitait certainement une initiation à Eviews et Stata (lister des procédures et commandes,…) pour mieux savourer la recette économétrique combien délicieuse que regorge cet ouvrage. Nous pensons avoir contribué à étancher sa soif à partir du moment où nous avons pris l'initiative de parcourir cet ouvrage en gardant à l'esprit un style concis et une approche orientée vers la pratique avec initiation aux logiciels Eviews 5 et Stata 9/10.

Dans ce manuel, nous ne nous limitons pas seulement à résumer les aspects théoriques tant soit peu – l'on notera que l'essentiel de nos résumés sur les aspects théoriques sont en manuscrit pour éviter justement le plagiat – mais aussi, sur base de mêmes séries et variables, nous faisons des analyses particulières de nature à s'ajouter ou compléter celles du Prof.

Par ailleurs, signalons qu'en plus de l'ouvrage du Professeur Kintambu, qui constitue notre pilori, nous avons consulté bien d'autres documents (Veuillez consulter nos références bibliographiques) pour la plupart axés sur la pratique de techniques économétriques et statistiques sur EViews et/ou Stata. Nous espérons rencontrer vos besoins au travers cette petite contribution.

Merci au Professeur Kintambu Mafuku Emmanuel Gustave (Université de Kinshasa) pour ses efforts de recherche qui nous inspirent.

Kinshasa, Mars 2018

CHAP 1 : MODELE DE REGRESSION LINEAIRE SIMPLE (FORMES FONCTIONNELLES)

1.1. Modèles logarithmiques

- **Modèle (1.1a) :** $Q_t = AP_t^{\beta_1} e^{\mu_t}$, avec Qt : *quantité et Pt : Prix.*

Modèle linéaire ou log-log (1.1b) : $\ln Q_t = \ln A + \beta_1 \ln P_t + \mu_t$ ou

$LQ_t = \beta_0 + \beta_1 LP_t + \mu_t$, Avec : $\beta_0 = \ln A$

- **Travail Demandé :**
  - Estimer le modèle linéaire (1.1b) par les MCO ;
  - Calculer l'élasticité prix de la demande (élasticité de la demande par rapport au prix).

a) Linéarisation

- Eviews : genr LQ=log(Q) ; genr LP=log(P) ;
- Stata : gen LQ=log(Q).

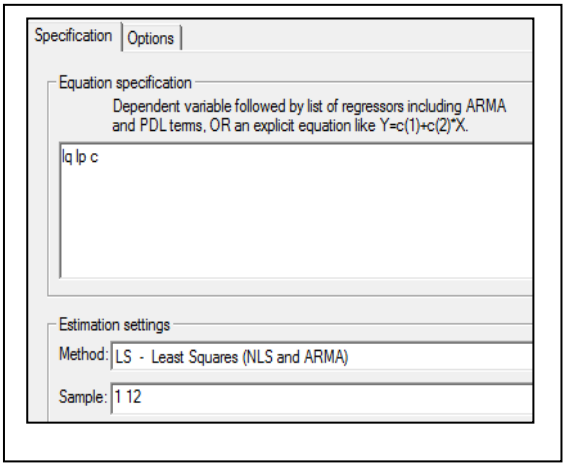
b) Estimation par les MCO

- **Eviews :** `ls lq c lp`

Dependent Variable: LQ
Method: Least Squares
Date: 10/12/13 Time: 17:46
Sample: 1 12
Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.773824	2.820948	-2.401258	0.0372
LP	1.966727	0.572769	3.433716	0.0064

R-squared	0.541082	Mean dependent var	2.910852
Adjusted R-squared	0.495191	S.D. dependent var	0.254430
S.E. of regression	0.180773	Akaike info criterion	-0.432142
Sum squared resid	0.326787	Schwarz criterion	-0.351325
Log likelihood	4.592854	F-statistic	11.79041
Durbin-Watson stat	1.652627	Prob(F-statistic)	0.006398



➢ **Stata :** `reg LQ LP`

```
reg LQ LP
      Source |       SS       df       MS              Number of obs =      12
-------------+------------------------------           F(  1,    10) =   11.79
       Model |  .385295076     1  .385295076           Prob > F      =  0.0064
    Residual |  .326787634    10  .032678763           R-squared     =  0.5411
-------------+------------------------------           Adj R-squared =  0.4952
       Total |   .71208271    11  .064734792           Root MSE      =  .18077
------------------------------------------------------------------------------
          LQ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          LP |   1.966726   .5727695     3.43   0.006     .6905157    3.242935
       _cons |  -6.773818   2.820949    -2.40   0.037    -13.05928   -.4883522
------------------------------------------------------------------------------
```

c) Elasticité

Elasticité : $E = \frac{dQ_t}{dP_t}\frac{P_t}{Q_t} = 1{,}96673$ (si « P » augmente de 1%, « Q » augmente de 1,9%).

1.2. Modèles semi-logarithmiques

- **Modèle (1.2a)/intérêt composé :** $Y_t = Y_0(1+r)^t$, avec Yt : *quantité vendue, r = taux de croissance de vente* et t = temps.

Modèle log-linéaire (1.2b) : $\ln Y_t = \ln Y_0 + t\ln(1+r) + \mu_t$ ou $LY = \beta_0 + \beta_1 t + \mu_t$, Avec : $\beta_0 = \ln Y_0, \ln(1+r) = \beta_1$.

- **Travail Demandé :**
  - Estimer le modèle log-linéaire (1.2b) par les MCO ;
  - Calculer « Yo » et « r » ;
  - Représenter la droite de régression du modèle 2.2b estimé.

a) Estimation du modèle 1.2b

➢ **Eviews :**

Dependent Variable: LYT
Method: Least Squares
Date: 10/12/13 Time: 18:59
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.552753	0.070942	35.98344	0.0000
T	0.075895	0.007803	9.726851	0.0000

R-squared	0.879195	Mean dependent var	3.159913
Adjusted R-squared	0.869903	S.D. dependent var	0.361981
S.E. of regression	0.130563	Akaike info criterion	-1.110357
Sum squared resid	0.221607	Schwarz criterion	-1.015950
Log likelihood	10.32768	F-statistic	94.61163
Durbin-Watson stat	1.635337	Prob(F-statistic)	0.000000

Sur Eviews :

genr lYT=log(YT) ;
ls YT c T

Dans l'output, Suivre :

`View/representations` :

Estimation Command:
==========
LS LYT C T

Estimation Equation:
==========
*LYT = C(1) + C(2)*T*

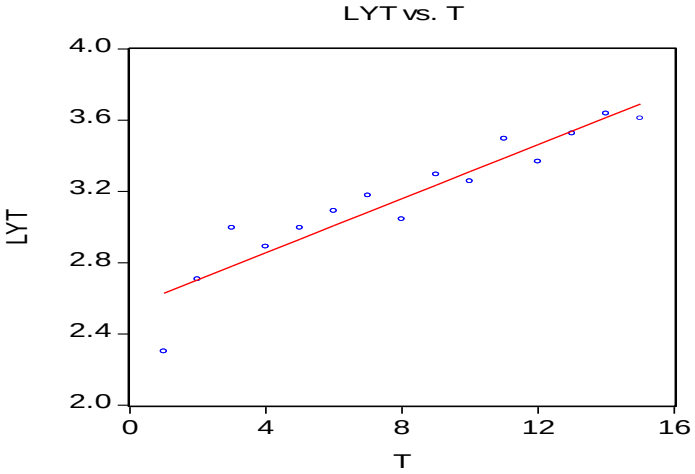
Substituted Coefficients:
==========
*LYT = 2.552752778 + 0.07589500036*T*

b) Calcul de « Yo et r »

- $Y_0 = e^{\beta_0} = e^{2,552753} = 12,84241032$;
- $e^{\ln(1+r)} = e^{\beta_1} \Rightarrow 1 + r = e^{0,075895} \Longrightarrow r = 1,078849289 - 1 = 0,078849289$;
- $\hat{Y}_t = 12,84241032(1,078849289)^t$

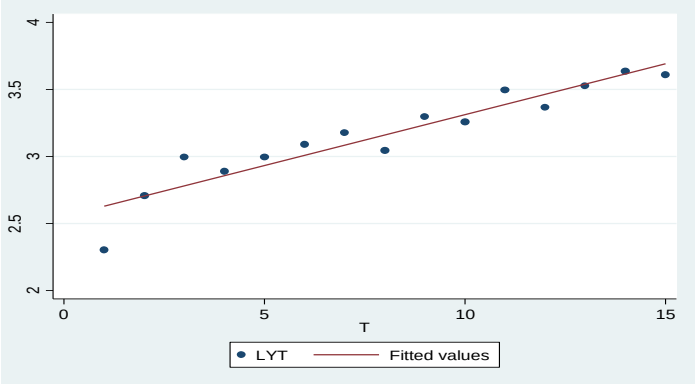
c) Droite de régression du modèle 2.2b estimé

Eviews :



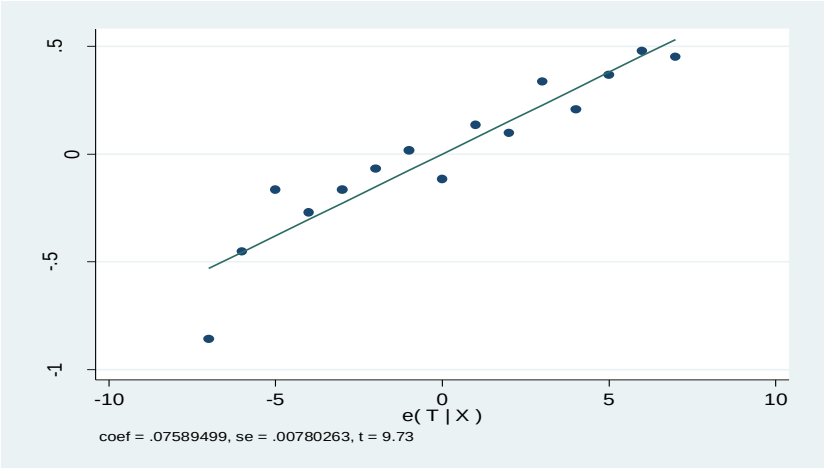
Nuage de points avec la droite de régression sur EViews : *Sélectionner 2 variables/clic droit/Open/as Group. Dans la boîte de dialogue, Suivre : View/Graph/Scatter With Regression.. (dans Global Fit Options : None ; None) Ok.*

Stata :



Commande Stata :

```
twoway(scatter LYT T) (lfit LYT T)
ou
{ reg LYT T
  avplots
```



1.2. Modèles semi-logarithmiques (suite)

Modèle linéaire-logarithmique (2.3)/Modèle de dépense d'Engel :

$DS_t = a_0 + a_1 \ln(RM)_t$, avec DS_t *: Dépenses en soins de Santé (DS) et* RM_t *: Niveau de salaire payé aux chefs de ménages. Soit :* $DS_t = a_0 + a_1 LRM_t + \mu_t$, avec $LRM_t = \ln(RM)_t$.

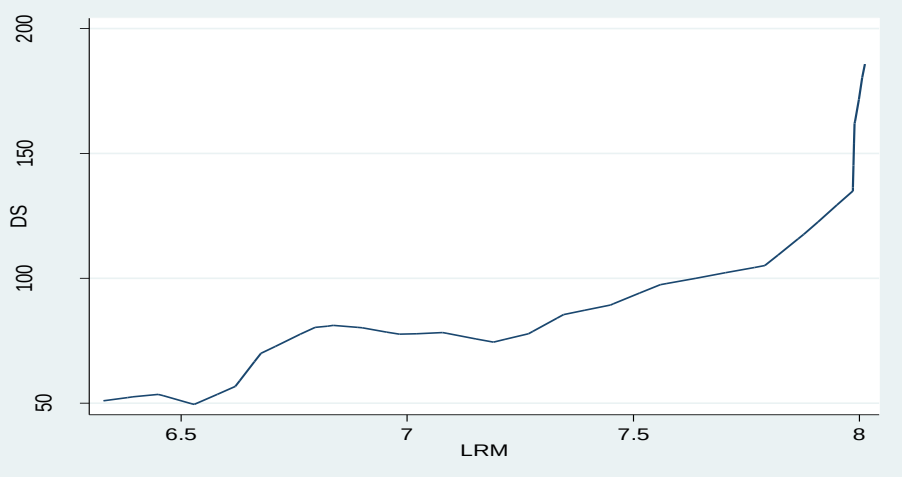
Travail Demandé : estimer le modèle 2.3 par les MCO

a) Estimation (avec Stata)

```
reg DS LRM

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  1,    19) =   72.43
       Model |  20412.5921     1  20412.5921           Prob > F      =  0.0000
    Residual |  5354.61239    19  281.821705           R-squared     =  0.7922
-------------+------------------------------           Adj R-squared =  0.7813
       Total |  25767.2045    20  1288.36023           Root MSE      =  16.788

------------------------------------------------------------------------------
          DS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         LRM |   56.67096   6.658841     8.51   0.000     42.73384    70.60807
       _cons |  -316.2399   47.71794    -6.63   0.000    -416.1147   -216.3651
------------------------------------------------------------------------------
```

b) Relation graphique entre DS et LRM

Commande Stata : `twoway line DS LRM, lpattern(l) xtitle() ytitle()`



CHAP 2 : MODELE DE REGRESSION LINEAIRE AVEC DE VARIABLES BINAIRES INDEPENDANTES

2.1. Modèle ANOVA (régression var quantitative sur une ou +++ var binaires)

- **Modèle (2.1a)** : $Y_i = \beta_1 + \beta_2 D_i + \mu_i$

  Avec :
  - Y_i : Salaire payé ;
  - $D_i = \begin{cases} 1, si\ le\ salaire\ est\ payé\ à\ un\ salarié\ national \\ 0, si\ le\ salaire\ est\ payé\ à\ un\ salarié\ étranger \end{cases}$

Travail Demandé :

- Estimer le modèle linéaire (2.1a) par les MCO ;
- Interpréter les résultats.

a) Estimation du modèle 2.1a par les MCO : distribution de revenus selon la nationalité

```
reg Yi Di

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =    2.58
       Model |  64.0666667     1  64.0666667           Prob > F      =  0.1466
    Residual |  198.333333     8  24.7916667           R-squared     =  0.2442
-------------+------------------------------           Adj R-squared =  0.1497
       Total |       262.4     9  29.1555556           Root MSE      =  4.9791

------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Di |  -5.166667    3.21401    -1.61   0.147    -12.57819    2.244854
       _cons |       63.5   2.489562    25.51   0.000     57.75906    69.24094
------------------------------------------------------------------------------
```

b) Interprétation

- Salaire moyen des étrangers : $SME = 63.5$;
- Salaire moyen des nationaux : $SMN = 63.5 - 5.17 = 58.33$;
- β_2 est non significatif (prob > 5%, |t-stat calculé| < 2) : c.à.d., pas de différence significative entre les salaires de nationaux et des étrangers.

- **Modèle (2.1b)** : $Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \mu_i$

  Avec :
  - Y_i : Salaire payé ;
  - $D_{1i} = \begin{cases} 1, si\ le\ salaire\ est\ payé\ à\ un\ salarié\ national \\ 0, si\ le\ salaire\ est\ payé\ à\ un\ salarié\ étranger \end{cases}$
  - $D_{2i} = \begin{cases} 1, si\ le\ salarié\ est\ un\ homme \\ 0, si\ le\ salarié\ est\ une\ femme \end{cases}$

Travail Demandé :

- Estimer le modèle <u>linéaire</u> (2.1a) par les MCO ;
- Interpréter les résultats.

a) Estimation du modèle linéaire (2.1b) : distribution de revenus selon la nationalité et selon le sexe

```
reg Yi D1i D2i

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  2,     7) =    1.14
       Model |  64.4666667     2  32.2333333           Prob > F      =  0.3728
    Residual |  197.933333     7  28.2761905           R-squared     =  0.2457
-------------+------------------------------           Adj R-squared =  0.0302
       Total |       262.4     9  29.1555556           Root MSE      =  5.3175

------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         D1i |  -5.166667   3.432455    -1.51   0.176    -13.28313    2.949799
         D2i |        -.4   3.363105    -0.12   0.909     -8.35248     7.55248
       _cons |        63.7   3.145897    20.25   0.000     56.26114    71.13886
------------------------------------------------------------------------------
```

b) Interprétation

Les deux variables explicatives binaires (D1i et D2i) ne sont pas significatives. D'où, pas de discrimination d'après la nationalité, ni d'après le sexe, dans la distribution de salaires.

2.2. Modèle ANCOVA (régression var quantitative sur mélange de <u>var binaires et quantitatives)</u>

- **Modèle (2.2)** : $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \mu_i$

  Avec :
  - Y_i : Salaire moyen des salariés ;
  - X_i : Nombre d'années d'expérience ;
  - $D_i = \begin{cases} 1, si\ le\ salarié\ est\ national \\ 0, si\ le\ salarié\ est\ étranger \end{cases}$

<u>Travail Demandé</u> :

- Estimer le modèle linéaire (2.2) par les MCO ;
- Interpréter les résultats.

a) Estimation du modèle 2.2 par les MCO : distribution de salaires selon la nationalité et selon le nombre d'années d'expérience

```
reg Yi Di Xi
      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  2,     7) =    1.25
       Model |  68.8952381     2   34.447619           Prob > F      =  0.3444
    Residual |  193.504762     7  27.6435374           R-squared     =  0.2626
-------------+------------------------------           Adj R-squared =  0.0519
       Total |       262.4     9  29.1555556           Root MSE      =  5.2577
------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
          Di |  -4.980952   3.422805    -1.46   0.189     -13.0746    3.112695
          Xi |   .7428571   1.777431     0.42   0.689      -3.4601    4.945814
       _cons |   61.08571   6.346699     9.62   0.000     46.07816    76.09327
------------------------------------------------------------------------------
```

b) Interprétation

La nationalité, ni le nombre d'années d'expérience ne sont de facteurs de discrimination dans la répartition de salaires.

2.3. Régression var quantitative sur une var quantitative et deux var binaires

- **Modèle (2.3)** : $Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 X_i + \mu_i$

  Avec :
  - Y_i : Salaire moyen des salariés ;
  - X_i : Nombre d'années d'expérience ;
  - $D_{1i} = \begin{cases} 1, si\ le\ salaire\ est\ payé\ à\ un\ salarié\ national \\ 0, si\ le\ salaire\ est\ payé\ à\ un\ salarié\ étranger \end{cases}$
  - $D_{2i} = \begin{cases} 1, si\ le\ salarié\ est\ un\ homme \\ 0, si\ le\ salarié\ est\ une\ femme \end{cases}$

Travail Demandé :

- Estimer le modèle linéaire (2.3) par les MCO ;
- Interpréter les résultats.

a) Estimation du modèle 2.3 par les MCO : distribution de salaires d'après la nationalité, le sexe et le nombre d'années d'expérience

```
reg Yi Xi D1i D2i
      Source |       SS       df       MS              Number of obs =      12
-------------+------------------------------           F(  3,     8) =    2.43
       Model |  741.116049     3  247.038683           Prob > F      =  0.1407
    Residual |  814.883951     8  101.860494           R-squared     =  0.4763
-------------+------------------------------           Adj R-squared =  0.2799
       Total |        1556    11  141.454545           Root MSE      =  10.093

------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xi |    10.1284   4.810268     2.11   0.068    -.9641033    21.22089
         D1i |  -15.39506   7.092354    -2.17   0.062    -31.75006    .9599348
         D2i |   6.669136   6.303822     1.06   0.321    -7.867503    21.20577
       _cons |   33.30123   9.737468     3.42   0.009     10.84659    55.75588
------------------------------------------------------------------------------
```

b) Interprétation

- Salaire Moyen d'une ouvrière étrangère :

$$E(Y/D1 = 0, D2 = 0, X) = \beta_0 + \beta_3 X_i = 33.30123$$

- Salaire Moyen d'une ouvrière nationale :

$$E(Y/D1 = 1, D2 = 0, X) = (\beta_0 + \beta_1) + \beta_3 X_i = 33.30123 - 15.39506 = 17.90623$$

- Salaire Moyen d'un ouvrier étranger :

$$E(Y/D1 = 0, D2 = 1, X) = (\beta_0 + \beta_2) + \beta_3 X_i = 33.30123 + 6.669136 = 39.97036$$

- Salaire Moyen d'un ouvrier national :

$$E(Y/D1 = 1, D2 = 1, X) = (\beta_0 + \beta_1 + \beta_2 +) + \beta_3 X_i$$
$$= 33.30123 + 6.669136 - 15.3 = 24.574536$$

Ni le sexe, ni la nationalité, ni l'ancienneté n'influencent la distribution de salaires (au seuil de 5%).

2.4. Autres utilisations de variables binaires

2.4.1. Variables binaires et test de stabilité de paramètres

- **Modèle (2.4)** : $INV_t = \alpha + \beta_0 D_i + \beta_1 PIB_t + \beta_2 (D_i PIB_t) + \mu_i$

  Avec :
  - INV_t : Investissement ;
  - PIB_t : Produit Intérieur Brut ;
  - $D_i = \begin{cases} 1, pour\ la\ sous - période\ 1988/1995\ (après\ le\ choc) \\ 0, pour\ la\ sous - période\ 1980/1987\ (avant\ le\ choc) \end{cases}$
  - β_0 : $intercepte\ différentiel$;
  - β_2 : $coefficient\ différentiel\ de\ la\ pente$

Travail Demandé :

- Estimer le modèle linéaire (2.4) par les MCO ;
- Interpréter les résultats (quel est l'impact du choc pétrolier en 1988).

a) Estimation du modèle 2.4 par les MCO

Commande : reg INVT PIBT Di DiPIBT

```
      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  3,    12) =  259.52
       Model |  12773.5472     3  4257.84907           Prob > F      =  0.0000
    Residual |  196.883196    12   16.406933           R-squared     =  0.9848
-------------+------------------------------           Adj R-squared =  0.9810
       Total |  12970.4304    15   864.69536           Root MSE      =  4.0505

------------------------------------------------------------------------------
        INVT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PIBT |   .2275952   .0372635     6.11   0.000     .1464049    .3087854
          Di |  -2.420614    15.8336    -0.15   0.881    -36.91906    32.07783
      DiPIBT |   .0033554   .0401496     0.08   0.935    -.0841231     .090834
       _cons |  -27.29802   13.37531    -2.04   0.064    -56.44032    1.844277
------------------------------------------------------------------------------
```

b) Interprétation

$\hat{\beta}_0$ et $\hat{\beta}_2$ ne sont pas statistiquement significatifs : l'investissement n'a pas subi les effets du choc pétrolier. Autant dire que les paramètres du modèle sont restés stables pour les deux sous-périodes.

2.4.2. Test de CHOW : Stabilité de paramètres

Dans Stata, pour calculer la statistique de Chow, procédons comme suit :

- estimer le modèle général ;
- récupérer la SCR et n ;
- estimer le modèle 1 pour la 1ère sous-période ;
- récupérer la SCR1 (scalar SCR1=e(rss)) ;
- estimer le modèle 2 pour la 2ème sous période ;
- récupérer la SCR2 (scalar SCR2=e(rss)) ;
- calcul de la statistique de Chow : scalar stat=((scr-(scr1+scr2))/(scr1+scr2))*((n-2*2)/2) ;
- display stat : obtenir la valeur calculée ;
- display F(2, n-2*2, stat) : c'est la statistique du test correspondant au « p-value » (1-probabilité du test) de l'hypothèse nulle (H0 : stabilité et H1 : instabilité).

Test de CHOW appliqué sur le modèle 2.4 :

Le test de CHOW n'est pas programmé dans Stata, mais il peut se calculer comme suit [$\frac{SCR-(SCR1+SCR2)}{SCR1+SCR2} \times \frac{n-2k}{k} \rightarrow F(k, n-2k)$; Avec : n (nombre d'observations), k (nombre de paramètres), SCR/SCR1/SCR2 (les sommes des carrés des résidus de deux sous échantillons et de l'échantillon total] :

- $\begin{cases} \textbf{reg } INVT\ PIBT \\ \textbf{scalar } scr = \textbf{e(rss)} \\ \textbf{scalar } n = \textbf{e(N)} \end{cases}$: *Estimation sur l'ensemble de l'échantillon/période et récupération de « SCR et n ».*

```
. reg INVT PIBT
      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  1,    14) =  899.64
       Model |   12771.681     1   12771.681           Prob > F      =  0.0000
    Residual |  198.749377    14  14.1963841           R-squared     =  0.9847
-------------+------------------------------           Adj R-squared =  0.9836
       Total |  12970.4304    15   864.69536           Root MSE      =  3.7678

------------------------------------------------------------------------------
        INVT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PIBT |   .2268125   .0075619    29.99   0.000     .2105938    .2430312
       _cons |  -27.21259   3.587799    -7.58   0.000    -34.90766   -19.51753
------------------------------------------------------------------------------
. scalar scr=e(rss)
. scalar n=e(N)
```

Estimation sur le 1er sous-échantillon (les pauvres) et récupération de SCR1.

- $\begin{cases} \textbf{reg } INVT\ PIBT \textbf{ if } Di == 1 \\ \textbf{scalar } scr1 = \textbf{e(rss)} \end{cases}$:

```
reg INVT PIBT if Di==1
      Source |       SS       df       MS              Number of obs =       8
-------------+------------------------------           F(  1,     6) =  135.36
       Model |  3916.85936     1  3916.85936           Prob > F      =  0.0000
    Residual |  173.615959     6  28.9359932           R-squared     =  0.9576
-------------+------------------------------           Adj R-squared =  0.9505
       Total |  4090.47532     7  584.353617           Root MSE      =  5.3792

------------------------------------------------------------------------------
        INVT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PIBT |   .2309506   .0198504    11.63   0.000     .1823784    .2795228
       _cons |  -29.71864   11.25328    -2.64   0.038    -57.25443   -2.182843
------------------------------------------------------------------------------
. scalar scr1=e(rss)
```

- $\begin{cases} \textbf{reg } INVT\ PIBT \textbf{ if } Di == 0 \\ \textbf{scalar } scr2 = \textbf{e(rss)} \end{cases}$: $\begin{cases} \textit{Estimation sur le } 2^{ème} \textit{ sous-échantillon (les riches) et récupération de SCR2.} \end{cases}$

```
reg INVT PIBT if Di==0
      Source |       SS       df       MS              Number of obs =       8
-------------+------------------------------           F(  1,     6) =  157.83
       Model |  612.047809     1  612.047809           Prob > F      =  0.0000
    Residual |  23.2672362     6  3.87787271           R-squared     =  0.9634
-------------+------------------------------           Adj R-squared =  0.9573
       Total |  635.315045     7  90.7592921           Root MSE      =  1.9692

------------------------------------------------------------------------------
        INVT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PIBT |   .2275952   .0181162    12.56   0.000     .1832664    .2719239
       _cons |  -27.29802   6.502599    -4.20   0.006    -43.20931   -11.38674
------------------------------------------------------------------------------
. scalar scr2=e(rss)
```

- $\begin{cases} scalar\ stat = ((\text{scr} - (\text{scr1} + \text{scr2}))/(\text{scr1} + \text{scr2})) * ((\text{n} - 2 * 2)/2) \\ display\ stat \\ display\ F(2, n - 2 * 2, stat) \end{cases}$

<u>NB</u> :

- 1ère expression (scalar…) : calculer la statistique de Fisher F(2,n-2*2)) ;
- 2ème expression (display F) : afficher la valeur F-calculé : **`stat =0.05687175. Le F-théorique/F(0.05 ;2 ;12)= 3.89 ;`**
- 3ème expression (display F(2,n-2*2,stat)) : calculer la probabilité associée à la statistique calculée de Fisher (<u>NB</u> : par définition, la p-value est le complément à 1 de la probabilité du test. C.à.d. « p-value = 1-prob du test ») : prob F (p-value nulle) = **`1-0.05503171=0.9449683`**.

L'on constate ainsi que la « p-value » associée au F-stat est > à 5% (Fc < Ft, soit **`0.05687175 < 3.89`**). D'où, on accepte l'hypothèse nulle selon la quelle les paramètres estimés sont stables sur les deux sous-périodes/les deux régressions ne sont pas différentes (même résultat avec l'utilisation de variables binaires dans le modèle).

2.4.3. Variables binaires et analyse de saisonnalité

- **Modèle (2.5) :** $Y_t = \beta_0 + \beta_1 X_t + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 D_{3i} + \mu_i$

  Avec :

  - Y_i : Ventes ;
  - X_i : Frais engagés pour la publicité ;
  - $D_{1i} = \begin{cases} 1, pour\ tous\ les\ T1 \\ 0, pour\ T2, T3\ et\ T4 \end{cases}$
  - $D_{2i} = \begin{cases} 1, pour\ tous\ les\ T2 \\ 0, pour\ tous\ les\ T1, T3\ et\ T4 \end{cases}$
  - $D_{3i} = \begin{cases} 1, pour\ tous\ les\ T3 \\ 0, pour\ tous\ les\ T1, T2\ et\ T4 \end{cases}$

Travail Demandé :

- Estimer le modèle linéaire (2.5) par les MCO ;
- Interpréter les résultats (existe-t-il un effet saisonnier ?).

a) Estimation du modèle 2.5 par les MCO : relation ventes et dépenses en publicité d'après les quatre saisons/trimestres

- Déclarer les données trimestrielles à Stata : *Taper edit/Saisir les codes numériques du 1er trimestre 1999 au dernier trimestre 2002 (en commençant par 156, avec « 1 » comme raison. NB : 1999-1960=156). Après avoir fermé le data editor, taper :* ***format var6 %tq.*** *Les données se présentent en partie comme suit :*

var12[11] =

	YT	XT	D1i	D2i	D3i	var6
1	130	43	1	0	0	1999q1
2	145	36	0	1	0	1999q2
3	156	52	0	0	1	1999q3
4	139	27	0	0	0	1999q4
5	150	37	1	0	0	2000q1
6	147	49	0	1	0	2000q2
7	159	40	0	0	1	2000q3
8	166	32	0	0	0	2000q4
9	210	70	1	0	0	2001q1
10	199	66	0	1	0	2001q2
11	194	49	0	0	1	2001q3
12	202	67	0	0	0	2001q4

- En présence de variables « dummy/dichotomiques », la multi-colinéarité est un problème récurrent. Pour le contourner : soit (i) *estimer le modèle sans constante/intercepte et y intégrer toutes les variables dummy (les quatre dans notre cas : T1, T2, T3 et T4)*, ou (ii) *inclure l'intercepte tout en ignorant une variable dummy (T4 est ignoré dans notre cas).*

  Les résultats d'estimation sont :

```
reg YT XT D1i D2i D3i
      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  4,    11) =    6.97
       Model |  6584.26076     4  1646.06519            Prob > F      =  0.0048
    Residual |  2596.17674    11  236.016067            R-squared     =  0.7172
-------------+------------------------------           Adj R-squared =  0.6144
       Total |   9180.4375    15  612.029167            Root MSE      =  15.363
------------------------------------------------------------------------------
          YT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          XT |   1.551496   .3028663     5.12   0.000     .8848922      2.2181
         D1i |       -12.5   10.86315    -1.15   0.274    -36.40963     11.40963
         D2i |  -16.23173   10.94831    -1.48   0.166     -40.3288     7.865338
         D3i |  -1.448504   10.86737    -0.13   0.896    -25.36743     22.47042
       _cons |   101.3039   16.30844     6.21   0.000     65.40929     137.1986
------------------------------------------------------------------------------
```

b) Interprétation

β_2, β_3 et β_4 sont non significatifs : les ventes ne subissent pas l'effet de variations saisonnières.

CHAP 3 : MODELE AVEC DES REPONSES QUALITATIVES : Régression d'une variable binaire dépendante sur des variables quantitatives

3.1. Modèle Linéaire de Probabilité/MLP

Modèle pondéré/Modèle linéaire de probabilité-MLP (3.1a) :

$$Y_i = \beta_0 + \beta_0 X_i + \mu_i$$

Avec :

- $Y_i = \begin{cases} 1, la\ famille\ possède\ une\ voiture \\ 0, la\ famille\ ne\ possède\ pas\ une\ voiture \end{cases}$
- X_i : niveau moyen de revenu ;

Travail Demandé :

- Estimer le modèle linéaire (3.1a) ;
- Interpréter les résultats.

a) Estimation du modèle linéaire de probabilité 3.1a

L'on peut montrer que le terme d'erreur du modèle 3.1 n'est pas homoscédastique. Pour corriger ce modèle d'hétéroscédasticité, procéder comme suit :

- Estimer le ***modèle 3.1a*** par les MCO et obtenir $\hat{Y}_i$. NB : $0 \leq \hat{Y}_i \leq 1$. En outre,

$$\hat{Y}_i = \begin{cases} 1, si\ \hat{Y}_i > 1 \\ 0, si\ \hat{Y}_i < 0 \end{cases}$$

- Estimer le ***modèle 3.1b*** ci-dessous par les MCO (modèle 3.1a corrigé d'hétéroscédasticité) :

$$\frac{Y_i}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}} = \frac{\beta_0}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}} + \beta_1 \frac{X_i}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}} + \frac{\mu_i}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}}$$

Résultats :

(i) Estimation du modèle 3.1a par les MCO : Régression non pondérée

```
reg Yi Xi
      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  1,    28) =    2.67
       Model |   .65285358     1   .65285358           Prob > F      =  0.1135
    Residual |  6.84714642    28  .244540944           R-squared     =  0.0870
-------------+------------------------------           Adj R-squared =  0.0544
       Total |         7.5    29   .25862069           Root MSE      =  .49451
------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xi |   .0153613   .0094014     1.63   0.113    -.0038967    .0346193
       _cons |   .0837098   .2703032     0.31   0.759    -.4699812    .6374009
------------------------------------------------------------------------------
```

(ii) Régression Pondérée (MLP)

```
reg YC XC
      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  1,    28) =    4.24
       Model |  4.46342617     1  4.46342617           Prob > F      =  0.0490
    Residual |  29.5057561    28    1.053777           R-squared     =  0.1314
-------------+------------------------------           Adj R-squared =  0.1004
       Total |  33.9691822    29  1.17135111           Root MSE      =  1.0265
------------------------------------------------------------------------------
          YC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          XC |   .0175328    .008519     2.06   0.049     .0000823    .0349832
       _cons |   .0593206   .5211676     0.11   0.910    -1.008243    1.126884
------------------------------------------------------------------------------
```

Les commandes STATA sont :

```
predict YF, xb
gen YD=1-YF
gen YR=sqrt(YF*YD)
gen YC=Yi/YR
gen XC=Xi/YR
gen CO=1/YR
```

b) Interprétation

Les modèles 3.1a et 3.1b estimés s'écrivent :

- *Régression non pondérée :* $\hat{Y}_i = .0837098 + .0153613X_i$*……[3.1a]*
- *Régression pondérée :*

$$\frac{Y_i}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}} = 0.0593206 * \frac{1}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}} + 0.0175328 * \frac{X_i}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}} \text{.....[3.1b]}$$

Autrement : $YC = 0.0593206 * CO + 0.0175328 * XC$

Avec :

- $YF = \hat{Y}_i$;
- $YD = 1 - \hat{Y}_i$;
- $YR = \sqrt{\hat{Y}_i(1 - \hat{Y}_i)}$;
- $YC = \frac{Y_i}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}}$;
- $XC = \frac{X_i}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}}$;
- $CO = \frac{1}{\sqrt{\hat{Y}_i(1-\hat{Y}_i)}}$.

Partant du modèle 3.1b, si l'on estime que « Xi = 15 usd » (NB : si Xi=15 YR=0.464168), l'on pourra calculer la probabilité de posséder une voiture (E(Yi=1)) comme suit :

$$\frac{Y_i}{\sqrt{\hat{Y}_i(1 - \hat{Y}_i)}} = 0.0593206 * \frac{1}{0.464168} + 0.0175328 * \frac{15}{0.464168} = 70\%$$

Par contre, dans le modèle pondéré (3.1a), cette probabilité se calcule comme suit :

$$\hat{Y}_i = 0.0837098 + 0.0153613 * 15 = 31.4\%$$

Note : la régression pondérée donne de meilleurs résultats que la régression non pondérée.

3.2. Modèle LOGIT

- ***Modèle Logit*** : $P_i = P(Z) = \frac{1}{1+e^{-Z}}$ ... ... [3.2a], avec $Z = \beta_0 + \beta_1 X_i$

Modèle Logit linéaire (après développements) :

$$Z = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 X_i + \mu_i \text{ ... ... } [3.2b]$$

NB : si $P_i = 0 \; ou \; P_i = 1, \Rightarrow \left(\frac{P_i}{1-P_i}\right) = 0 \; ou \; \infty$. Dans ce cas, les MCO sont inefficaces. La méthode d'estimation adéquate dépend de la nature de données. *Question* : Les observations de « X_i » sont regroupées en classes ou pas ?

Hypothèse : *La fonction de répartition de l'erreur suit une loi de type logistique (Modèle Logit = Fonction logistique cumulée).*

- Si les données sont regroupées (considérer les valeurs moyennes de classes de X_i), *le modèle 3.2b* devient :

$$\ln\left(\frac{P_i}{1 - P_i}\right) = \ln\left(\frac{\frac{n_i}{N}}{1 - \frac{n_i}{N}}\right) = L_i = \beta_0 + \beta_1 X_i + \mu_i \text{ ... ... } [3.2c]$$

Avec :

- N_i : nombre total d'observations pour la classe « i » ;
- n_i : fréquence d'apparition/occurrence de X ;
- $\frac{n_i}{N}$: probabilité « P_i » pour chaque classe.

Pour corriger *le modèle 3.2c* d'hétéroscédasticité, sous réserve que $\sigma_\mu = \frac{1}{N_i \hat{P}_i(1-\hat{P}_i)}$, il tient de diviser ce modèle par « σ_μ », ce qui donne (modèle à estimer) :

$$L_i^* = \beta_0\sqrt{W_i} + \beta_1 X_i^* + \mu_i^* \dots\dots [3.2d]$$

Avec :

- $\sqrt{W_i} = \sqrt{N_i \hat{P}_i(1-\hat{P}_i)}$;
- $L_i = \ln\left(\frac{P_i}{1-P_i}\right)$ et $L_i^* = \sqrt{W_i} L_i$;
- $X_i^* = X_i\sqrt{W_i}$;
- $\mu_i^* = \mu_i\sqrt{W_i}$

- Si les données sont non regroupées, appliquer la méthode du Maximum de Vraisemblance au modèle suivant :

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \dots\dots [3.2e]$$

Avec :

- $Y_i = \begin{cases} 1, recours\ au\ centre\ médical \\ 0, recours\ à\ l'automédication \end{cases}$
- $X_i = revenu$

NB : La question est : quelle est la probabilité que les familles possédant un certain revenu recourent au centre médical ou à l'automédication, au cas où le paludisme frappait un membre de famille.

Travail Demandé :

- Estimer le modèle Logit pour les données regroupées (3.2d), interpréter les résultats et calculer les probabilités ;
- Estimer le modèle Logit pour les données non regroupées (3.2e) et interpréter les résultats.

1) *Les données sont regroupées*

a) *Estimation du modèle 3.2d (Logit pondéré)*

Modèle (rappel) : $L_i^* = \beta_0\sqrt{W_i} + \beta_1 X_i^* + \mu_i^*$

Avec (infos, sur 10 avenues, relatives aux familles disposant d'un véhicule en fonction du revenu du ménage) :

- X_i : revenu du ménage ;
- N_i : nombre d'habitants sur une avenue ;
- n_i : nombre de familles disposant d'une voiture sur chaque avenue.

Résultats d'estimation :

```
reg Lr Wr Xr, noconstant
      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  2,     8) =    9.38
       Model |  12.8240669     2  6.41203346           Prob > F      =  0.0080
    Residual |  5.47059877     8  .683824847           R-squared     =  0.7010
-------------+------------------------------           Adj R-squared =  0.6262
       Total |  18.2946657    10  1.82946657           Root MSE      =  .82694
------------------------------------------------------------------------------
          Lr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Wr |  -.0682658   .3801634    -0.18   0.862    -.9449243    .8083926
          Xr |   .0153586   .0100544     1.53   0.165     -.007827    .0385442
------------------------------------------------------------------------------
```

Les commandes Stata relatives à la transformation de séries sont :

```
gen p=ni/Ni
gen p1=1-p
gen p2=p/p1
gen Li=ln(p2)
gen p3=p*p1
gen W=Ni*p3
gen Wr=sqrt(W)
gen Lr=Wr*Li
gen Xr=Wr*Xi
```

predict Lf, xb : Pour obtenir les valeurs estimées « Lf » (logits pondérés)

Les valeurs estimées (appelées « *logits pondérés* » parce qu'estimées sur le *modèle pondéré 3.2d*) sont consignées dans le tableau suivant :

Lf
0.2913868
0.3725826
0.4283201
0.6928309
0.9895403
0.9087647
1.295467
1.50705
1.774217
1.741995

b) *Interprétation de résultats*

Le modèle 3.2d estimé se présente comme suit :

$$L_i^* = -0.0682658\sqrt{W_i} + 0.0153586\, X_i^*$$

Sachant que $L_i = \ln\left(\frac{P_i}{1-P_i}\right)$ et $L_i^* = \sqrt{W_i}L_i$, il y a lieu de préciser ce qui suit (anti-log) :

$\frac{L_i^*}{\sqrt{W_i}} = \ln\left(\frac{P_i}{1-P_i}\right) = -0.0682658 + 0.0153586\, X_i \Rightarrow \frac{P_i}{1-P_i} = e^{-0.0682658} e^{0.0153586\, X_i}$

Autant dire que « $e^{0.0153586} = 1{,}01548$ » est le réel coefficient de la variable « X_i ». De ce fait, si X_i ↗ d'une unité, les probabilités pondérées en faveur de posséder une voiture ↗ de 1.5% $[soit: (1{,}01548 - 1) * 100]$.

c) *Calcul de probabilités*

Il est question de calculer les probabilités « $\frac{P_i}{1-P_i}$ » associées aux différents niveaux de revenu. Pour la 1ère observation, nous avons : $\ln\left(\frac{P_i}{1-P_i}\right) = 0.291387$ (*Cfr tableau de logits pondérés ci-dessus*). En passant par l'anti-log, la valeur de « p » pour cette observation est calculée comme suit :

$$\frac{P_i}{1-P_i} = e^{0.291387} \Longrightarrow (1 + e^{0.291387})P_i = e^{0.291387} \Rightarrow P_i = \frac{1{,}338282339}{2{,}338282339} = 0{,}572357$$

Procéder ainsi pour les 9 autres observations jusqu'à obtenir les probabilités suivantes :

Obs	Prob
1	**0.57233566**
2	0.591284403
3	0.605472429
4	0.666596401
5	0.728997054
6	0.712747371
7	0.785071632
8	0.818623605
9	0.854981313
10	0.850940290

2) *Les données sont non regroupées*

a) *Estimation du modèle 3.2e*

- ***Résultats sur Eviews :*** commande : `logit yi c xi`

Dependent Variable: Yi
Method : ML - Binary Logit (Quadratic hill climbing)
Date: 10/15/13 Time: 22:36
Sample: 1 40
Included observations: 40
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	3.532044	1.770789	1.994616	0.0461
Xi	-0.240169	0.112754	-2.130024	0.0332

Mean dependent var	0.425000	S.D. dependent var	0.500641
S.E. of regression	0.470578	Akaike info criterion	1.289927
Sum squared resid	8.414876	Schwarz criterion	1.374371
Log likelihood	-23.79854	Hannan-Quinn criter.	1.320459
Restr. log likelihood	-27.27418	Avg. log likelihood	-0.594964
LR statistic (1 df)	6.951283	McFadden R-squared	0.127433
Probability(LR stat)	0.008376		

Obs with Dep=0	23	Total obs	40
Obs with Dep=1	17		

- ***Résultats sur Stata :*** commande : `logit Yi Xi`

```
logit Yi Xi
Iteration 0:   log likelihood = -27.274184
Iteration 1:   log likelihood =  -24.03776
Iteration 2:   log likelihood =  -23.80572
Iteration 3:   log likelihood = -23.798552
Iteration 4:   log likelihood = -23.798543

Logistic regression                             Number of obs   =         40
                                                LR chi2(1)      =       6.95
                                                Prob > chi2     =     0.0084
Log likelihood = -23.798543                     Pseudo R2       =     0.1274

------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xi |  -.2401688   .1127545    -2.13   0.033    -.4611635   -.0191742
       _cons |   3.532044   1.770795     1.99   0.046     .0613496    7.002738
------------------------------------------------------------------------------
```

Pour les probabilités et résidus prédits, les commandes/chemins sont :

- Eviews : dans l'output, suivre :Views/Actual, Fitted, Residual/ Actual, Fitted, Residual Table ;
- Stata : probabilités : predict Yf, pr.

Pour calculer les probabilités prévues/valeurs ajustées (ex : le cas de la 5ème famille pour laquelle $X_i = 17$), se référant à l'équation estimée « $\hat{y}_i = 3{,}532044 - 0{,}240164\ X_i$ », procéder comme suit :

(i) Remplacer la valeur de *Xi* (soit 17) dans l'équation estimée et obtenir « $-z = -0{,}550744$ » (*Cfr équation 3.2a*) ;

(ii) Sachant que « $P_i = P(Z) = \frac{1}{1+e^{-Z}}$ », remplacer « z » par sa valeur et calculer la probabilité prévue « P5 » pour que la 5ème famille, disposant d'un revenu moyen de 17 Unités Monétaires/UM – quand un membre de famille est atteint du paludisme – recours à un centre médical (Y=1), comme suit :

$$P_5 = \frac{1}{1 + e^{-(0{,}550744)}} = 0{,}365691$$

NB : Procéder ainsi pour les 39 autres familles restantes pour construire le tableau reprenant les probabilités prédites.

b) *Interprétation de résultats (inférence)*

- *Significativité individuelle de paramètres* : se référer à la table de la loi normale centrée réduite ($Statistique\ de\ Wald\ :\ W = \hat{\beta}_i / \delta_{\hat{\beta}_i} \sim Z_{\alpha/2}\ , N \geq 30$), à la place de la table de student, pour tester la significativité de paramètres estimés : tous les coefficients sont statistiquement significatifs (*prob<0.05*).
- *Significativité conjointe de paramètres estimés* : La statistique du test est la ***raison de vraisemblance « LRstat »*** (comparable au *F-stat*) *qui se calcule comme suit :*

$$LRstat = -2\ ln\left(\frac{Lac}{Lsc}\right) = -2(LLac - LLsc) \sim \chi^2_{dl}$$

ou

$$LRstat = 2\ ln\left(\frac{Lsc}{Lac}\right) = 2(LLsc - LLac) \sim \chi^2_{dl}$$

Avec :

- *Lsc : Likelihood sans contrainte ;*
- *Lac : Likelihood avec contrainte (tous les paramètres sont nuls, constante exclue).*

Les hypothèses du test sont :

Ho : Non Significativité conjointe de paramètres estimés $[prob > 0.05 / LRstat < \chi^2_{dl}(tab)]$
H1 : Significativité conjointe de paramètres estimés $[prob < 0.05 / LRstat > \chi^2_{dl}(tab)]$

Dans notre modèle, la probabilité associée à $LRstat = 0.008376 < 0.05$ (d'où, les paramètres pris conjointement sont statistiquement significatifs).

- *Bonté/qualité globale de l'ajustement* : Le R^2 n'est pas efficace/approprié pour juger de la qualité de l'ajustement d'un modèle Logit. Recourir plutôt au ***pseudo R^2 de McFadden*** calculé comme suit :

$$R^2 = 1 - \frac{loglikelihood}{restr.loglikelihood}$$

Ou encore au ***comptage* R^2**, appelé autrement *Taux de Bonnes Prévisions/TBP*, dont la formule est :

$$comptage\ R^2 = \frac{\sum de\ bonnes\ prédictions}{Nombre\ d'observations}$$

NB : les observations correctes/bonnes prédictions (à gauche) et incorrectes (à droite) sont celles qui sont :

$$\begin{cases}(i)\ les\ \hat{y}_i > 0.5\ pour\ les\ y_i = 1\\(ii)\ les\ \hat{y}_i < 0.5\ pour\ les\ y_i = 0\end{cases} \qquad \begin{cases}(a)\ les\ \hat{y}_i < 0.5\ pour\ les\ y_i = 1\\(b)\ les\ \hat{y}_i > 0.5\ pour\ les\ y_i = 0\end{cases}$$

Dans notre modèle, le $comptage\ \boldsymbol{R^2} = \frac{\mathbf{27}}{\mathbf{40}} = \mathbf{0,68}$ (la prédiction est bonne > 0.50).

3.3. Modèle PROBIT

- ***Modèle Probit*** :

$$P_i = F(I) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{\frac{-Z^2}{2}} dZ \quad \ldots\ldots [3.3a]$$

Sous réserve que « *aller ou non à un centre médical est un indicateur "I_i" de revenu non observable tel que : $I_i = \beta_0 + \beta_1 X_i$* », le modèle 3.3a devient :

$$P_i = F(I) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0+\beta_1 X_i} e^{\frac{-Z^2}{2}} dZ \quad \ldots\ldots [3.3b]$$

En fait, l'on admet l'existence d'un seuil « I_i^* » tel que :

$$\begin{cases}I_i^* > I_i \Longrightarrow la\ famille\ recourt\ à\ un\ centre\ médical\\I_i^* < I_i \Longrightarrow la\ famille\ recourt\ à\ l'automédication\end{cases}$$

Si bien que :

$$P_i = P(Y = 1|X) = P(I_i^* \leq I_i) = P(Z_i \leq \beta_0 + \beta_1 X_i) = F(\beta_0 + \beta_1 X_i) \ldots . [3.3c]$$

Avec : $I_i = F^{-1}(I_i) = F^{-1}(P_i) = \beta_0 + \beta_1 X_i$ et F^{-1} : *l'inverse de la fonction de densité cumulée normale*

NB : L'indice « I_i^* » n'étant pas connu (il est supposé qu'il suit une distribution normale), le modèle Probit permet de calculer « $P(I_i^* \leq I_i)$ » à partir d'une fonction de densité normale centrée réduite (Cfr 3.3c).

Hypothèse : La fonction de répartition de l'erreur suit une loi de type normale (Modèle Probit = Fonction de densité cumulée normale).

- *Nous considérons les données ayant servi à l'estimation du modèle Logit pour estimer le modèle Probit comme suit :*

- **Travail Demandé :**

  - Estimer le *modèle Probit 3.3c* ;
  - Comparer les résultats issus du modèle Probit à ceux obtenus avec Logit.

1) *Estimation du modèle Probit 3.3c*

- *Sur Eviews* : probit Yi c Xi

```
Estimation Command:
====================
BINARY(D=N)  YI XI C

Estimation Equation:
====================
YI = 1-@CNORM(-(C(1)*XI + C(2)))
Substituted Coefficients:
====================
YI = 1-@CNORM(-(-0.1482421178*XI + 2.181726435))
```

- *Sur Stata* : `probit Yi Xi`

a) *Estimation du modèle Probit 3.3c sur Stata et Eviews*

```
Commande: probit Yi Xi
Iteration 0:   log likelihood = -27.274184
Iteration 1:   log likelihood =   -23.9185
Iteration 2:   log likelihood = -23.714322
Iteration 3:   log likelihood = -23.710557
Iteration 4:   log likelihood = -23.710555
Probit regression                                 Number of obs   =         40
                                                  LR chi2(1)      =       7.13
                                                  Prob > chi2     =     0.0076
Log likelihood = -23.710555                       Pseudo R2       =     0.1307
------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xi |  -.1482421   .0658704    -2.25   0.024    -.2773457   -.0191385
       _cons |   2.181726   1.043756     2.09   0.037      .136002    4.227451
------------------------------------------------------------------------------
```

Dependent Variable: YI
Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 40
Included observations: 40
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	2.181726	1.043754	2.090269	0.0366
XI	-0.148242	0.065870	-2.250518	0.0244

Mean dependent var	0.425000	S.D. dependent var	0.500641
S.E. of regression	0.470262	Akaike info criterion	1.285528
Sum squared resid	8.403576	Schwarz criterion	1.369972
Log likelihood	-23.71055	Hannan-Quinn criter.	1.316060
Restr. log likelihood	-27.27418	Avg. log likelihood	-0.592764
LR statistic (1 df)	7.127259	McFadden R-squared	0.130659
Probability(LR stat)	0.007592		

Obs with Dep=0	23	Total obs	40
Obs with Dep=1	17		

b) *Fréquences de la variable dépendante avec Eviews*

Dans l'output, suivre le chemin : View/Dependent Variable Frequencies.

Dependent Variable: YI
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 10/20/13 Time: 10:11
Sample: 1 40
Included observations: 40
Frequencies for dependent variable

Value	Count	Percent	Cumulative Count	Cumulative Percent
0	23	57.00	23	57.50
1	17	42.00	40	100.00

c) *Quelques statistiques descriptives de variables explicatives*

Dans l'output, suivre le chemin : View/Categorical Regressor Stats.

Dependent Variable: YI
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 10/20/13 Time: 10:11
Sample: 1 40
Included observations: 40
Descriptive statistics for explanatory variables

Variable	Mean Dep=0	Mean Dep=1	Mean All
XI	18.04348	14.52941	16.55000
C	1.000000	1.000000	1.000000

Variable	Standard Deviation Dep=0	Standard Deviation Dep=1	Standard Deviation All
XI	5.182824	2.672023	4.601839
C	0.000000	0.000000	0.000000
Observations	23	17	40

d) *Critères de sélection/information*

```
Commande: estat ic

-----------------------------------------------------------------------------
       Model |     Obs    ll(null)   ll(model)      df          AIC          BIC
-------------+---------------------------------------------------------------
           . |      40   -27.27418   -23.71056       2     51.42111     54.79887
-----------------------------------------------------------------------------
```

e) Estimation du modèle Probit 3.3c avec les effets/impacts marginaux (élasticités) et probabilités prédites sur Stata (commande : **dprobit**)

```
Commande : dprobit Yi Xi

Iteration 0:   log likelihood = -27.274184
Iteration 1:   log likelihood =   -23.9185
Iteration 2:   log likelihood = -23.714322
Iteration 3:   log likelihood = -23.710557
Iteration 4:   log likelihood = -23.710555

Probit regression, reporting marginal effects           Number of obs =     40
                                                        LR chi2(1)    =   7.13
                                                        Prob > chi2   = 0.0076
Log likelihood = -23.710555                             Pseudo R2     = 0.1307
------------------------------------------------------------------------------
      Yi |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
      Xi |  -.0569973   .0244915    -2.25   0.024     16.55     -.105 -.008995
---------+--------------------------------------------------------------------
  obs. P |       .425
 pred. P |   .3929338  (at x-bar)
------------------------------------------------------------------------------
    z and P>|z| correspond to the test of the underlying coefficient being 0
```

Autrement, après la commande probit, taper : **mfx**

```
mfx

Marginal effects after probit
      y  = Pr(Yi) (predict)
         =   .3929338
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.      z    P>|z|  [    95% C.I.   ]       X
---------+--------------------------------------------------------------------
      Xi |  -.0569973      .02449   -2.33   0.020     -.105 -.008995     16.55
------------------------------------------------------------------------------
```

- Nous constatons que quand les familles augmentent leur revenu d'1 UM en moyenne, elles perdent 0,057 point de chance à recourir à un centre médical.
- En outre, relevons qu'en moyenne, les probabilités prédites/valeurs ajustées = 0.3929338. A titre illustratif, pour obtenir la valeur ajustée relative à la 5ème observation (famille), faire : $\begin{cases} probit\ Yi\ Xi \\ matrix\ xvalue = (17) \\ dprobit\ Yi\ Xi, at(xvalue) \end{cases}$

```
probit Yi Xi
matrix xvalue=(17)
dprobit Yi Xi, at(xvalue)

Iteration 0:   log likelihood = -27.274184
Iteration 1:   log likelihood =   -23.9185
Iteration 2:   log likelihood = -23.714322
Iteration 3:   log likelihood = -23.710557
Iteration 4:   log likelihood = -23.710555

Probit regression, reporting marginal effects           Number of obs =     40
                                                        LR chi2(1)    =   7.13
                                                        Prob > chi2   = 0.0076
Log likelihood = -23.710555                             Pseudo R2     = 0.1307
------------------------------------------------------------------------------
          Yi |      dF/dx    Std. Err.      z    P>|z|         x    [    95% C.I.   ]
-------------+----------------------------------------------------------------
          Xi |  -.0558492   .0233146    -2.25   0.024        17   -.101545 -.010153
-------------+----------------------------------------------------------------
      obs. P |       .425
     pred. P |   .3929338  (at x-bar)
     pred. P |   .3675348  (at x)
------------------------------------------------------------------------------
    z and P>|z| correspond to the test of the underlying coefficient being 0
```

La valeur ajustée relative à la 5ème observation (famille) est « 0.3675348 ». Aussi, constatons que quand la 5ème famille augmente son revenu d'1 UM, elle perd « 0,056 » point de chance à recourir à un centre médical.

2) *Comparaison de résultats*

Les coefficients estimés ne sont pas directement comparables (les équations de ces deux modèles diffèrent) ; mais, de façon générale, les résultats de ces deux modèles sont presque semblables.

3) *Tableau de prédictions (bonnes et mauvaises) avec Stata et Eviews*

Commande : lstat

```
Probit model for Yi

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |         8             6  |         14
     -     |         9            17  |         26
-----------+--------------------------+-----------
   Total   |        17            23  |         40

Classified + if predicted Pr(D) >= .5
True D defined as Yi != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   47.06%
Specificity                     Pr( -|~D)   73.91%
Positive predictive value       Pr( D| +)   57.14%
Negative predictive value       Pr(~D| -)   65.38%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   26.09%
False - rate for true D         Pr( -| D)   52.94%
False + rate for classified +   Pr(~D| +)   42.86%
False - rate for classified -   Pr( D| -)   34.62%
--------------------------------------------------
Correctly classified                        62.50%
--------------------------------------------------
```

Sur Eviews, dans l'output, suivre : *View/Expectation-Prediction Table (Prediction evaluation : 0.5) ok* :

Dependent Variable: YI
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 10/20/13 Time: 10:11
Sample: 1 40
Included observations: 40
Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	17	9	26	23	17	40
P(Dep=1)>C	6	8	14	0	0	0
Total	23	17	40	23	17	40
Correct	17	8	25	23	0	23
% Correct	73.91	47.06	62.50	100.00	0.00	57.50
% Incorrect	26.09	52.94	37.50	0.00	100.00	42.50
Total Gain*	-26.09	47.06	5.00			
Percent Ga...	NA	47.06	11.76			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	14.72	8.34	23.06	13.23	9.78	23.00
E(# of Dep=1)	8.28	8.66	16.94	9.78	7.22	17.00
Total	23.00	17.00	40.00	23.00	17.00	40.00
Correct	14.72	8.66	23.39	13.23	7.22	20.45
% Correct	64.01	50.97	58.47	57.50	42.50	51.13
% Incorrect	35.99	49.03	41.53	42.50	57.50	48.88
Total Gain*	6.51	8.47	7.34			
Percent Ga...	15.33	14.73	15.03			

*Change in "% Correct" from default (constant probability) specification
**Percent of incorrect (default) prediction corrected by equation

Le taux de bonne prédiction est donc de 62,5% $[TBP = (8 + 17)/40]$: la prédiction est bonne.

4) *Qualité de l'ajustement (test de Hosmer-Lemeshow et surface ROC)*

- Les hypothèses du test d'Hosmer-Lemeshow/$HL \sim \chi^2_\alpha$ sont :

H0 : L'ajustement est bon (prob>0.5)
H1 : L'ajustement n'est pas bon (prob<0.5)

Sur EViews, dans l'output, suivre : *View/Goodness-of-Fit Test (Hosmer-Lemeshow) ok* :

Dependent Variable: YI
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 10/20/13 Time: 10:11
Sample: 1 40
Included observations: 40
Andrews and Hosmer-Lemeshow Goodness-of-Fit Tests
Grouping based upon predicted risk (randomize ties)

	Quantile of Risk		Dep=0		Dep=1		Total	H-L
	Low	High	Actual	Expect	Actual	Expect	Obs	Value
1	0.0171	0.0637	4	3.84749	0	0.15251	4	0.15856
2	0.0844	0.2628	3	3.29591	1	0.70409	4	0.15093
3	0.2628	0.3133	3	2.79747	1	1.20253	4	0.04877
4	0.3675	0.4246	2	2.47280	2	1.52720	4	0.23677
5	0.4246	0.4833	2	2.18423	2	1.81577	4	0.03423
6	0.4833	0.4833	2	2.06685	2	1.93315	4	0.00447
7	0.4833	0.5423	2	1.94874	2	2.05126	4	0.00263
8	0.5423	0.6005	1	1.65623	3	2.34377	4	0.44375
9	0.6005	0.6005	2	1.59810	2	2.40190	4	0.16832
10	0.6565	0.8403	2	1.19030	2	2.80970	4	0.78415
		Total	23	23.0581	17	16.9419	40	2.03258

H-L Statistic:	2.0326	Prob. Chi-Sq(8)	0.9800
Andrews Statistic:	6.8807	Prob. Chi-Sq(10)	0.7367

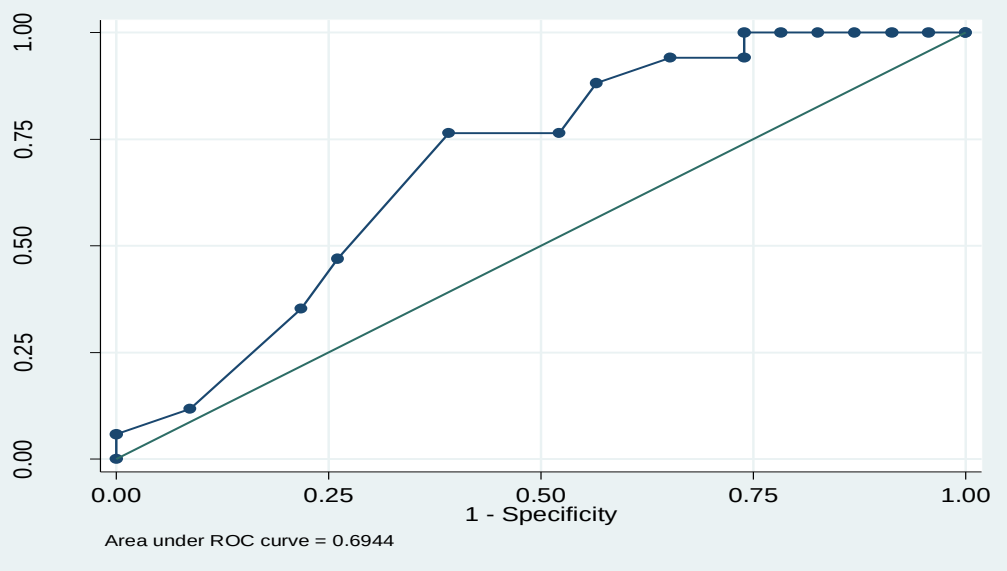
L'ajustement est bon du fait que la probabilité associée à la statistique H-L est > à 5% (soit : 0.9800). La même conclusion est tirée sous Stata (résultats ci-dessous).

***Sur Stata**, après l'estimation du modèle probit, taper :* `estat gof`

```
Probit model for YI, goodness-of-fit test

       number of observations =        40
 number of covariate patterns =        16
            Pearson chi2(14) =     10.55
                  Prob > chi2 =    0.7210
```

- *Test de bonne prédictions avec la Surface ROC (commande : `lroc`) :*

```
lroc
Probit model for Yi
number of observations =       40
area under ROC curve   =   0.6944
```



- *Test de bonnes prédictions (commande : **linktest**) :*

```
commande:linktest
Iteration 0:   log likelihood = -27.274184
Iteration 1:   log likelihood = -23.923205
Iteration 2:   log likelihood = -23.608593
Iteration 3:   log likelihood = -23.556906
Iteration 4:   log likelihood = -23.553804
Iteration 5:   log likelihood = -23.553789

Probit regression                                 Number of obs   =         40
                                                  LR chi2(2)      =       7.44
                                                  Prob > chi2     =     0.0242
Log likelihood = -23.553789                       Pseudo R2       =     0.1364

------------------------------------------------------------------------------
          Yi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   .7855453   .5850031     1.34   0.179    -.3610397     1.93213
      _hatsq |  -.3266764   .6320572    -0.52   0.605    -1.565486    .9121331
       _cons |   .0484364   .2372082     0.20   0.838    -.4164831    .5133559
------------------------------------------------------------------------------
```

Les hypothèses sont :

H0 : L'ajustement est bon (prob>0.5)
H1 : L'ajustement n'est pas bon (prob<0.5)

CHAP 4 : Modèle de données de Panel

Modèle de Panel :

Soit le modèle de Panel (fonction de production Cobb-Douglass) suivant :

$$Y_{it} = AX_{1it}^{\beta_1} X_{1it}^{\beta_1} e^{\mu_{it}} \dots\dots\dots\dots [4.1a]$$

Modèle de Panel linéaire :

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \mu_{it} \dots\dots\dots [4.1b]$$

Avec : $y_{it} = \ln Y_{it}$; $x_{1it} = \ln X_{1it}$; $x_{2it} = \ln X_{2it}$ *et* $\beta_0 = \ln A$

« i = 1, 2 et 3ème entreprise (A, B et C) et t = 1, 2, 3, 4 et 5ème année »

Travail Demandé :

- *Estimer le modèle à effets fixes (MàF) : différents cas :*
  - *Cas 1 : les coefficients β_0, β_1 et β_2 sont tous constants (modèle contraint) ;*
  - *Cas 2 : « β_0 » varie entre entreprises, tandis que « β_1 et β_2 » sont constants ;*
- *Effectuer le test d'homogénéité d'Hsiao ;*
  - *Cas 3 : « β_0 » varie avec le temps et entre entreprises, tandis que « β_1 et β_2 » sont constants ;*
  - *Cas 4 : « β_0, β_1 et β_2 » varient tous entre entreprises ;*
  - *Cas 5 : Estimateur Within-group ;*
- *Estimer le modèle à effets aléatoires (MàA) ;*
- *Effectuer le test d'Hausman.*

4.1. Estimation du modèle à effets fixes : différents cas

1) *Cas 1 : les coefficients « β_0, β_1 et β_2 » sont tous constants (modèle contraint)*

Dans ce cas, appliquer les MCO au modèle contraint suivant (sur de données empilées les unes sur les autres) :

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \mu_t \dots\dots\dots [4.1c]$$

Dependent Variable: LYT
Method: Least Squares
Date: 10/20/13 Time: 15:36
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.021442	2.664974	2.259475	0.0433
LX1T	1.004464	0.164797	6.095149	0.0001
LX2T	-1.305445	0.678869	-1.922971	0.0785

R-squared	0.758527	Mean dependent var	5.356386
Adjusted R-squared	0.718281	S.D. dependent var	0.357456
S.E. of regression	0.189728	Akaike info criterion	-0.309596
Sum squared resid	0.431960	Schwarz criterion	-0.167986
Log likelihood	5.321968	F-statistic	18.84746
Durbin-Watson stat	0.320106	Prob(F-statistic)	0.000198

- *Commandes Eviews* :

$$\begin{cases} create\ u\ 1\ 15 \\ data\ yt\ x1t\ x2t \\ gen\ Lyt = \log(yt) \\ gen\ Lx1t = \log(x1t) \\ gen\ Lx2t = \log(x2t) \\ ls\ Lyt\ c\ Lx1t\ Lx2t \end{cases}$$

- *Commentaires* :
  - *Tous les coefficients sont statistiquement significatifs, sauf celui de x2t ;*
  - *Au regard du R^2 et du F-stat, le modèle semble globalement bon. Mais la statistique de DW est très faible, ce qui amène à présumer une éventuelle autocorrélation des erreurs (d'ordre 1) ou erreur de spécification du modèle (d'où le cas 2).*

2) *Cas 2 : « β_0 » varie entre entreprises, tandis que « β_1 et β_2 » sont constants*

Dans ce cas, le « *modèle 4.1c* » devient :

$$y_t = \beta_{0i} + \beta_1 x_{1it} + \beta_2 x_{2it} + \mu_{it} \dots\dots\dots [4.1d]$$

Notons que « β_{0i} » reflète les spécificités des entreprises prises globalement. Lorsque l'on se propose de saisir les effets spécifiques propres à chaque entreprise, il sera question d'incorporer « $N-1$ » variables dummy[1] dans l'expression 4.1d. Cette dernière deviendra :

$$y_t = \alpha_0 + \alpha_1 D_{1it} + \alpha_2 D_{2it} + \beta_1 x_{1t} + \beta_2 x_{2t} + \mu_t \dots\dots\dots [4.1e] \rightarrow R^2_{NC}$$

En outre, les variables sous-étude peuvent subir communément les effets liés au temps (changement climatique, modification de politiques publiques, effet saisonnier, etc.). A cet effet, l'on peut construire jusqu'à t-1 variables temporelles – pour saisir ces effets temporels – à intégrer dans le modèle 4.1d comme suit :

$$y_t = \lambda_0 + \lambda_1 var(90) + \lambda_2 var(91) + \lambda_3 var(92) + \lambda_4 var(93) + \beta_1 x_{1t} + \beta_2 x_{2t} + \mu_t \dots\dots\dots [4.1f]$$

Avec :

$$\begin{cases} var(90) \longrightarrow \begin{cases} 1 : pour\ l'année\ 90 \\ 0 : ailleurs \end{cases} \\ var(91) \longrightarrow \begin{cases} 1 : pour\ l'année\ 91 \\ 0 : ailleurs \end{cases} \\ var(92) \longrightarrow \begin{cases} 1 : pour\ l'année\ 92 \\ 0 : ailleurs \end{cases} \\ var(93) \longrightarrow \begin{cases} 1 : pour\ l'année\ 93 \\ 0 : ailleurs \end{cases} \end{cases}$$

[1] On introduit (n-1) variables dummy – sous forme d'une matrice unitaire – pour contourner la « trappe de variables binaires.

Les résultats d'estimation du ***modèle 4.1e*** se présentent comme suit :

Dependent Variable: LYT
Method: Least Squares
Date: 10/20/13 Time: 16:25
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.421176	0.449187	9.842615	0.0000
LX1T	-0.077257	0.109050	-0.708453	0.4948
LX2T	0.319383	0.074812	4.269109	0.0016
D1I	-0.523122	0.013428	-38.95834	0.0000
D2I	0.363665	0.070582	5.152373	0.0004

R-squared	0.998530	Mean dependent var	5.356386
Adjusted R-squared	0.997943	S.D. dependent var	0.357456
S.E. of regression	0.016213	Akaike info criterion	-5.144752
Sum squared resid	0.002629	Schwarz criterion	-4.908735
Log likelihood	43.58564	F-statistic	1698.735
Durbin-Watson stat	2.124257	Prob(F-statistic)	0.000000

- *Commentaires* :
  - *Tous les coefficients sont statistiquement significatifs, sauf celui de x1t ;*
  - *Au regard du R^2 et du F-stat, le modèle est globalement bon, surtout que la statistique de DW s'est améliorée. Toutefois, le test d'Hsiao devra nous renseigner sur le modèle adéquat entre les expressions « 4.1c et 4.1e ».*

Les résultats d'estimation du ***modèle 4.1f*** se présentent comme ci-dessous (avec EViews, les variables temporelles peuvent être automatiquement générées et intégrées dans le modèle estimé : *Cfr output à droite*) :

Dependent Variable: LYT
Method: Least Squares
Date: 10/20/13 Time: 18:04
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.39264	3.413526	4.802261	0.0014
LX1T	1.238219	0.135925	9.109576	0.0000
LX2T	-3.923613	0.861495	-4.554424	0.0019
V90	-0.652273	0.187085	-3.486507	0.0082
V91	-0.497024	0.155549	-3.195290	0.0127
V92	-0.191595	0.125989	-1.520722	0.1668
V93	-0.214177	0.125280	-1.709592	0.1257

R-squared	0.913222	Mean dependent var	5.356386
Adjusted R-squared	0.848138	S.D. dependent var	0.357456
S.E. of regression	0.139299	Akaike info criterion	-0.799664
Sum squared resid	0.155234	Schwarz criterion	-0.469240
Log likelihood	12.99748	F-statistic	14.03147
Durbin-Watson stat	1.214193	Prob(F-statistic)	0.000737

Dependent Variable: LYT?
Method: Pooled Least Squares
Date: 12/22/13 Time: 17:27
Sample: 1990 1994
Included observations: 5
Cross-sections included: 3
Total pool (balanced) observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LX1T?	-0.244620	0.163350	-1.497518	0.1849
LX2T?	0.369800	0.216069	1.711489	0.1378
C--1990	4.976628	1.055262	4.716011	0.0033
C--1991	4.975221	1.067602	4.660181	0.0035
C--1992	4.953235	1.070669	4.626297	0.0036
C--1993	4.973861	1.075338	4.625392	0.0036
C--1994	4.982936	1.089576	4.573282	0.0038
Fixed Effects (Cross)				
_A--C	-0.512139			
_B--C	0.491549			
_C--C	0.020591			

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.999032	Mean dependent var	5.356386
Adjusted R-squared	0.997741	S.D. dependent var	0.357456
S.E. of regression	0.016988	Akaike info criterion	-5.028934
Sum squared resid	0.001732	Schwarz criterion	-4.604104
Log likelihood	46.71701	F-statistic	774.0870
Durbin-Watson stat	2.517729	Prob(F-statistic)	0.000000

- *Commentaires* :
  - *Tous les coefficients sont statistiquement significatifs, sauf ceux de V92 et V93 ;*
  - *Au regard du R^2 et du F-stat, le modèle apparaît globalement bon, mais la statistique de DW est faible.*

3) ***Test d'homogénéité d'Hsiao***

La statistique d'Hsiao est basée sur le F de Fisher et est calculée comme suit :

$$F = \frac{(R_{NC}^2 - R_c^2)/n - 1}{(1 - R_{NC}^2)/nT - k - 1} \sim F_{(n-1,nt-k-1)}^{\alpha}$$

Ho : Homogénéité ($F_{cal} < F_{tab}; prob > 5\%$) : $\alpha_1 = \alpha_2 = 0$ *(modèle contraint)→Rejet de la structure en panel ;*

H1 : Hétérogénéité ($F_{cal} > F_{tab}; prob < 5\%$) : $\alpha_1 \neq \alpha_2 \neq 0$ *(modèle non contraint)→la structure en panel est acceptée.*

Avec :

- R^2_{NC} : Coefficient de détermination du modèle non contraint ($R^2_{NC} = 0.998530$) ;
- R^2_c : Coefficient de détermination du modèle contraint ($R^2_{NC} = 0.758527$)

Dans notre cas :

$$F_c = \frac{(0.998530 - 0.758527)/2}{(1 - 0.758527)/(15 - 5)} = 816{,}336 > F^{0,05}_{(2,10)} = 4{,}10$$

Conclusion : Rejet de Ho→le modèle non contraint est adéquat à la structure de données sous-étude.

Calcul de paramètres structurels (intercepte) :

- Pour l'entreprise C : 4,421176 ;
- Pour l'entreprise A : 4,421176 – 0,523122 = 3,898054 ;
- Pour l'entreprise B : 4,421176 – 0,363665 = 4,7884841.

4) *Cas 3 : « β_0 » varie avec le temps et entre entreprises, tandis que « β_1 et β_2 » sont constants*

Le *modèle 4.1f* deviendra :

$$y_t = \gamma_0 + \gamma_1 var(90) + \gamma_2 var(91) + \gamma_3 var(92) + \gamma_4 var(93) + \varphi_1 D_{1it} + \varphi_2 D_{2it} + \beta_1 x_{1t} + \beta_2 x_{2t} + \mu_t \dots\dots\dots [4.1g]$$

Dependent Variable: LYT
Method: Least Squares
Date: 10/20/13 Time: 17:54
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.003527	1.061144	4.715221	0.0033
LX1T	-0.244620	0.163350	-1.497518	0.1849
LX2T	0.369800	0.216069	1.711489	0.1378
D1I	-0.532730	0.023103	-23.05874	0.0000
D2I	0.470958	0.107592	4.377259	0.0047
V90	-0.006308	0.040556	-0.155540	0.8815
V91	-0.007715	0.029961	-0.257519	0.8054
V92	-0.029701	0.023542	-1.261643	0.2539
V93	-0.009075	0.020137	-0.450694	0.6680

R-squared	0.999032	Mean dependent var	5.356386
Adjusted R-squared	0.997741	S.D. dependent var	0.357456
S.E. of regression	0.016988	Akaike info criterion	-5.028934
Sum squared resid	0.001732	Schwarz criterion	-4.604104
Log likelihood	46.71701	F-statistic	774.0870
Durbin-Watson stat	2.274687	Prob(F-statistic)	0.000000

5) *Cas 4 : « β_0, β_1 et β_2 » varient tous entre entreprises*

Le ***modèle 4.1h*** ci-dessous sera une combinaison des modèles du cas 2 et 3 :

$$y_t = \Phi_0 + \Phi_1 D_{1i} + \Phi_2 D_{2i} + \Phi_3 x_{1t} + \Phi_4 x_{2t} + \Phi_5 D_{1i} x_{1t} + \Phi_6 D_{1i} x_{2t} + \Phi_7 D_{2i} x_{1t} + \Phi_8 D_{2i} x_{2t} + \Phi_9 V_{90} + \Phi_{10} V_{91} + \Phi_{11} V_{92} + \Phi_{12} V_{93} + \mu_t \dots\dots\dots [4.1h]$$

Les résultats d'estimation de ce modèle sont :

Dependent Variable: LYT
Method: Least Squares
Date: 10/20/13 Time: 18:39
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.968816	0.616916	11.29622	0.0077
LX1T	-0.162264	0.085166	-1.905273	0.1970
LX2T	-0.178724	0.146593	-1.219182	0.3471
D1I	-0.688831	0.477123	-1.443717	0.2856
D2I	0.073445	1.305379	0.056263	0.9602
D1IX1T	-0.388581	0.148935	-2.609068	0.1208
D1IX2T	0.468434	0.102359	4.576392	0.0446
D2IX1T	0.232920	0.264608	0.880244	0.4716
D2IX2T	-0.195738	0.095862	-2.041876	0.1779
V90	-0.088810	0.025208	-3.523094	0.0720
V91	-0.056589	0.016143	-3.505531	0.0726
V92	-0.054595	0.013471	-4.052770	0.0558
V93	-0.049553	0.013092	-3.784999	0.0633

R-squared	0.999948	Mean dependent var	5.356386
Adjusted R-squared	0.999635	S.D. dependent var	0.357456
S.E. of regression	0.006826	Akaike info criterion	-7.417648
Sum squared resid	9.32E-05	Schwarz criterion	-6.804005
Log likelihood	68.63236	F-statistic	3198.927
Durbin-Watson stat	3.160072	Prob(F-statistic)	0.000313

Les commandes E-Views sont :

$$\begin{cases} genr\ D1iX1t = D1i * LX1t \\ genr\ D1iX2t = D1i * LX2t \\ genr\ D2iX1t = D2i * LX1t \\ genr\ D2iX2t = D2i * LX2t \end{cases}$$

6) *Cas 5 : Estimateur Within-group*

Le *modèle Within* est le suivant :

$$\left(\ln Y_{it} - \overline{\ln Y_i}\right) = \left(\ln A_i - \overline{\ln A_i}\right) + \beta_1\left(\ln X_{1it} - \overline{\ln X_{1i}}\right) + \beta_2\left(\ln X_{2it} - \overline{\ln X_{2i}}\right) + \left(\varepsilon_{it} - \bar{\varepsilon}_i\right) \dots\dots [4.1i]$$

Avec : $\overline{\ln Y_i}$ = La moyenne des observations pour chaque entreprise (groupe) « i ».

NB : Le R^2 d'une telle estimation renseigne sur l'efficacité du modèle, les effets individuels fixes exclus.

Posons :

« $\left(\ln Y_{it} - \overline{\ln Y_i}\right) = LYTC$; $\left(\ln X_{1it} - \overline{\ln X_{1i}}\right) = LX1TC$; $\left(\ln X_{2it} - \overline{\ln X_{2i}}\right) = LX2TC$

$(\varepsilon_{it} - \bar{\varepsilon}_i) = e$ », ce qui nous amène à écrire *l'expression 4.1i* autrement (modèle à estimer) :

$$LYTC = \beta_1 LX1TC + \beta_2 LX2TC + e \ldots\ldots [4.1j]$$

Les résultats d'estimation du ***modèle 4.1j*** sur Eviews se présentent comme ci-dessous :

Dependent Variable: LYTC
Method: Least Squares
Date: 10/20/13 Time: 21:26
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LX1TC	-0.077257	0.095643	-0.807761	0.4338
LX2TC	0.319383	0.065615	4.867533	0.0003

R-squared	0.675314	Mean dependent var	3.58E-16
Adjusted R-squared	0.650339	S.D. dependent var	0.024048
S.E. of regression	0.014220	Akaike info criterion	-5.544752
Sum squared resid	0.002629	Schwarz criterion	-5.450345
Log likelihood	43.58564	Durbin-Watson stat	2.124257

Commentaire : Le R^2 est élevé (*mesure de la variabilité intra-individuelle ou l'effet dans le groupe*) et DW est acceptable. En outre, seule *LX2TC* est significative.

4.2. *Estimation du modèle à effets aléatoires (MàA)*

- ***Modèle à effets aléatoires/MàA*** : Soient deux modèles (*toujours « log-log »*) :

$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \mu_{it} .. [4.2a]$ et $y_{it} = \beta_{0i} + \beta_1 x_{1it} + \beta_2 x_{2it} + \mu_{it} .. [4.2b]$

Avec, $\beta_{0i} = \beta_0 + \varepsilon_{it} \ldots [4.2c], i = 1; 2; 3$ (ou A ; B ; C).

Remplaçons $[4.2c]$ dans $[4.2b]$: $y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \varepsilon_{it} + \mu_{it} \ldots [4.2d]$

Posons $v_{it} = \varepsilon_{it} + \mu_{it}$; ainsi, ***4.2a devient (le MàA à estimer)*** :

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \varepsilon_{it} + \mu_{it} \ldots [4.2e]$$

- **Travail Demandé :**

  - *Estimer le modèle à effets aléatoires (MàA) sur Eviews et Stata ;*
  - *Etablir la différence entre MàF et MàA.*

1) *Estimation du Modèle à effets aléatoires (4.2e) sur Eviews et Stata*

- **Sur Eviews, suivre le chemin** : *Object/New Object…/Pool ⟶(écrire les noms des individus en colonne, précédés chacun d'une sous barre) ⟶Sheet ⟶(saisir les noms des variables suivi d'un « ? » chacun) ⟶(copier et coller les observations) ⟶Estimate ⟶(cross section=random, period=none) ⟶(Method : LS..) ⟶(Options : White Cross-section) ⟶ok.*

Malheureusement pour nous, N<T (*ou 3 entreprises < 5 années*)→l'estimation du MàA n'est pas possible avec Eviews (*la condition est : N>T*).

Sur Stata, la procédure est :

- `Saisie de données et clic sur l'onglet « preserve »`
- `tsset id yr` : déclarer la structure de données en panel à Stata ;
- `xtreg LYit LX1it LX2it, re` : estimer le MàA

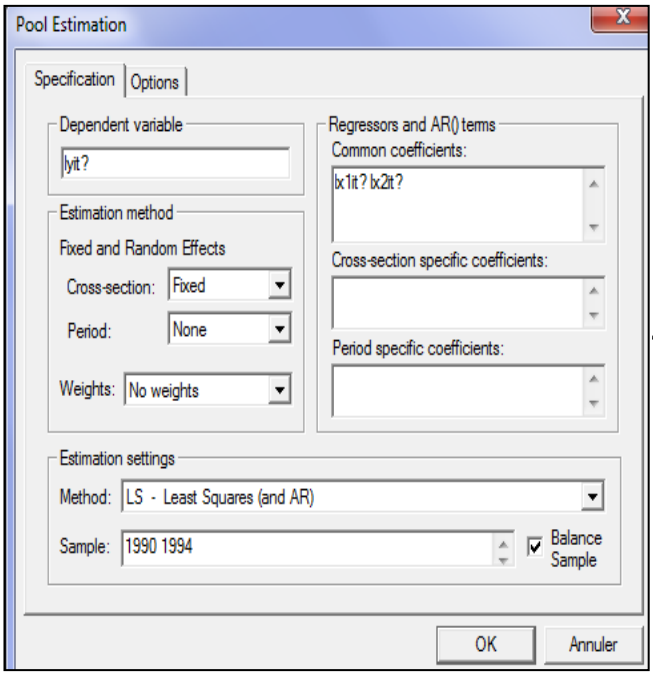
NB : *A la suite de « tsset… », Stata nous confirme la déclaration de données en panel par le résultat suivant :*

```
tsset id yr
       panel variable: id, 1 to 3
        time variable: yr, 1990 to 1994
```

Après estimation du MàA 4.2e, les résultats sont :

```
xtreg LYit LX1it LX2it, re

Random-effects GLS regression                   Number of obs      =        15
Group variable (i): id                          Number of groups   =         3

R-sq:  within  = 0.5731                         Obs per group: min =         5
       between = 0.8330                                        avg =       5.0
       overall = 0.7585                                        max =         5

Random effects u_i ~ Gaussian                   Wald chi2(2)       =     37.69
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       LYit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+----------------------------------------------------------------
      LX1it |    1.004464   .1647972     6.10   0.000     .6814672    1.327461
      LX2it |   -1.305446   .6788684    -1.92   0.054    -2.636004    .0251116
      _cons |    6.021445   2.664972     2.26   0.024     .7981955    11.24469
------------+----------------------------------------------------------------
     sigma_u |           0
     sigma_e |   .01621345
         rho |           0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

A titre d'information, voici ***les résultats d'estimation du MàF sur Eviews et Stata*** (*à gauche : capture de la procédure à suivre sur Eviews ; et à droite : résultats*).



Dependent Variable: LYIT?
Method: Pooled Least Squares
Date: 10/22/13 Time: 21:55
Sample: 1990 1994
Included observations: 5
Cross-sections included: 3
Total pool (balanced) observations: 15
White cross-section standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.368024	0.371130	11.76953	0.0000
LX1IT?	-0.077257	0.089856	-0.859780	0.4100
LX2IT?	0.319383	0.062448	5.114355	0.0005
Fixed Effects (Cross)				
_A--C	-0.469970			
_B--C	0.416817			
_C--C	0.053152			

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.998530	Mean dependent var	5.356386
Adjusted R-squared	0.997943	S.D. dependent var	0.357456
S.E. of regression	0.016213	Akaike info criterion	-5.144752
Sum squared resid	0.002629	Schwarz criterion	-4.908735
Log likelihood	43.58564	F-statistic	1698.735
Durbin-Watson stat	2.419663	Prob(F-statistic)	0.000000

```
xtreg LYit LX1it LX2it, fe
Fixed-effects (within) regression               Number of obs      =        15
Group variable (i): id                          Number of groups   =         3

R-sq:  within  = 0.6753                         Obs per group: min =         5
       between = 0.9969                                        avg =       5.0
       overall = 0.4518                                        max =         5

                                                F(2,10)            =     10.40
corr(u_i, Xb)  = -0.7126                        Prob > F           =    0.0036

------------------------------------------------------------------------------
       LYit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+----------------------------------------------------------------
      LX1it |  -.0772559   .1090501    -0.71   0.495    -.3202347    .1657228
      LX2it |   .3193816   .0748126     4.27   0.002     .1526888    .4860744
      _cons |   4.368024   .4684272     9.32   0.000     3.324303    5.411745
------------+----------------------------------------------------------------
    sigma_u |  .44577627
    sigma_e |  .01621345
        rho |  .99867888   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(2, 10) =   816.60                 Prob > F = 0.0000
```

2) ***Différence entre MàF et MàA***

a) (i) dans le MàF, l'effet propre à chaque individu est pris en compte (l'estimation se fait par les MCO), alors que (ii) dans le MàA, l'intercepte renseigne sur la moyenne de tous les termes indépendants, et l'effet individuel est aléatoire (l'estimation se fait par les MCG) ;

b) (i) dans le MàF :

- R^2 ***within*** renseigne sur l'importance de la variabilité intra-individu (ou intra-groupe) ;
- R^2 ***Between*** renseigne sur l'importance des effets fixes dans le modèle (la statistique de Hsiao/Fisher est directement calculée (elle est arrondie) après estimation : Cfr dernière ligne dans l'output de l'estimation).

(ii) dans le MàA :

- R^2 ***within*** renseigne sur l'importance de la variabilité inter-individu (ou intergroupe) ;
- R^2 ***Between*** renseigne sur l'importance des effets aléatoires dans le modèle.

Dans tous les cas, le R^2 ***overall*** renseigne sur la bonté globale de l'ajustement (ou du modèle).

c) Pour tester la significativité conjointe de paramètres, l'on recourt à la statistique de Fisher (elle est distribuée suivant la loi de Fisher) si le modèle est à effets fixes, ou à la statistique de Wald (elle est distribuée comme un χ^2) pour un modèle à effets aléatoires.

4.3. Test d'Hausman

Le test d'Hausman aide à choisir la spécification adéquate entre le MàF et MàA.

- *La statistique du test est :*

$$W = (\beta_F - \beta_A)' var[(\beta_F - \beta_A)^{-1}](\beta_F - \beta_A) \sim \chi^2_{dl=2}$$

Avec :

- $\beta_F = \beta_{MCO}$: Matrice de paramètres estimés du MàF (le MàF étant estimé par les MCO, on peut écrire « β_{MCO} ») ;
- $\beta_A = \beta_{MCG}$: Matrice de paramètres estimés du MàA (le MàA étant estimé par les MCG, on peut écrire « β_{MCG} ») ;
- $var\ \beta_A\ et\ var\ \beta_F$: Les matrices de variances-covariances du MàA et MàF, respectivement.

- *Les hypothèses du test sont :*

H_0: $\beta_F - \beta_A = 0$ *(Pas de différence entre MàF et MàA) : Retenir le MàA* $(W > \chi^2; p > 5\%)$

H_1: $\beta_F - \beta_A \neq 0$ *(Différence entre MàF et MàA) : Retenir le MàF* $(W > \chi^2; p < 5\%)$

- *Commandes/syntaxes Stata (test d'Hausman) :*

$$\begin{cases} xtreg\ LYit\ LX1it\ LX2it, re \\ est\ store\ re \\ xtreg\ LYit\ LX1it\ LX2it, fe \\ hausman\ re \end{cases}$$

- *Exécution de syntaxes→Résultats (sur Stata) :*

```
                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       re           fe         Difference          S.E.
-------------+----------------------------------------------------------------
       LX1it |    1.004464    -.0772559         1.08172        .1235565
       LX2it |   -1.305446     .3193816        -1.624828        .6747336
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       83.08
                Prob>chi2 =      0.0000
```

Décision : Rejet de Ho (p<0.05)→Le MàF est différent du MàA : d'où, le Modèle à effets fixes est approprié.

CHAP 5 : Selection de modèles

- ***Test de Mackinnon-White-Davidson (MWD)*** :

Ce test renseigne sur la meilleure spécification entre un modèle linéaire et un modèle non linéaire (log-log). Soient les deux modèles suivants :

$$Y_t = \theta_0 + \theta_1 X_t + e_t \dots\dots\dots [5.1]$$
$$\ln Y_t = \emptyset_0 + \emptyset_1 \ln X_t + e_t \dots\dots\dots [5.2]$$

Les hypothèses à tester sont dans ce cas :

Ho : Bonne spécification (modèle linéaire 5.1)
Ho : Mauvaise spécification (modèle non linéaire 5.2)

Les étapes du test sont :

(i) Obtenir « $\hat{Y}_t$ » (appelé A_t) et « $\widehat{\ln Y_t}$ » (appelé « A_t ») en estimant respectivement les modèles (5.1) et (5.2) par les MCO ;

(ii) Estimer par les MCO le modèle (5.1) en y incorporant la variable W_{1t} (soit : $W_{1t} = \ln A_t - \ln B_t$), ce qui revient à estimer :

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 W_{1t} + e_t \dots\dots\dots [5.3]$$

Et tester :

$H_0: \alpha_2 = 0$ ($|t_c| < |t_t|, prob > 5\%$) *: rejeter le modèle 5.1*
$H_1: \alpha_2 \neq 0$ ($|t_c| > |t_t|, prob < 5\%$) *: retenir le modèle 5.1*

(iii) Si le « modèle 5.1 » est rejeté, estimer le « modèle 5.2 » en y incorporant la variable « W_{2t} » (soit : $W_{2t} = B_t - antilog\ de\ A_t$), ce qui revient à estimer :

$$\ln Y_t = \beta_0 + \beta_1 \ln X_t + \beta_2 \ln W_{2t} + e_t \dots\dots\dots [5.4]$$

Et tester :

$H_0: \beta_2 = 0$ ($|t_c| < |t_t|, prob > 5\%$) *: rejeter le modèle 5.2*
$H_1: \beta_2 \neq 0$ ($|t_c| > |t_t|, prob < 5\%$) *: retenir le modèle 5.2*

- **<u>Travail Demandé</u> :** *Décider sur le modèle adéquat/meilleur entre les expressions 5.1 et 5.2 en recourant au test de Mackinnon-White-Davidson (MWD).*

<u>Solution (suivre les étapes du test)</u> :

1) Obtenir « $\hat{Y}_t$ » (appelé A_t) et « $\widehat{\ln Y_t}$ » (appelé « A_t ») en estimant respectivement les modèles (5.1) et (5.2) par les MCO. ***<u>5.1 estimé par les MCO</u>*** *:*

```
Source |       SS       df       MS                  Number of obs =      15
-------------+------------------------------          F(  1,    13) =  298.75
       Model |  52156324.7     1  52156324.7          Prob > F      =  0.0000
    Residual |  2269569.04    13  174582.234          R-squared     =  0.9583
-------------+------------------------------          Adj R-squared =  0.9551
       Total |  54425893.7    14  3887563.84          Root MSE      =  417.83

------------------------------------------------------------------------------
          Yt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xt |   99.87273   5.778212    17.28   0.000     87.38967    112.3558
       _cons |   1983.075   276.5384     7.17   0.000      1385.65      2580.5
------------------------------------------------------------------------------
```

Le modèle 5.2 estimé par les MCO :

```
Source |       SS       df       MS                  Number of obs =      15
-------------+------------------------------          F(  1,    13) =  219.91
       Model |  1.15442318     1  1.15442318           Prob > F      =  0.0000
    Residual |  .068243255    13  .005249481          R-squared     =  0.9442
-------------+------------------------------          Adj R-squared =  0.9399
       Total |  1.22266644    14  .087333317          Root MSE      =  .07245

------------------------------------------------------------------------------
         LYt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         LXt |    .655584   .0442083    14.83   0.000     .5600777    .7510902
       _cons |   6.296354   .1644735    38.28   0.000     5.941031    6.651678
------------------------------------------------------------------------------
```

2) *Estimer par les MCO le modèle 5.3 :* $Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 W_{1t} + e_t$

```
      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------          F(  2,    12) =  158.02
       Model |    52434978     2    26217489          Prob > F      =  0.0000
    Residual |  1990915.72    12  165909.644          R-squared     =  0.9634
-------------+------------------------------          Adj R-squared =  0.9573
       Total |  54425893.7    14  3887563.84          Root MSE      =  407.32

------------------------------------------------------------------------------
          Yt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xt |   94.59296   6.951727    13.61   0.000     79.44644    109.7395
         W1t |   10297.53   7945.786     1.30   0.219    -7014.849    27609.91
       _cons |   2208.182    320.695     6.89   0.000     1509.448    2906.916
------------------------------------------------------------------------------
```

Le constat est que le coefficient associé à « W1t » n'est pas statistiquement significatif (prob>0.05). Décision : rejeter le modèle 5.1 (il n'est pas meilleur/adéquat) et passer à l'étape suivante.

3) Parce que le « *modèle 5.1* » est rejeté, estimer le « modèle 5.4 » :

$$\ln Y_t = \beta_0 + \beta_1 \ln X_t + \beta_2 \ln W_{2t} + e_t$$

……(à compléter)……

CHAP 6 : Modèles dynamiques/modèles à variables décalées

6.1. Modélisation

Modèle d'Almon

Hypothèses :

(i) Le décalage est fini (sa longueur est connue);
(ii) L'évolution de paramètres est définie par un polynôme dont la forme est connue.

Modélisation :

Soit (forme quadratique sur « a_i ») :

$$a_i = \alpha_0 + i\alpha_1 + i^2\alpha_2 + \cdots + i^q\alpha_q = \sum_{j=0}^{q} \alpha_j i^j \ldots\ldots [6.1a]$$

et considérons la fonction :

$$Y = Xa + \varepsilon \ldots\ldots [6.2a]$$

Si « $a = H\alpha$ », alors $Y = XH\alpha + \varepsilon \ldots. [6.2b]$ ou $Y = Z\alpha + \varepsilon \ldots. [6.2c]$ (sous l'hypothèse que « $XH = Z$ »).

Si l'on suppose que « $\boldsymbol{q = 2}$ *et* $\boldsymbol{i = 3}$ » (*Cfr Bourbonnais, p.191*), l'expression « 6.1a » ($a = H\alpha$) peut s'écrire – sous forme matricielle – comme suit :

$$a_i = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 2^2 \\ 1 & 3 & 3^2 \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\alpha}} \ldots\ldots [6.1b]$$

Autrement, écrivons « 6.1b » – forme générale – comme suit (NB : si $i = 3 \rightarrow \alpha_{q=2}$ $(q = 3 - 1\ ou\ q = h - 1).\ càd, h > q$) :

$$\left\{ \begin{array}{l} a_0 = \alpha_0 + 0 + 0 + \cdots + 0 \\ a_1 = \alpha_0 + \alpha_1 + \alpha_2 + \cdots + \alpha_q \\ a_2 = \alpha_0 + 2\alpha_1 + 2^2\alpha_2 + \cdots + 2^q\alpha_q \\ a_3 = \alpha_0 + 3\alpha_1 + 3^2\alpha_2 + \cdots + 3^q\alpha_q \\ \vdots \\ a_h = \alpha_0 + h\alpha_1 + h^2\alpha_2 + \cdots + h^q\alpha_q \end{array} \right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6.1c]$$

Etant donné « ***6.1c*** », la fonction « ***6.2b*** » peut s'écrire (forme générale) :

$$\left\{ \begin{array}{l} a_0 : (1)\ Y_t = \gamma + \alpha_0 X_t + \varepsilon \\ a_1 : (2)\ Y_t = \gamma + \alpha_0 X_{t-1} + \alpha_1 X_{t-1} + \alpha_2 X_{t-1} + \cdots + \alpha_q X_{t-1} + \varepsilon \\ a_2 : (3)\ Y_t = \gamma + \alpha_0 X_{t-2} + 2\alpha_1 X_{t-2} + 2^2\alpha_2 X_{t-2} + \cdots + 2^q\alpha_q X_{t-2} + \varepsilon \\ a_3 : (4)\ Y_t = \gamma + \alpha_0 X_{t-3} + 3\alpha_1 X_{t-3} + 3^2\alpha_2 X_{t-3} + \cdots + 3^q\alpha_q X_{t-3} + \varepsilon \\ \vdots \\ a_h : (h)\ Y_t = \gamma + \alpha_0 X_{t-h} + h\alpha_1 X_{t-h} + h^2\alpha_2 X_{t-h} + \cdots + h^q\alpha_q X_{t-h} + \varepsilon \end{array} \right\} \ldots\ldots [6.2d]$$

Ou simplement :

$$\begin{aligned} Y_t = \gamma + \alpha_0(X_t + X_{t-1} + X_{t-2} + X_{t-3} + \cdots + X_{t-h}) + \varepsilon \\ \alpha_1(0X_t + X_{t-1} + 2X_{t-2} + 3X_{t-3} + \cdots + hX_{t-h}) + \varepsilon \\ \alpha_2(0X_t + X_{t-1} + 2^2X_{t-2} + 3^2X_{t-3} + \cdots + h^2X_{t-h}) + \varepsilon \\ \ldots\ \ldots\ \ldots \\ \alpha_q(0X_t + X_{t-1} + 2^qX_{t-2} + 3^qX_{t-3} + \cdots + h^qX_{t-h}) + \varepsilon \end{aligned} \quad \ldots\ldots [6.2e]$$

Dans « 6.2*e* », remplaçons les parenthèses par la variable « Z_{qt} », ce qui amène à réécrire cette expression comme suit :

$$\boldsymbol{Y_t = \gamma + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \cdots + \alpha_q Z_{qt} + \varepsilon} \ldots\ldots [\boldsymbol{6.2f}]$$

Avec :

$$Z_{jt} = \sum_{i=0}^{h} i^j X_{t-i} = PDL$$

NB : après estimation : $\hat{Y}_t = \hat{\gamma} + \hat{a}_0 X_t + \hat{a}_1 X_{t-1} + \hat{a}_2 X_{t-2} + .. + \hat{a}_h X_{t-h} \ldots [6.2g]$

Avec :

- $\hat{a}_0 = \hat{\alpha}_0$;
- $\hat{a}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2$;
- $\hat{a}_2 = \hat{\alpha}_0 + 2\hat{\alpha}_1 + 4\hat{\alpha}_2$;

  ……
- $\hat{a}_h = \hat{\alpha}_0 + h\hat{\alpha}_1 + h^2\hat{\alpha}_2 + \cdots + h^q\hat{\alpha}_q$

Note : *L'hypothèse fondamentale de modèles autorégressifs – au départ à retards distribués – est que le nombre de retards n'est pas connu ou est plus élevé. Parmi ces modèles, on compte ceux de KOYCK, NERLOVE et CAGAN.*

▶ *Modèle de KOYCK*

Partant de l'expression « 6.3*a* » suivante :

$$Y_t = \gamma + \alpha_0 X_t + \alpha_1 X_{t-1} + \cdots + e_t \ldots\ldots [6.3a]$$

Pour contourner les inconvénients liés à l'estimation de ce type de modèle (modèles à retards distributifs) – *soient la perte en termes de degré de liberté et d'éventuelles multi-colinéarités* – *KOYCK* formule l'hypothèse d'une décroissance géométrique des paramètres « α_j », tel que $\alpha_j = \alpha_0 \lambda^j$; ce qui lui permet de déboucher sur le modèle autorégressif[1] suivant:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + v_t \ldots\ldots [6.3b]$$

Avec : $\beta_0 = \alpha(1-\lambda)$; $\beta_1 = \alpha_0$; $\beta_2 = \lambda$; $e_t - \lambda e_{t-1} = v_t$

[1] *KOYCK a procédé comme suit : (i) étant donné* $\alpha_j = \alpha_0 \lambda^j$, *« 6.3a » devient :* $Y_t = \gamma + \alpha_0 X_t + \alpha_0 \lambda X_{t-1} + \alpha_0 \lambda^2 X_{t-2} + \cdots + e_t .. [1]$; *(ii) multiplier «* λ *» par «* Y_{t-1} *», ce qui donne :* $\lambda Y_{t-1} = \lambda\gamma + \alpha_0 \lambda X_{t-1} + \alpha_0 \lambda^2 X_{t-2} + \cdots + e_{t-1} .. [2]$; *(iii) et faire* $[1] - [2] : Y_t = \alpha(1-\lambda) + \alpha_0 X_t + \lambda Y_{t-1} + (e_t - \lambda e_{t-1})$.

Modèle de Nerlove

Ce modèle – appelé aussi « modèle d'ajustement partiel » – est, comme le modèle de *CAGAN*, un remède au modèle de *KOYCK* dont les développements sont purement algébriques. *NERLOVE et CAGAN* font plutôt des hypothèses sur le comportement des agents économiques, qui leur permettent de déboucher également sur de modèles autorégressifs d'ordre 1. En effet, NERLOVE définit :

$$Y_t^* = \alpha_0 + \alpha_1 X_t + e_t \dots\dots [6.4a]$$

Avec : Y_t^* : le niveau désiré ou attendu de la variable « Y_t ».

Il estime que, « Y_t^* » n'étant pas observable, les variations de « Y_t » sont liées à celles de « Y_t^* » comme suit :

$$Y_t - Y_{t-1} = \lambda\,(Y_t^* - Y_{t-1}) \dots\dots [6.4b]$$

Partant : $Y_t = \lambda Y_t^* + (1-\lambda) Y_{t-1} \dots\dots [6.4c]$

Avec :

- $\lambda \in [0,1]$: coefficient d'ajustement partiel ;
- $Y_t - Y_{t-1}$: variations réelles (réalisées) ;
- $Y_t^* - Y_{t-1}$: variations attendues (désirées).

« *6.4a* » dans « *6.4c* » implique :

$$Y_t = \lambda\alpha_0 + \lambda\alpha_1 X_t + (1-\lambda) Y_{t-1} + \lambda e_t \dots\dots [6.4d]$$

Posons : $\psi_0 = \lambda\alpha_0$; $\psi_1 = \lambda\alpha_1$; $\psi_2 = (1-\lambda)$ et $v_t = \lambda e_t$. Ainsi, « 6.4d » devient :

$$\boldsymbol{Y_t = \psi_0 + \psi_1 X_t + \psi_2 Y_{t-1} + v_t \dots\dots [6.4e]}$$

Modèle de CAGAN

Le modèle de CAGAN est celui qui modélise les anticipations des agents économiques, considérant que ces derniers font des anticipations adaptatives : c'est-à-dire que les agents se servent des erreurs de prévision passées pour formuler les anticipations courantes. De ce fait :

$$Y_t = \alpha_0 + \alpha_1 X_t^* + e_t \dots\dots [6.5a]$$

Avec : $X_t^* =$ le niveau attendu qui n'est pas observable. Comme chez NERLOVE, CAGAN estime que les anticipations adaptatives sur « X_t » se forment comme suit :

$$X_t^* - X_{t-1}^* = \delta(X_t - X_{t-1}^*) \dots\dots [6.5b]$$

or de « 6.5a » on peut exprimer « X_t^* » comme suit :

$$X_t^* = \frac{Y_t}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{e_t}{\alpha_1} \dots\dots [6.5c]$$

En outre, de « 6.5c », on peut écrire :

$$X_{t-1}^* = \frac{Y_{t-1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{e_{t-1}}{\alpha_1} \dots\dots [6.5d]$$

Remplaçons « X_t^* et X_{t-1}^* » par leurs expressions dans « 6.5a », ce qui correspond à :

$$\left(\frac{Y_t}{\alpha_1}-\frac{\alpha_0}{\alpha_1}-\frac{e_t}{\alpha_1}\right)-\left(\frac{Y_{t-1}}{\alpha_1}-\frac{\alpha_0}{\alpha_1}-\frac{e_{t-1}}{\alpha_1}\right)=\delta\left[X_t-\left(\frac{Y_{t-1}}{\alpha_1}-\frac{\alpha_0}{\alpha_1}-\frac{e_{t-1}}{\alpha_1}\right)\right]$$

Autrement :

$$\frac{(Y_t-Y_{t-1}+\delta Y_{t-1})+(-e_{t-1}-e_{t-1}-\delta e_{t-1})-\delta\alpha_0-\delta\alpha_1 X_t}{\alpha_1}=0$$

Après réarrangement, il s'ensuit que :

$$Y_t=\delta\alpha_0+\delta\alpha_1 X_t+(1-\delta)Y_{t-1}+[e_t-(1-\delta)e_{t-1}]\ldots\ldots[6.5e]$$

Si on pose : $\theta_0=\delta\alpha_0$; $\theta_1=\delta\alpha_1$; $\theta_2=(1-\delta)$; $v_t=[e_t-(1-\delta)e_{t-1}]$, alors « 6.5e » devient :

$$\boldsymbol{Y_t=\theta_0+\theta_1 X_t+\theta_2 Y_{t-1}+v_t\ldots\ldots[6.5f]}$$

Comme on peut le constater, les approches de KOYCK, NERLOVE et CAGAN permettent de passer d'un modèle à retards distribués à un modèle autorégressif d'ordre 1 (virtuellement identique pour tous). Cependant, la différence entre ces deux dernières approches réside dans l'interprétation de paramètres. En effet :

(i) Dans le modèle de CAGAN, on estime une relation d'équilibre ou de long terme (*équation 6.5f*). En outre – étant donné les expressions « 6.5e et 6.5f » – on note que :

$$\delta=1-\hat{\theta}_2\Rightarrow\alpha_1=\frac{\hat{\theta}_1}{\delta}=\frac{\hat{\theta}_1}{(1-\hat{\theta}_2)}$$

On dira que l'accroissement de « Y_t » correspondant à une croissance soutenue de « X_t » est de « α_1 ».

(ii) Le modèle de NERLOVE estime une relation de court-terme (équation 6.4e). Partant de l'expression (6.4d), on note que « α_1 » exprime l'élasticité de long terme de « X_t » vers « Y_t », et que l'élasticité à court-terme est traduite par le coefficient « $\lambda\alpha_1$». En effet :

$$\lambda=1-\hat{\psi}_2\Rightarrow\alpha_1=\frac{\hat{\psi}_1}{\lambda}=\frac{\hat{\psi}_1}{(1-\hat{\psi}_2)}$$

6.1. Estimation et diagnostic (applications)

6.1.1. Le modèle d'Almon

Trois étapes pour l'estimer :

- *Détermination du nombre de décalage optimal ;*
- *Détermination du degré du polynôme ;*
- *Estimation par l'OLS.*

- *Détermination du nombre de décalage optimal* : pour ce faire, l'on recourt à plusieurs critères dont les plus courants sont : le critère d'AKAIKE (AIC) et celui de SCHWARZ (SIC), à minimiser. Ci-dessous, leurs expressions :

$$AIC(h) = \ln\left(\frac{SCR(h)}{T}\right) + \frac{2h}{T} \; ; SIC(h) = \ln\left(\frac{SCR(h)}{T}\right) + \frac{h \ln T}{T}$$

Avec : T= nombre d'observations ; $i = T/4$: nombre de retards/décalages de tâtonnement (appelé aussi « *décalage maximum* ») ; SCR(h) = Somme de Carrés de Résidus pour le modèle à « *h* » décalages.

Aussi, l'on recourt à deux autres critères (à maximiser) pour retenir le modèle optimal : $le\ R^2\ ou\ R^2 - ajusté\ et\ le\ F - stat\ de\ Fisher.$

- *Commande d'estimation sur Eviews (degré du polynôme)* :

Commande : *LS Y C PDL(X, h, q, r)*

Avec :

- Y et X : les variables dépendante et explicative, respectivement ;
- PDL : Polynomial Distributed Lag ;
- h : décalage optimal ;
- q : ordre du polynôme. *NB* : $2 \leq q \leq 4$; $q < h$; $q = h - 1$ ($tâtonner\ de\ h - 1\ à\ 2$) ;
- r : les restrictions, à savoir :

$$\begin{cases} r = 0 : \text{pas de restriction} \\ r = 1 : 1^{er} \text{ coefficient du polynôme égal à 0} ; \\ r = 2 : \text{dernier coefficient du polynôme égal à 0} ; \\ r = 3 : 1^{er} \text{ et dernier coefficients du polynôme sont nuls.} \end{cases}$$

Note : (i) *Madalla et Rao (1971, cité par Bourbonnais) proposent de retenir le paramètre « r » qui maximise le « $R^2 - corrigé$ » ; (ii) la présence d'une contrainte entraîne la perte d'un degré de polynôme (en Z : Cfr équation 6.2f) ; et dans l'output : $PDLoi = \alpha_i$; et $a_i = coefficients\ en\ bas$ ($avec\ le\ Lag\ Distributed$).*

Cas pratique I *: Dépenses de publicité (DP) et Recettes de vente (RV) : approche par le modèle d'Almon (sur Eviews)*

a) Détermination du décalage optimal

$$i = T/4 = 36/4 = 9 \approx 10\ mois\ de\ tâtonnement\ (décalage\ maximum)$$

i = décalage	***AIC(h)***	***SIC(h)***	***F-stat***	***R^2***
0 : LRV=f(LDP)	1.680505	1.768479	9.579159	0.196664
1 : LRV=f(LDP_t, LDP_{t-1})	1.538048	1.671363	9.185618	0.325012
…	…	…	…	…
h=6 : LRV=f(LDP_t, LDP_{t-1},.., LDP_{t-6})	***0.901222***	***1.274874***	***9.914517***	***0.682719***
…	…	…	…	…
10 :LRV=f(LDP_t, LDP_{t-1},.., LDP_{t-10})	0.942234	1.522894	5.235875	0.650812

Le décalage optimal est donc « 6 » (min AIC et SIC ; max F-stat et R^2). Le modèle s'écrit alors :

$$LRV_t = \gamma + a_0 LDP_t + a_1 LDP_{t-1} + a_2 LDP_{t-2} + a_3 LDP_{t-3} + a_4 LDP_{t-4} + a_5 LDP_{t-5} + a_6 LDP_{t-6} + u_t$$

NB : Partir du décalage maximum (de bas en haut) pour retenir le « h^* » optimal qui maximise la « *F-stat* » (selon F-stat, $h = h^* + 1$: c.à.d. *la ligne suivante*).

Les résultats d'estimation sur Eviews sont les suivants (pour le retard 6) :

Dependent Variable: LRV
Method: Least Squares
Date: 11/02/13 Time: 22:55
Sample (adjusted): 1990M07 1992M12
Included observations: 30 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-35.79269	6.611568	-5.413646	0.0000
LDP	0.487594	0.268794	1.814008	0.0833
LDP(-1)	0.727153	0.263393	2.760719	0.0114
LDP(-2)	0.778385	0.282720	2.753204	0.0116
LDP(-3)	-0.005328	0.299840	-0.017770	0.9860
LDP(-4)	0.750772	0.329881	2.275889	0.0329
LDP(-5)	1.183883	0.329049	3.597896	0.0016
LDP(-6)	0.408567	0.319810	1.277528	0.2147

R-squared	0.759304	Mean dependent var	15.54118
Adjusted R-squared	0.682719	S.D. dependent var	0.602943
S.E. of regression	0.339624	Akaike info criterion	0.901222
Sum squared resid	2.537574	Schwarz criterion	1.274874
Log likelihood	-5.518324	F-statistic	9.914517
Durbin-Watson stat	1.099340	Prob(F-statistic)	0.000015

Commandes Eviews :

Create m 1990:01 1992:12
Data RV DP
Genr LRV=log(RV)
Genr LDP=log(DP)
LS LRV c LDP
LS LRV c LDP LDP(-1) LDP(-2) LDP(-3) LDP(-4) LDP(-5) LDP(-6)

b) *Détermination du degré du polynôme en Z et estimation*

Rappelons que : $2 \leq q \leq 4$; $q < h$; $q = h - 1$ (*tâtonner de* $h - 1$ *à* 2), ce qui nous amène aux résultats ci-après :

Degré du polynôme (q)	Nombre de contraintes/restrictions (r)	Dernier coefficient ($\alpha_i = PDLoi$)	Décision
$q = 4$	$r = 0$	-0.025013	NS
	$r = 1$	-0.012709	NS
	$r = 2$	0.000503	NS
$q = 3$	$r = 0$	-0.011738	NS
	$r = 1$	0.006186	NS
	$r = 2$	-0.011689	NS
$\boldsymbol{q = 2}$	$r = 0$	-0.008647	NS
	$\boldsymbol{r = 1}$	$\boldsymbol{-0.035462}$	**S**
	$r = 2$	-0.020031	NS

NB : NS=Non Significatif et S=Significatif.

On en déduit que notre polynôme est de degré 2 (correspondant à r=1). Reste à estimer ce polynôme avec la commande : ***LS LRV c PDL(LDP,6,2,1).*** *Ci-dessous, nous présentons les résultats de cette estimation :*

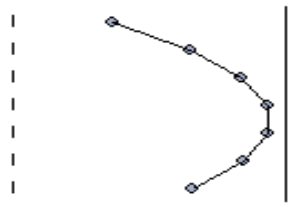
Dependent Variable: LRV
Method: Least Squares
Date: 11/03/13 Time: 00:22
Sample (adjusted): 1990M07 1992M12
Included observations: 30 after adjustments

Les « α_i » (Cfr « 6.2*f*).

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-31.87636	6.700183	-4.757535	0.0001
PDL01	0.320186	0.080572	3.973925	0.0005
PDL02	-0.035462	0.015730	-2.254345	0.0325

R-squared	0.675946	Mean dependent var	15.54118
Adjusted R-squared	0.651942	S.D. dependent var	0.602943
S.E. of regression	0.355715	Akaike info criterion	0.865265
Sum squared resid	3.416394	Schwarz criterion	1.005385
Log likelihood	-9.978979	F-statistic	28.15967
Durbin-Watson stat	1.053623	Prob(F-statistic)	0.000000

Les « a_i » (Cfr « 6.2*g* »).

Lag Distribution of LDP	i	Coefficient	Std. Error	t-Statistic
	0	0.28472	0.06546	4.34957
	1	0.49852	0.10145	4.91384
	2	0.64140	0.11048	5.80555
	3	0.71335	0.10073	7.08190
	4	0.71438	0.10087	7.08235
	5	0.64449	0.15585	4.13532
	6	0.50367	0.26538	1.89788
	Sum of Lags		4.00054	0.56486	7.08235

Soit :

$$\hat{Y}_t = \underset{\downarrow}{-31{,}87636} + \underset{\downarrow}{0{,}320186}Z_{0t} - \underset{\downarrow}{0{,}035462}Z_{1t}$$
$$\gamma \qquad \hat{a}_1 \qquad \hat{a}_2$$

Enfin, le modèle d'Almon estimé (à 6 retards) se présente comme suit (dans l'output généré par Eviews, suivre : *View/Representations*) :

$$\widehat{\boldsymbol{LRV}}_t = \mathbf{-31,87636} + \mathbf{0,28472}\boldsymbol{LDP}_t + \mathbf{0,49852}\boldsymbol{LDP}_{t-1} + \mathbf{0,64140}\boldsymbol{LDP}_{t-2} + \mathbf{0,71335}\boldsymbol{LDP}_{t-3} + \mathbf{0,71438}\boldsymbol{LDP}_{t-4} + \mathbf{0,64449}\boldsymbol{LDP}_{t-5} + \mathbf{0.50367}\boldsymbol{LDP}_{t-6}$$

Partant de paramètres estimés ci-haut, l'on peut calculer le retard moyen/RM (nombre de périodes requis en moyenne pour obtenir/ressentir le 100% des effets) et le Retard médian/RMé (nombre de périodes requis pour que 50% des effets soit ressenti) comme suit :

$$RM_Y = \frac{\sum_{i=0}^{h} i a_i}{\sum_{i=0}^{h} a_i} = 3{,}25\ mois$$

$$RMé = i + \frac{0{,}5 - \sum a_i^*}{a_m^*}, avec\ a_i^* = \frac{a_i}{\sum a_i} \Rightarrow RMé = 2 + \frac{0{,}5 - 0{,}35676}{0{,}5344 - 0{,}35676}$$

a_m^* : Paramètre normalisé.

Si l'on considère le modèle de KOYCK :

$$RM = \frac{\lambda^1}{(1-\lambda^1)} \quad et \quad RM\acute{e} = -\frac{\ln 2}{\ln \lambda^1}$$

Cas pratique II *: Estimation de la fonction d'investissement par l'approche d'Almon (sur Eviews)*

1) *Détermination du décalage optimal*

Les statistiques de AIC et SIC calculées sont consignées dans le tableau suivant (en fait, h=32/4=8 et *nombre de retards de tâtonnement=8-1=7*) :

Lag	***1***	***2***	***3***	***4***	***5***	***6***	***7***
AIC	*8.63218*	*8.60541*	*8.61335*	*8.54802*	*8.40523*	***8.194987***	*8.24391*
SIC	*8.77095*	*8.79223*	*8.84909*	*8.8335*	*8.74119*	***8.582094***	*8.68270*

Résultat : « 6 » est le décalage optimal.

Les résultats d'estimation sur Eviews sont les suivants (pour le retard 6) :

Dependent Variable: I
Method: Least Squares
Date: 11/03/13 Time: 13:00
Sample (adjusted): 1971 1996
Included observations: 26 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-187.9825	42.49238	-4.423911	0.0003
Y	0.397762	0.097417	4.083091	0.0007
Y(-1)	0.050798	0.158758	0.319971	0.7527
Y(-2)	-0.248443	0.142073	-1.748695	0.0974
Y(-3)	0.108946	0.103342	1.054219	0.3057
Y(-4)	-0.082457	0.086958	-0.948247	0.3556
Y(-5)	-0.006957	0.090495	-0.076881	0.9396
Y(-6)	0.146283	0.067186	2.177301	0.0430

R-squared	0.859420	Mean dependent var	80.47308
Adjusted R-squared	0.804750	S.D. dependent var	29.12107
S.E. of regression	12.86776	Akaike info criterion	8.194987
Sum squared resid	2980.428	Schwarz criterion	8.582094
Log likelihood	-98.53484	F-statistic	15.72012
Durbin-Watson stat	2.138390	Prob(F-statistic)	0.000002

Commandes Eviews :

Create a 1965 1996
Data I Y
LS I c Y
LS I c Y Y(-1) Y(-2) Y(-3) Y(-4) Y(-5) Y(-6)

2) *Détermination du degré du polynôme en Z et estimation*

Le nombre de décalage est « 6 » avec deux contraintes ($r = 2, c.à.d: a_1 = 0$ et $a_7 = 0$), pour un polynôme de degré « 3 » (la détermination du degré du polynôme est toutefois subjective). Le tableau ci-dessous en dit plus.

Degré du polynôme (q)	Nombre de contraintes/restrictions (r)	Dernier coefficient ($\alpha_i = PDLoi$)	Décision (à 5%)	AIC	SIC	R^2	F-Stat
$q = 4$	$r = 0$	0.007812	NS	-	-	-	-
	$r = 1$	-0.004938	NS	-	-	-	-
	$r = 2$	-0.014823	NS	-	-	-	-
$\boldsymbol{q = 3}$	$r = 0$	-0.009433	NS	-	-	-	-
	$r = 1$	0.014067	S (1%)	8.492	8.685	0.7425	21.1560
	$\boldsymbol{r = 2}$	**0.038246**	**S (1%)**	**8.1170**	**8.311**	**0.8231**	**34.1257**
$q = 2$	$r = 0$	0.038307	S (1%)	8.1797	8.373	0.81167	31.6051
	$r = 1$	-0.007350	S (1%)	9.2911	9.436	0.3820	7.10912
	$r = 2$	-0.023732	S (1%)	8.876	9.021	0.5918	16.675

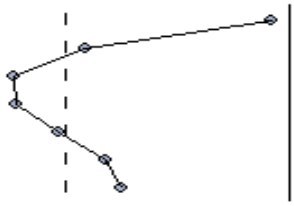
NB : NS=Non Significatif et S=Significatif.

Estimons ce polynôme avec la commande : ***LS I c PDL(Y,6,3,2).*** *Ci-dessous, nous présentons les résultats de cette estimation :*

Dependent Variable: I
Method: Least Squares
Date: 11/03/13 Time: 14:40
Sample (adjusted): 1971 1996
Included observations: 26 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-207.5100	41.59804	-4.988458	0.0001
PDL01	-0.097085	0.020775	-4.673180	0.0001
PDL02	0.055903	0.015147	3.690692	0.0013
PDL03	0.038246	0.006171	6.198205	0.0000

R-squared	0.823119	Mean dependent var	80.47308
Adjusted R-squared	0.798998	S.D. dependent var	29.12107
S.E. of regression	13.05591	Akaike info criterion	8.116996
Sum squared resid	3750.047	Schwarz criterion	8.310550
Log likelihood	-101.5210	F-statistic	34.12573
Durbin-Watson stat	2.278069	Prob(F-statistic)	0.000000

Lag Distribution of Y	i	Coefficient	Std. Error	t-Statistic
	0	0.39096	0.05149	7.59369
	1	0.03640	0.01164	3.12719
	2	-0.10320	0.02672	-3.86311
	3	-0.09709	0.02077	-4.67318
	4	-0.01447	0.00882	-1.64078
	5	0.07540	0.01942	3.88331
	6	0.10330	0.02270	4.54965
	Sum of Lags	0.39130	0.05620	6.96315

Soit :

$$\hat{Y}_t = \underset{\downarrow}{-207.51} - \underset{\downarrow}{0.097085}Z_{0t} + \underset{\downarrow}{0.055903}Z_{1t} + \underset{\downarrow}{0.038246}Z_{2t}$$
$$\gamma \qquad \hat{a}_1 \qquad \hat{a}_2 \qquad \hat{a}_3$$

Le modèle d'Almon estimé (à 6 retards) se présente comme suit (dans l'output généré par Eviews, suivre : *View/Representations*) :

```
  Estimation Command:
=====================
  LS I C PDL(Y,6,3,2)

  Estimation Equation:
=====================
  I = C(1) + C(2)*PDL01 + C(3)*PDL02 + C(4)*PDL03

  Forecasting Equation:
=====================
  I = C(1) + C(5)*Y + C(6)*Y(-1) + C(7)*Y(-2) + C(8)*Y(-3) + C(9)*Y(-4) + C(10)*Y(-5) + C(11)*Y(-6)

  Substituted Coefficients:
=====================
  I = -207.5100336 + 0.3909618822*Y + 0.03640217958*Y(-1) - 0.1032031975*Y(-2) - 0.09708524323*Y(-3)
                   - 0.01447495208*Y(-4) + 0.07539668172*Y(-5) + 0.1032986638*Y(-6)
```

6.1.2. Le modèle de CAGAN/NERLOVE

Cas pratique III *: Estimation des fonctions de consommation agrégée à court et long terme : recours aux modèles de CAGAN et de NERLOVE (sur Stata et Eviews)*

TD : Appliquer les modèles de CAGAN et de NERLOVE et commenter les résultats.

1) *Modèle de CAGAN*

- *Modélisation (consommation liée au PIB permanent/d'équilibre)* :

$$\begin{cases} C_t = \alpha_0 + \alpha_1 PIB_t^* + e_t \dots\dots\dots\dots\dots\dots\dots\dots\dots (a) \rightarrow [Cfr\ 6.5a] \\ PIB_t^* - PIB_{t-1}^* = \delta(PIB_t - PIB_{t-1}^*) \dots\dots\dots\dots (b) \rightarrow [Cfr\ 6.5b] \\ C_t = \theta_0 + \theta_1 PIB_t + \theta_2 C_{t-1} + v_t \dots\dots\dots\dots (c) \rightarrow [Cfr\ 6.5f] \\ avec : \theta_0 = \delta\alpha_0\ ;\ \theta_1 = \delta\alpha_1\ ;\ \theta_2 = (1-\delta);\ v_t = [e_t - (1-\delta)e_{t-1}] \dots\dots (d) \end{cases}$$

Avec : Ct (consommation) ; PIBt* (le PIB attendu/anticipé ou potentiel) et PIBt (*PIB permanent/de long terme ou d'équilibre*).

- *Estimation* : nous estimons le modèle (c) comme suit :

Commandes (Stata à gauche et Eviews à droite) :

```
tsset  annee
reg Ct PIBt L.Ct
```

```
Create a 1967 1987
LS Ct C PIBt Ct(-1)
```

Résultats (Eviews d'abord, et Stata ensuite) :

Dependent Variable: CT
Method: Least Squares
Date: 11/03/13 Time: 17:03
Sample (adjusted): 1968 1987
Included observations: 20 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	424.2198	754.4856	0.562264	0.5813
PIBT	0.858170	0.016583	51.75058	0.0000
CT(-1)	-0.154943	0.034211	-4.529099	0.0003

R-squared	0.999743	Mean dependent var	90924.04
Adjusted R-squared	0.999712	S.D. dependent var	175003.1
S.E. of regression	2967.555	Akaike info criterion	18.96635
Sum squared resid	1.50E+08	Schwarz criterion	19.11571
Log likelihood	-186.6635	F-statistic	33029.79
Durbin-Watson stat	1.246975	Prob(F-statistic)	0.000000

```
   Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    17) =33029.79
       Model |  5.8175e+11     2  2.9087e+11           Prob > F      =  0.0000
    Residual |   149708562    17  8806386.03           R-squared     =  0.9997
-------------+------------------------------           Adj R-squared =  0.9997
       Total |  5.8190e+11    19  3.0626e+10           Root MSE      =  2967.6

------------------------------------------------------------------------------
          Ct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PIBt |   .8581702   .0165828    51.75   0.000     .8231835    .8931569
          Ct |
          L1. |  -.1549433   .0342106    -4.53   0.000    -.2271214   -.0827652
       _cons |   424.2197   754.4856     0.56   0.581    -1167.606    2016.045
------------------------------------------------------------------------------
```

- *Commentaires* :
  - En RDC (entre 1967 et 1987), la consommation autonome est de « 424 UM » et la propension marginale à consommer est égale à « 0.86 » (si le PIB s'accroît d'1 *um*, alors la consommation varie – dans le même sens – de l'ordre de 0.82 *um*) ;
  - De (d), trouvons « δ » : $\theta_2 = (1-\delta) \rightarrow \delta = 1 + 0.154943 = 1.154943$. *Ce qui revient à dire que,* ***pour une croissance économique soutenue/durable ou stable (à long terme),*** *la propension marginale à consommer vaut :*

$$\theta_1 = \delta\alpha_1 \rightarrow \hat{\alpha}_1 = \frac{\hat{\theta}_1}{\hat{\delta}} = \frac{0.858170}{1.154943} = 0.743040998 \approx 0.74\ um$$

2) *Modèle de NERLOVE*

- *Modélisation (consommation permanente/de long terme liée au PIB)* :

$$\begin{cases} C_t^* = \alpha_0 + \alpha_1 PIB_t + e_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1) \rightarrow [Cfr\ 6.4a] \\ C_t - C_{t-1} = \lambda\,(C_t^* - C_{t-1}) \ldots\ldots\ldots\ldots\ldots (2) \rightarrow [Cfr\ 6.4b] \\ C_t = \psi_0 + \psi_1 PIB_t + \psi_2 C_{t-1} + v_t \ldots\ldots\ldots (3) \rightarrow [Cfr\ 6.4e] \\ avec : \psi_0 = \lambda\alpha_0\ ;\ \psi_1 = \lambda\alpha_1\ ;\ \psi_2 = (1-\lambda)\ et\ v_t = \lambda e_t \ldots\ldots (4) \end{cases}$$

NB : "α_1": *capte le long terme* et "ψ_1": *traduit le court terme*.

- *Estimation :*

Les résultats d'estimation du modèle de NERLOVE (*Modèle* $(3) \rightarrow [Cfr\ 6.4e]$) sont semblables à ceux obtenus avec le modèle de CAGAN (*Modèle* $(c) \rightarrow [Cfr\ 6.5f]$), seule l'interprétation des paramètres diffère.

- *Interprétations/commentaires* : Rappelons que le *modèle de NERLOVE* estime une relation de court-terme (équation 6.4e) et que – partant de l'expression (6.4d) – « α_1 » exprime l'élasticité de long terme de « PIB_t » vers « C_t », l'élasticité à court-terme se traduisant par le coefficient « $\lambda\alpha_1$ ». Ainsi :

$$\hat{\lambda} = 1 - \hat{\psi}_2 = 1.154943 \Rightarrow \hat{\alpha}_1 = \frac{\hat{\psi}_1}{\hat{\lambda}} = \frac{\hat{\psi}_1}{(1-\hat{\psi}_2)} = \frac{0.858170}{1.154943} = 0.743040998$$
$$\approx 0.74\ um$$

Ce qui revient à dire que, ***pour une croissance économique conjoncturelle (à court terme),*** *la propension marginale à consommer vaut :* $0.74\ um$.

6.1.3. Modèles de CAGAN et de NERLOVE : Méthodes d'estimation et diagnostic

Cas pratique IV : *Estimation d'un modèle autorégressif (modèle de CAGAN/NERLOVE)*

Travail Demandé :

- *Estimer le modèle (i) qui suit par les MCO/OLS :*

$$C_t = \pi_0 + \pi_1 Y_t + \pi_2 C_{t-1} + u_t \dots\dots\dots (i)$$

- *Calculer la statistique « h » de Durbin ;*
- *Estimer le modèle (i) par l'approche/méthode de variables instrumentales :*

Etapes à suivre :

- *Régresser* C_t *sur* $Y_{t-1} \rightarrow \hat{C}_t$ *;*
- *Régresser* C_t *sur* Y_t *et* $\hat{C}_{t-1}$

- *Estimer le modèle (i) par l'approche/méthode de Wallis :*

Etapes à suivre :

- *Régresser* C_t *sur* Y_t *et* Y_{t-1} *, et générer les résidus* $(\hat{u}_t)$ *;*
- *Calculer le coefficient de corrélation simple entre «* $\hat{u}_t$ *et* $\hat{u}_{t-1}$ *» qui représente «* $\hat{\rho}$ *» :*

$$r_{\hat{u}_t\hat{u}_{t-1}} = \frac{\sum \hat{u}_t\hat{u}_{t-1}/(T-1)}{\sum \hat{u}_t^2/T} + \frac{k}{T} = \hat{\rho}$$

Avec : k=nombre de paramètres et T=taille de l'échantillon.

- *Insérer la valeur calculée de «* $\hat{\rho}$ *» dans l'expression suivante :*

$$\boldsymbol{C_t - \rho C_{t-1}} = \pi_0(1-\rho) + \pi_1(\boldsymbol{Y_t - \rho Y_{t-1}}) + \pi_2(\boldsymbol{C_{t-1} - \rho C_{t-2}}) + (u_t - \rho u_{t-1})$$

- *Commenter les résultats.*

1) Estimation par les MCO

```
Commande: reg Ct Yt L.Ct

      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  2,    12) = 4787.28
       Model |  117873.554     2  58936.7769           Prob > F      =  0.0000
    Residual |   147.73338    12   12.311115           R-squared     =  0.9987
-------------+------------------------------           Adj R-squared =  0.9985
       Total |  118021.287    14  8430.09194           Root MSE      =  3.5087
------------------------------------------------------------------------------
          Ct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Yt |    .328648   .0947103     3.47   0.005     .1222919    .5350041
          Ct |
         L1. |   .6082769   .1449243     4.20   0.001      .292514    .9240399
       _cons |  -.7552286   6.016684    -0.13   0.902    -13.86446      12.354
------------------------------------------------------------------------------
```

Statistique de Durbin-Watson et test d'autocorrélation d'ordre 1.

```
Commande: estat durbin

Durbin's alternative test for autocorrelation
------------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+----------------------------------------------------------------
       1     |          0.004               1                   0.9507
------------------------------------------------------------------------------
                        H0: no serial correlation
```

```
Commande : dwstat
Durbin-Watson d-statistic(  3,    15) =  1.803214
```

2) Diagnostic sur « le modèle dynamique i » estimé : Calcul de la statistique h de Durbin

Nous insistons sur le test d'autocorrélation : Pour ce faire, dans le cas de modèles dynamiques, l'on recourt à *la statistique h de Durbin*[1] exprimée comme suit :

$$h = \left(1 - \frac{d^*}{2}\right)\sqrt{\frac{T}{1 - T[var(\hat{\pi}_2)]}} = \hat{\rho}\sqrt{\frac{T}{1 - T[var(\hat{\pi}_2)]}}$$

$$= \left(1 - \frac{1.803214}{2}\right)\sqrt{\frac{16}{1 - 16[(0.1449243)^2]}}$$

$$h = \mathbf{0.483010006}$$

Avec :

- T : Taille de l'échantillon/nombre d'observations ;
- $var(\hat{\pi}_2)$: variance du coefficient, associé à Ct-1, estimé ;
- d* : statistique de Durbin-Watson calculée ($h \sim loi\ normale$ pour un échantillon grande de taille) ; et

[1] Le test de Durbin-Watson traditionnel sous-estime le risque d'autocorrélation dans ce type de modèle : il n'est donc pas efficace.

- $\hat{\rho} = 1 - \frac{d^*}{2}$.

Les hypothèses du test sont :

Ho : absence d'autocorrélation$(h < z\,, prob > 5\%)$
H1 : Présence d'autocorrélation$(h > z\,, prob < 5\%)$

L'on notera que sous Ho, ***h*** est normalement distribuée (**z** est la valeur lue sur la table de la loi normale, pour $\alpha = 5\%$ et $z = \pm 1{,}96$), soit recourir au **t** de student.

Conclusion : *Suivant la table de la loi normale, $h < z$ (d'où, accepter Ho : absence d'autocorrélation de résidus au seuil de 5%).*

3) Estimer le modèle (i) par l'approche/méthode de variables instrumentales

▶ *1ère Etape : Régresser C_t sur $Y_{t-1} \rightarrow \hat{C}_t$*

Commande : **reg Ct L.Yt**

```
      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  1,    13) = 2082.21
       Model |  117289.007     1  117289.007           Prob > F      =  0.0000
    Residual |  732.279835    13  56.3292181           R-squared     =  0.9938
-------------+------------------------------           Adj R-squared =  0.9933
       Total |  118021.287    14  8430.09194           Root MSE      =  7.5053

------------------------------------------------------------------------------
          Ct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Yt |
         L1. |   .8185523   .0179384    45.63   0.000     .7797987    .8573059
       _cons |  -1.726778   8.133361    -0.21   0.835    -19.29784    15.84428
------------------------------------------------------------------------------
```

Après avoir régressé C_t sur Y_{t-1}, faire `predict Cf, xb` pour obtenir « $\hat{C}_t$ » (nommé « Cf »).

▶ *2ème Etape : Régresser C_t sur Y_t et $\hat{C}_{t-1}$*

Commande : **reg Ct Yt L.Cf**

```
      Source |       SS       df       MS              Number of obs =      14
-------------+------------------------------           F(  2,    11) = 2355.19
       Model |  103946.932     2  51973.4661           Prob > F      =  0.0000
    Residual |  242.744132    11  22.0676483           R-squared     =  0.9977
-------------+------------------------------           Adj R-squared =  0.9972
       Total |  104189.676    13  8014.59048           Root MSE      =  4.6976

------------------------------------------------------------------------------
          Ct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Yt |   .5311512   .0951844     5.58   0.000     .3216517    .7406506
          Cf |
         L1. |    .305554   .1460783     2.09   0.060    -.0159622    .6270702
       _cons |   6.427719   7.370421     0.87   0.402    -9.794469    22.64991
------------------------------------------------------------------------------
```

4) Estimer le modèle (i) par l'approche/méthode de Wallis

- *1ère Etape : Régresser C_t sur Y_t et Y_{t-1} , et générer les résidus ($\hat{u}_t$) ;*

```
      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  2,    12) = 2551.14
       Model |  117744.365     2  58872.1826           Prob > F      =  0.0000
    Residual |  276.921944    12  23.0768286           R-squared     =  0.9977
-------------+------------------------------           Adj R-squared =  0.9973
       Total |  118021.287    14  8430.09194           Root MSE      =  4.8038

------------------------------------------------------------------------------
          Ct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Yt |
         --. |   .5044591   .1135632     4.44   0.001     .2570262    .7518921
         L1. |   .2503492   .1284274     1.95   0.075    -.0294702    .5301685
       _cons |   12.44653   6.105837     2.04   0.064    -.8569489      25.75
------------------------------------------------------------------------------
```

- *Calculer le coefficient de corrélation simple entre « $\hat{u}_t$ et $\hat{u}_{t-1}$ » qui représente « $\hat{\rho}$ » :*

$$r_{\hat{u}_t\hat{u}_{t-1}} = \hat{\rho} = \frac{\sum \hat{u}_t\hat{u}_{t-1}/(T-1)}{\sum \hat{u}_t^2/T} + \frac{k}{T} = \frac{165.7887876/(16-1)}{276.921951/16} + \frac{2}{16} = 0.76359644$$

Avec : k=nombre de paramètres et T=taille de l'échantillon.

- *Insérer la valeur calculée de « $\hat{\rho}$ » dans l'expression suivante :*

$$\boldsymbol{C_t - \rho C_{t-1}} = \pi_0(1-\rho) + \pi_1(\boldsymbol{Y_t - \rho Y_{t-1}}) + \pi_2(\boldsymbol{C_{t-1} - \rho C_{t-2}}) + (u_t - \rho u_{t-1})$$

Et estimer :

$$\Delta \boldsymbol{C_t} = \gamma_0 + \gamma_1 \Delta \boldsymbol{Y_t} + \gamma_2 \Delta \boldsymbol{C_{t-1}} + \Delta u_t$$

Avec :

$$\Delta \boldsymbol{C_t} = \boldsymbol{C_t - \rho C_{t-1}} ; \Delta \boldsymbol{Y_t} = \boldsymbol{Y_t - \rho Y_{t-1}} ; \Delta \boldsymbol{C_{t-1}} = \boldsymbol{C_{t-1} - \rho C_{t-2}} \text{ et } \gamma_0 = \pi_0(1-\rho)$$

```
      Source |       SS       df       MS              Number of obs =      14
-------------+------------------------------           F(  2,    11) =  326.85
       Model |  10500.8274     2  5250.41372           Prob > F      =  0.0000
    Residual |  176.701514    11   16.063774           R-squared     =  0.9835
-------------+------------------------------           Adj R-squared =  0.9804
       Total |   10677.529    13  821.348382           Root MSE      =   4.008

------------------------------------------------------------------------------
         DCt1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          DYt |   .4472641   .0942496     4.75   0.001     .2398222     .654706
         DCt2 |   .4422425   .1544908     2.86   0.015     .1022106    .7822745
       _cons |  -.5953346   4.708378    -0.13   0.902    -10.9584    9.767735
------------------------------------------------------------------------------
```

5) Commenter les résultats

Précisons ce qui suit :

- Les estimateurs des variables instrumentales sont asymptotiquement consistantes et efficaces sur des grands échantillons, mais biaisés sur des petits échantillons ;

- Les estimateurs de Wallis sont consistants et asymptotiquement efficaces, mais biaisés.

Pour ce cas pratique IV, voici les commandes Stata utilisées :

```
tsset annee
reg Ct Yt L.Ct
reg Ct L.Yt
predict Cf, xb
reg Ct Yt L.Cf
```

Méthode des variables instrumentales

```
reg Ct Yt L.Yt
predict e, re
gen de=L.e
gen ede=e*de
summarize ede, detail
gen e2=e^2
gen dCt1=L.Ct
gen dYt=L.Yt
gen dCt2=L2.Ct
gen DCt1=(Ct-0.76359644*dCt1)
gen DYt=(Yt-0.76359644*dYt)
gen DCt2=(dCt1-0.76359644*dCt2)
reg DCt1 DYt DCt2
```

Méthode/Approche de Wallis

CHAP 7 : Modèles d'équations simultanées

Il s'agit d'un ensemble d'équations inter-liées formant ainsi un système. L'identificabilité des équations et, partant, le choix de la méthode d'estimation adéquate sont les deux grandes préoccupations dans ce type de modèles. Si l'on peut trouver de solutions (valeurs) uniques ou multiples de paramètres structurels à partir de coefficients réduits, on dit que le modèle (équation) est identifié (juste ou sur-identifié). Il s'en suit que la méthode des Doubles moindres Carrés (DMC) est couramment employée pour estimer un modèle juste ou sur-identifié.

7.1. L'identification

7.1.1. Notion

Identifier une équation structurelle, c'est en étudier la possibilité de trouver les paramètres. En effet, une équation peut être :

- Juste/exactement identifiée, si la solution est unique ;
- Sur-identifiée, si les solutions sont multiples ;
- Sous-identifiées, si l'équation n'est pas soluble.

7.1.2. Procédure

Définissons :

- K* : le nombre de variables exogènes (prédéterminées) présentes dans l'équation à identifier ;
- K** : le nombre de variables exogènes exclues (absentes) de l'équation à identifier ;
- K*+K**=K : le nombre total des variables exogènes dans le système ;

- G° : le nombre des variables endogènes présentes dans l'équation à identifier ;
- G°° : le nombre des variables endogènes exclues de l'équation à identifier ;
- G°+G°°=G : le nombre total des variables endogènes dans le système.

Pour une équation, l'on tient compte de certaines restrictions sur les paramètres structurels, à savoir :

- Les restrictions d'exclusion : affecter un coefficient nul aux variables endogènes n'apparaissant pas dans d'autres équations ;
- Les restrictions linéaires : en cas de combinaison linéaire entre les variables explicatives ou le fait que deux ou plusieurs variables partagent un même paramètre (paramètre identique) ;
- La normalisation : affecter le coefficient « 1 » à la variable dépendante exprimant une équation à estimer ;
- Une équation ne peut pas contenir « G » variables endogènes et « K » variables exogènes en même temps.

Deux procédures permettent d'identifier une équation, à savoir : la condition d'ordre (nécessaire) et la condition de rang (suffisante).

a) L'identification par la condition d'ordre

Il est à préciser que la condition d'ordre est nécessaire, mais non suffisante pour identifier une équation. Toutefois, si « r » correspond au nombre de restrictions dans l'équation à identifier, alors les conclusions suivantes sont à tirer :

(i) $K - K^* > G^\circ - 1 + r$: l'équation est sur-identifiée (ou $K^{**} > G^\circ - 1, si\ r = 0$) ;
(ii) $K - K^* = G^\circ - 1 + r$: l'équation est juste-identifiée ;
(iii) $K - K^* < G^\circ - 1 + r$: l'équation est sous-identifiée.

Ces conditions peuvent être respectées sans que l'équation ne soit identifiée ; d'où, il tient de recourir à la condition d'identification suffisante : *la condition de rang*.

b) L'identification par la condition de rang

Ici, il est plutôt question de comparer le rang d'une matrice (Γ) au nombre de variables endogènes dans le système moins un $(G - 1)$. En effet, la matrice « Γ » est tirée de la matrice des coefficients structurels, après avoir supprimé d'abord la ligne relative à l'équation faisant l'objet de l'identification, ensuite les colonnes des coefficients non nuls figurant sur la ligne supprimée. Ainsi, si :

- $r(\Gamma) \geq G - 1$ et que $K^{**} = G^\circ - 1$: l'équation est juste-identifiée (sous l'hypothèse que $r = 0$) ;
- $r(\Gamma) \geq G - 1$ et que $K^{**} > G^\circ - 1$: l'équation est sur-identifiée (sous l'hypothèse que $r = 0$) ;

- $r(\Gamma) \geq G - 1$ et que $K^{**} < G^{\circ} - 1$: l'équation est sous-identifiée (sous l'hypothèse que $r = 0$) ;

7.2. Méthodes d'estimation

Les méthodes d'estimation sont tributaires aux statuts (caractéristiques/natures) des équations après identification. L'on notera qu'une équation sous-identifiée n'est pas estimable, et que les MCO ne sont pas appropriées pour l'estimation des modèles multi-équationnels sous forme structurelle. Selon le statut de l'équation après identification, les méthodes d'estimation sont :

Si l'équation est juste identifiée	Si l'équation est sur-identifiée
○ *Les Moindres Carrés Indirects (MCI) ;* ○ *Les Doubles Moindres Carrés (DMC).*	○ *Les DMC ;* ○ *Les Triples Moindres Carrés (TMC).*

En outre, on distingue aussi :

- Les méthodes applicables aux équations individuelles : MCI, DMC, la méthode du Maximum de Vraisemblance à information limitée (MVIL), etc. ;
- Les méthodes applicables à l'ensemble du système : TMC, les méthodes du Maximum de vraisemblance à information Complète (MVIC), etc.

7.3. Applications (illustrations)

Cas I *: Estimation d'un modèle à équations simultanées par la méthode de MCI*

*a) **Modèle***

Soit le modèle d'équilibre sur le marché des biens et services d'une province donnée :

$$\begin{cases} Q_d = b_0 + b_1 P_t + b_2 Y_t + e_{1t} \\ Q_o = a_0 + a_1 P_t + e_{2t} \\ Q_d = Q_o \end{cases} \quad \ldots\ldots [7.1]$$

Avec :

- Q_d : quantité demandée ;
- Q_o : quantité offerte ;
- P_t : prix ;
- Y_t : revenu.

NB :
. *1ère équation* : K*=2 ; K**=0 ; G°=1 ; G°°=1.
. *2ème équation* : K*=1 ; K**=1 ; G°=2.

Travail demandé :

- Identifier les équations ;
- Estimer l'équation juste identifiée par les MCI et calculer les paramètres structurels.

b) ***Identification :***

Le « *modèle 7.1* », qui est sous sa forme structurelle, peut également se présenter sous forme réduite (matricielle) en écrivant :

$$\begin{cases} Q_d - b_0 - b_1 P_t - b_2 Y_t = e_{1t} \\ Q_o - a_0 - a_1 P_t = e_{2t} \dots\dots\dots [7.2] \\ Q_d - Q_o = 0 \end{cases}$$

Autrement (sous forme matricielle), écrivons :

$$\begin{bmatrix} Q_d & Q_o & Q \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} Q_d \\ Q_o \\ Q \end{bmatrix} + \begin{bmatrix} U & P_t & Y_t \\ -b_0 & -b_1 & -b_2 \\ -a_0 & -a_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U \\ P_t \\ Y_t \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \\ 0 \end{bmatrix} \dots\dots\dots [7.3]$$

Avec : $Q = Q_d - Q_o$.

Etant donné « l'expression 7.3 », construisons notre tableau des paramètres structurels comme suit (utile pour vérifier la condition de rang) :

Equations	***Constantes***	Q_o	Q_d	Y_t	P_t
Eq.1	$-b_0$	0	1	$-b_2$	$-b_1$
Eq.2	$-a_0$	1	0	0	$-a_1$
Eq.3	0	-1	1	0	0

Après avoir barré/supprimé l'équation à identifier et les colonnes de paramètres non nuls, il en ressort la matrice carrée :

$\Gamma = \begin{bmatrix} 1 & -b_2 \\ 1 & 0 \end{bmatrix}$ dont le rang $r(\Gamma) = 2$. Rappelons que, pour ce qui est de la deuxième équation (*Cfr forme structurelle/modèle 7.1*) : K*=1 ; K**=1 ; G°=2. Reste à comparer le rang de notre matrice carrée (Γ) au nombre de variables endogènes dans le système moins un ($G - 1 = 2 - 1 = 1$).

Conclusion : $r(\Gamma) \geq G - 1$ et que $K^{**} = G^{\circ} - 1$: la 2ème équation est juste-identifiée (sous l'hypothèse que $r = 0$).

c) ***Estimation et calcul de paramètres structurels***

▶ Estimation :

Parce que le marché est équilibré (c.à.d. : $Q_d = Q_o$), les équations de la forme réduite s'expriment comme suit :

$$\begin{cases} P_t = \left(\frac{a_0 - b_0}{b_1 - a_1}\right) - \left(\frac{b_2}{b_1 - a_1}\right) Y_t + \frac{e_{2t} - e_{1t}}{b_1 - a_1} \dots\dots\dots [7.4] \\ Q_t = \left(\frac{a_0 b_1 - a_1 b_0}{b_1 - a_1}\right) - \left(\frac{a_1 b_2}{b_1 - a_1}\right) Y_t + \frac{b_1 e_{2t} - a_1 e_{1t}}{b_1 - a_1} \dots\dots\dots [7.5] \end{cases}$$

Posons :

$\lambda_0 = \left(\frac{a_0 - b_0}{b_1 - a_1}\right)$; $\lambda_1 = \left(\frac{b_2}{b_1 - a_1}\right)$; $\lambda_2 = \left(\frac{a_0 b_1 - a_1 b_0}{b_1 - a_1}\right)$; $\lambda_3 = \left(\frac{a_1 b_2}{b_1 - a_1}\right)$; $z_{1t} = \left(\frac{e_{2t} - e_{1t}}{b_1 - a_1}\right)$ et

$$z_{2t} = \left(\frac{b_1 e_{2t} - a_1 e_{1t}}{b_1 - a_1}\right)$$

Ainsi, les équations réduites « 7.4 et 7.5 » deviennent :

$$\begin{cases} P_t = \lambda_0 - \lambda_1 Y_t + z_{1t} \dots\dots\dots [7.4]^* \\ Q_t = \lambda_2 - \lambda_3 Y_t + z_{2t} \dots\dots\dots [7.5]^* \end{cases}$$

Estimons « 7.4* et 7.5* » par les MCO. Voici les résultats de nos estimations (respectivement) :

Dependent Variable: PT
Method: Least Squares
Date: 11/10/13 Time: 14:38
Sample: 1980 1995
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1814.892	119.1216	-15.23562	0.0000
YT	0.732897	0.045843	15.98697	0.0000

R-squared	0.948068	Mean dependent var	89.31250
Adjusted R-squared	0.944359	S.D. dependent var	28.51834
S.E. of regression	6.727032	Akaike info criterion	6.766614
Sum squared resid	633.5415	Schwarz criterion	6.863187
Log likelihood	-52.13291	F-statistic	255.5832
Durbin-Watson stat	0.639244	Prob(F-statistic)	0.000000

Dependent Variable: QT
Method: Least Squares
Date: 11/10/13 Time: 14:40
Sample: 1980 1995
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1660.204	67.00880	-24.77591	0.0000
YT	0.678508	0.025788	26.31098	0.0000

R-squared	0.980177	Mean dependent var	102.6875
Adjusted R-squared	0.978762	S.D. dependent var	25.96592
S.E. of regression	3.784118	Akaike info criterion	5.615971
Sum squared resid	200.4737	Schwarz criterion	5.712545
Log likelihood	-42.92777	F-statistic	692.2679
Durbin-Watson stat	1.127630	Prob(F-statistic)	0.000000

Partant, calculons les paramètres structurels de la 2ème équation ($Q_o = a_0 + a_1 P_t + e_{2t}$) comme suit :

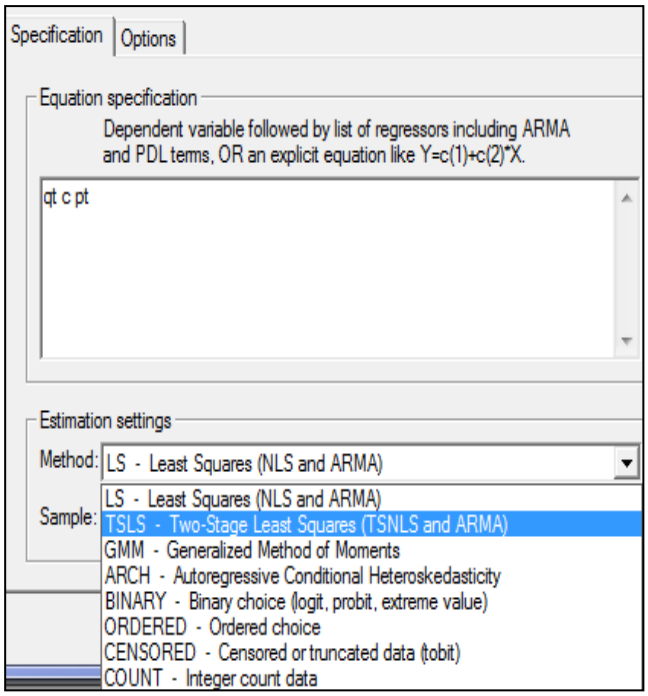
$$\hat{a}_1 = \frac{\lambda_3}{\lambda_1} = \frac{0.678508}{0.732897} = 0.925789026$$

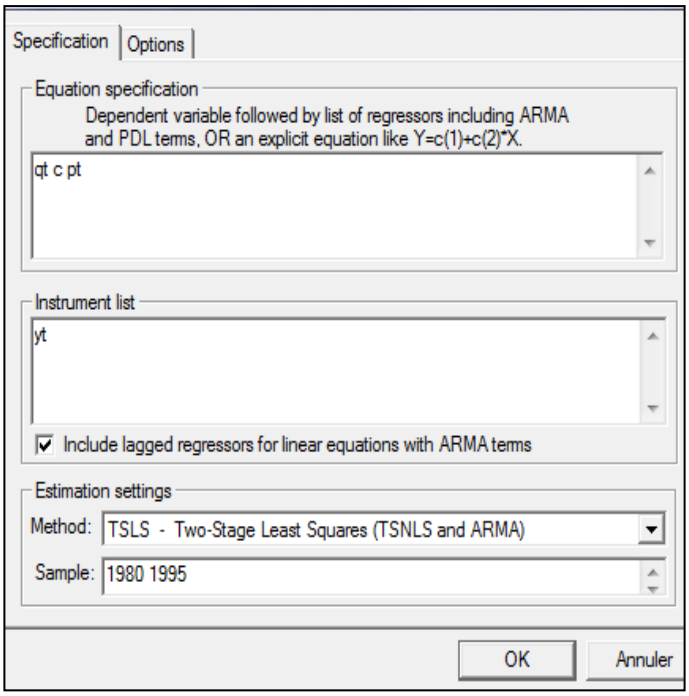
$$\hat{a}_0 = \lambda_2 - \hat{a}_1 \lambda_0 = -1660.204 - (0.925789026) * (-1814.892) = 20.00309698$$

D'où l'expression :

$$\boldsymbol{Q_o = 20.00309698 + 0.925789026 P_t} \dots\dots\dots [\boldsymbol{7.6}]$$

Les paramètres structurels estimés (tel qu'exprimés dans *l'équation 7.6* ci-dessus) sont directement obtenus si l'on recourt à la méthode des Doubles Moindres Carrés (DMC). Pour ce faire, dans *Eviews*, suivre : *Quick/Estimate Equation…(qt c pt)→Method : TSLS…Cfr les figures ci-dessous…cliquer sur ok (figure à droite). Soit, faire :* ***tsls Qt c Pt @ Yt***





*Voici les résultats d'**estimation** du modèle d'équilibre sur le marché des biens et services d'une province donnée (**modèle 7.1**) par les doubles moindres carrés/**DMC** (Sur **Eviews**) :*

Dependent Variable: QT
Method: Two-Stage Least Squares
Date: 11/10/13 Time: 15:07
Sample: 1980 1995
Included observations: 16
Instrument list: YT

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.00295	3.146997	6.356203	0.0000
PT	0.925789	0.033740	27.43881	0.0000

R-squared	0.981774	Mean dependent var	102.6875
Adjusted R-squared	0.980472	S.D. dependent var	25.96592
S.E. of regression	3.628578	Sum squared resid	184.3321
F-statistic	752.8883	Durbin-Watson stat	0.535230
Prob(F-statistic)	0.000000		

*Sur **Stata**, ces mêmes résultats se présentent plutôt comme suit (**estimation** du modèle d'équilibre sur le marché des biens et services d'une province donnée (**modèle 7.1**) par les doubles moindres carrés/**DMC**) :*

```
Commande : reg3  (eq1: Qt Pt Yt) (eq2: Qt Pt), 2sls

Two-stage least-squares regression
------------------------------------------------------------------------------
Equation          Obs  Parms        RMSE    "R-sq"     F-Stat        P
------------------------------------------------------------------------------
eq1                16      2    2.085354    0.9944    1156.31   0.0000
eq2                16      1    3.562656    0.9824     782.80   0.0000
------------------------------------------------------------------------------
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
          Pt |   .4766549     .08285     5.75   0.000     .3066609     .646649
          Yt |   .3291693   .0623614     5.28   0.000     .2012143    .4571242
       _cons |  -795.1272   154.8317    -5.14   0.000    -1112.816   -477.4387
-------------+----------------------------------------------------------------
eq2          |
          Pt |   .9024648   .0322555    27.98   0.000     .8362819    .9686476
       _cons |   22.08612   3.015363     7.32   0.000      15.8991    28.27313
------------------------------------------------------------------------------
Endogenous variables:  Qt
Exogenous variables:   Pt Yt
------------------------------------------------------------------------------
```

Cas II *: Estimation d'un modèle à équations simultanées par la méthode de DMC*

a) ***Modèle***

Soit le modèle du multiplicateur/accélérateur d'un pays donné :

$$\begin{cases} C_t = b_0 + b_1 Y_t + e_{1t} \\ I_t = a_0 + a_1 (Y_t - Y_{t-1}) + e_{2t} \ldots\ldots [7.7] \\ Y_t = C_t + I_t + G_t \end{cases}$$

Avec : a_1 = facteur de proportionnalité de l'accroissement du revenu (accélérateur).

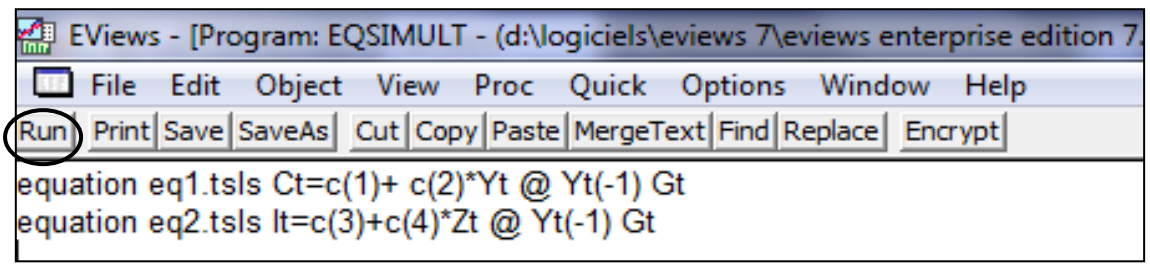
NB : *multiplicateur (mesure de l'effet d'une variation de l'investissement sur le revenu) et accélérateur (mesure l'effet sur l'investissement de l'augmentation du revenu ou de la consommation).*

b) ***Estimation***

Précisons que :

- *La 1ère équation est sur-identifiée ;*
- *La 2ème équation est juste identifiée.*

Nous estimons ces deux équations par la méthode de doubles moindres carrés/DMC (sur Eviews). Procédure : ***File/New/Program..Cfr figure ci-dessous..cliquer sur « Run » :***



Les résultats sont les suivants :

Dependent Variable: CT
Method: Two-Stage Least Squares
Date: 11/10/13 Time: 16:33
Sample (adjusted): 1968 1987
Included observations: 20 after adjustments
CT=C(1)+ C(2)*YT
Instrument list: YT(-1) GT

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	252.7037	1072.934	0.235526	0.8165
C(2)	0.783710	0.004367	179.4807	0.0000

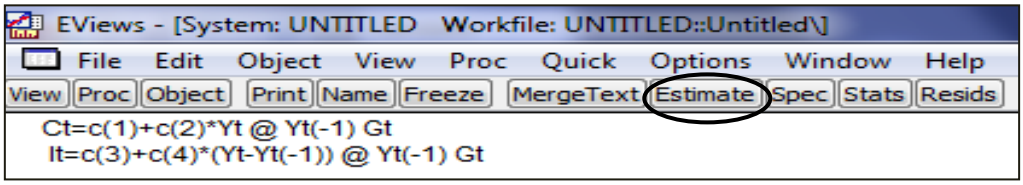
R-squared	0.999446	Mean dependent var	90969.04
Adjusted R-squared	0.999416	S.D. dependent var	175079.7
S.E. of regression	4232.542	Sum squared resid	3.22E+08
Durbin-Watson stat	1.982982		

Dependent Variable: IT
Method: Two-Stage Least Squares
Date: 11/10/13 Time: 16:33
Sample (adjusted): 1968 1987
Included observations: 20 after adjustments
IT=C(3)+C(4)*ZT
Instrument list: YT(-1) GT

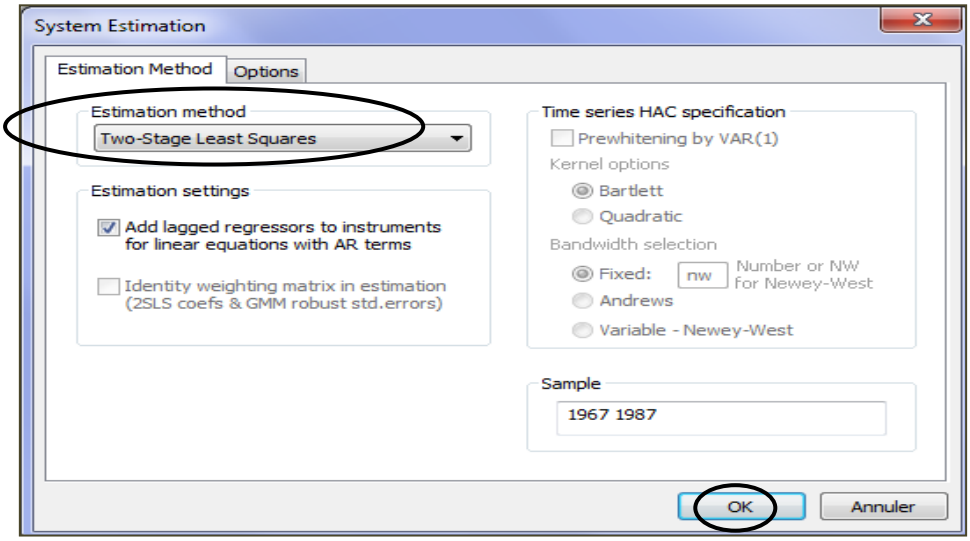
	Coefficient	Std. Error	t-Statistic	Prob.
C(3)	-1829.544	4127.696	-0.443236	0.6629
C(4)	0.660670	0.042965	15.37707	0.0000

R-squared	0.925561	Mean dependent var	26592.52
Adjusted R-squared	0.921426	S.D. dependent var	58882.71
S.E. of regression	16505.46	Sum squared resid	4.90E+09
Durbin-Watson stat	1.940915		

Une autre façon de procéder pour estimer notre système c'est de suivre : ***Object/New Object…/System→Ok→reprendre les équations comme ci-dessous :***



Ensuite, cliquer sur « **Estimate** », et choisir la méthode ***TSLS*** dans la boîte qui apparaît comme suit :



Cliquer sur « Ok » pour obtenir (*les mêmes résultats qu'avant*) :

System: UNTITLED
Estimation Method: Two-Stage Least Squares
Date: 09/07/14 Time: 12:55
Sample: 1968 1987
Included observations: 20
Total system (balanced) observations 40

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	252.7037	1072.934	0.235526	0.8151
C(2)	0.783710	0.004367	179.4807	0.0000
C(3)	-1829.544	4127.696	-0.443236	0.6602
C(4)	0.660670	0.042965	15.37707	0.0000

Determinant residual covariance 1.70E+15

Equation: CT=C(1)+C(2)*YT
Instruments: YT(-1) GT C
Observations: 20

R-squared	0.999446	Mean dependent var	90969.04
Adjusted R-squared	0.999416	S.D. dependent var	175079.7
S.E. of regression	4232.542	Sum squared resid	3.22E+08
Durbin-Watson stat	1.982982		

Equation: IT=C(3)+C(4)*(YT-YT(-1))
Instruments: YT(-1) GT C
Observations: 20

R-squared	0.925561	Mean dependent var	26592.52
Adjusted R-squared	0.921426	S.D. dependent var	58882.71
S.E. of regression	16505.46	Sum squared resid	4.90E+09
Durbin-Watson stat	1.940915		

Cas III : Estimation d'un modèle à équations simultanées par la méthode de DMC et Simulation/Scénarios

c) Modèle

Soit le modèle du multiplicateur/accélérateur d'un pays donné :

$$\begin{cases} C_t = b_0 + b_1 Y_t + e_{1t} \\ I_t = a_0 + a_1 Y_{t-1} + a_2 Y_t + e_{2t} \dots \dots [7.8] \\ Y_t = C_t + I_t + G_t \end{cases}$$

Avec : a_2 = accélérateur et Gt = dépenses gouvernementales.

NB : *multiplicateur (mesure de l'effet d'une variation de l'investissement/It sur le revenu/Yt) et accélérateur (mesure l'effet sur l'investissement de l'augmentation du revenu ou de la consommation/Ct).*

d) Identification : dans le système, *G=3 et K=2.*

Equations	Identification	Conclusion	Méthode
1ère équation	g'=2 (Ct et Yt) ; k'=0 et r=0. (G-g')+(K-k')+r ? (g'-1) (3-2)+(2-0)+0>2-1 3 > 1	La 1ère équation est sur-identifiée	Doubles Moindres Carrés (DMC)
2ème équation	g'=2 (It et Yt) ; k'=1 (Yt-1) et r=0. (G-g')+(K-k')+r ? (g'-1) (3-2)+(2-1)+0>2-1 2 > 1	La 2ème équation est sur-identifiée	Doubles Moindres Carrés/DMC

e) Estimation :

Nous estimons ces deux équations (séparément et on les enregistre dans le workfile) par la méthode de doubles moindres carrés/DMC. Sur EViews (6), taper :

```
create a 1967 1987
data Ct Gt It Yt
equation eq1.tsls Ct c Yt @ Yt(-1) Gt
equation eq2.tsls It c Yt Yt(-1) @ Yt(-1) Gt
```

Dependent Variable: CT
Method: Two-Stage Least Squares
Date: 09/07/14 Time: 13:10
Sample (adjusted): 1968 1987
Included observations: 20 after adjustments
Instrument list: YT(-1) GT

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	252.7037	1072.934	0.235526	0.8165
YT	0.783710	0.004367	179.4807	0.0000

R-squared	0.999446	Mean dependent var	90969.04
Adjusted R-squared	0.999416	S.D. dependent var	175079.7
S.E. of regression	4232.542	Sum squared resid	3.22E+08
F-statistic	32213.32	Durbin-Watson stat	1.982982
Prob(F-statistic)	0.000000	Second-Stage SSR	5.32E+09

Dependent Variable: IT
Method: Two-Stage Least Squares
Date: 09/07/14 Time: 13:10
Sample (adjusted): 1968 1987
Included observations: 20 after adjustments
Instrument list: YT(-1) GT

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3178.841	3083.016	-1.031082	0.3169
YT	0.478260	0.078715	6.075804	0.0000
YT(-1)	-0.351815	0.126089	-2.790207	0.0126

R-squared	0.961953	Mean dependent var	26592.52
Adjusted R-squared	0.957477	S.D. dependent var	58882.71
S.E. of regression	12142.27	Sum squared resid	2.51E+09
F-statistic	221.6616	Durbin-Watson stat	1.472678
Prob(F-statistic)	0.000000	Second-Stage SSR	5.15E+08

f) ***Simulations/Scénarios***

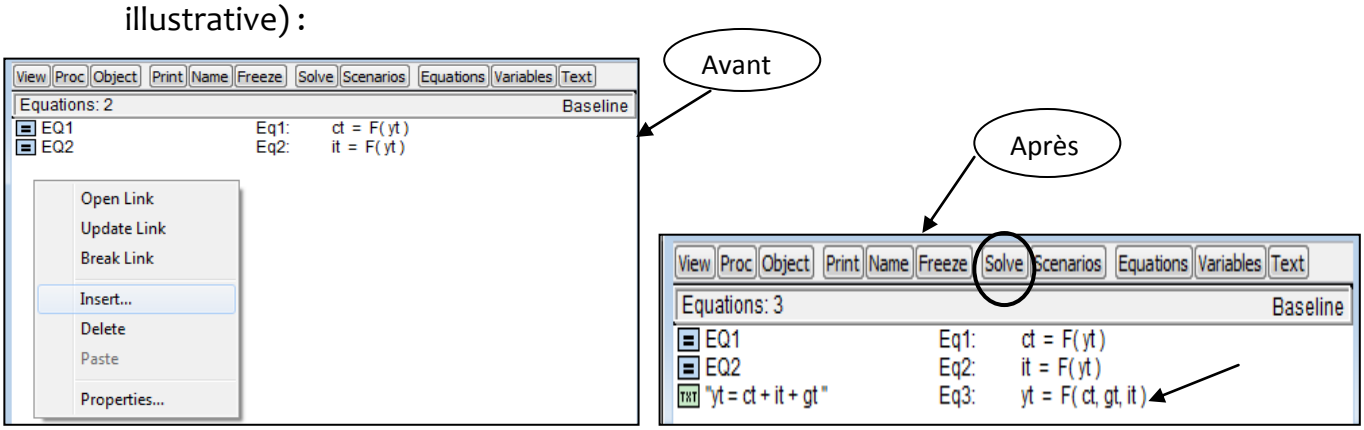
Travail demandé (*les questions sont ordonnées selon les étapes à suivre pour une simulation*) :

- Trouver la solution initiale du modèle statique ou Résoudre le modèle (obtenir le *Baseline ou solution de base*) ;
- Analyser ou élaborer le 1er Scénario : Si « Gt » augmente de « 1% », comment réagiront Ct et It ? (autrement dit, quel est l'impact dans tout le système si Gt varie de 1% positivement ?) ;
- Analyser le 2ème Scénario : Si « Yt-1 » augmente de « 2% », *toutes choses restant égales par ailleurs (c.à.d. bloquer le modèle)*, comment réagira le système (tenir compte des canaux de transmission dans l'analyse) ?
- Analyser le 3ème Scénario : Si « Gt » augmente de « 1% », l*ever l'hypothèse « ceteris paribus » (c.à.d. ne pas bloquer le modèle)*, et que « Yt-1 » augmente en même temps de 1,5%, comment réagira le système ?
- Analyser le 4ème Scénario : Si « Gt » baisse de « 3% », *toutes choses restant égales par ailleurs (c.à.d. bloquer le modèle)*, comment réagira le système ? Pour ce scénario, envisager une simulation hors échantillon (***sur 1 an, c.à.d. « 1988 »***).

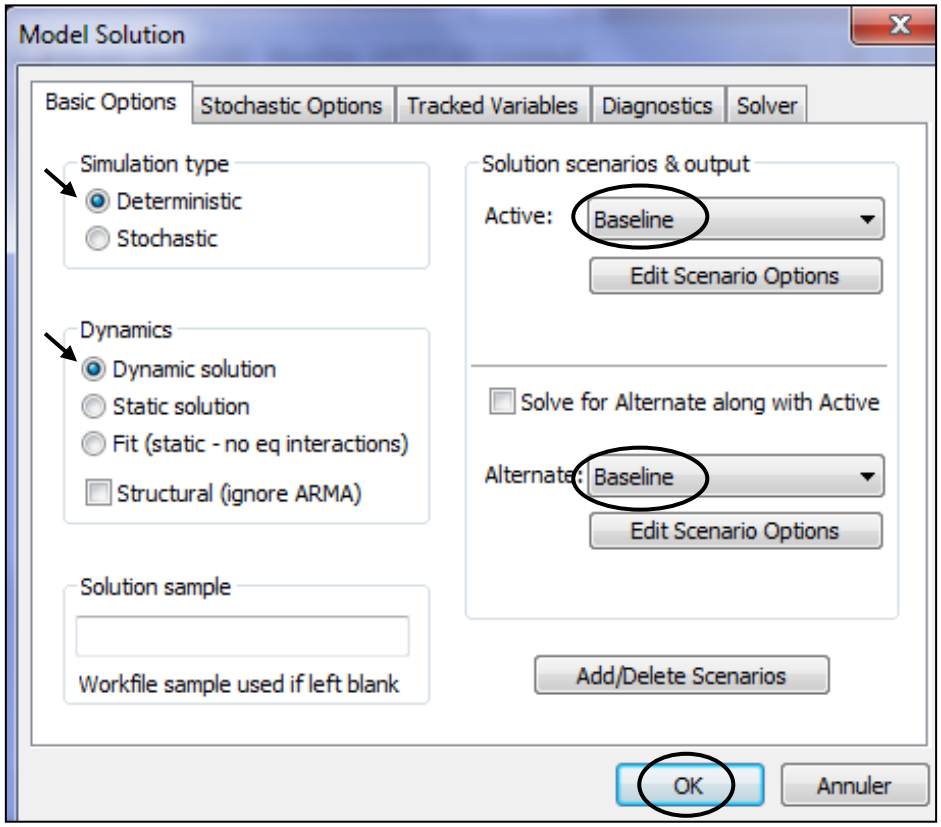
a) ***Recherche de la solution initiale (Baseline)***

Après avoir estimé nos deux équations séparément et les avoir enregistré dans le *Workfile/fichier de travail*, procéder comme suit :

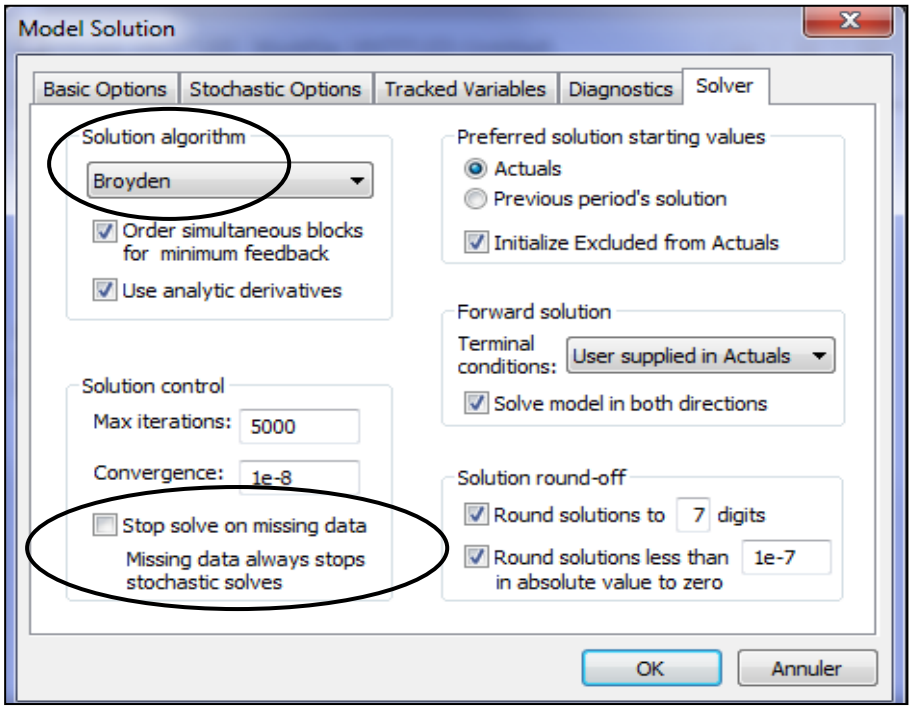
- Sélectionner les équations enregistrées dans le Workfile→Clic droit→Open as Model→Ok→un espace s'ouvre (nous l'appelons « **Boîte A** ») ;
- Dans l'espace « Boite A », faire clic droit→insert (pour insérer la relation d'équilibre) : Yt=Ct+It+Gt→ok (la boîte de dialogue A ci-dessous est illustrative) :



- Dans la boîte de dialogue ci-haut (à droite), cliquer sur « **Solve** » pour résoudre le modèle (simulation statique ou solution de base/Baseline)→la boîte de dialogue ci-dessous apparaît :

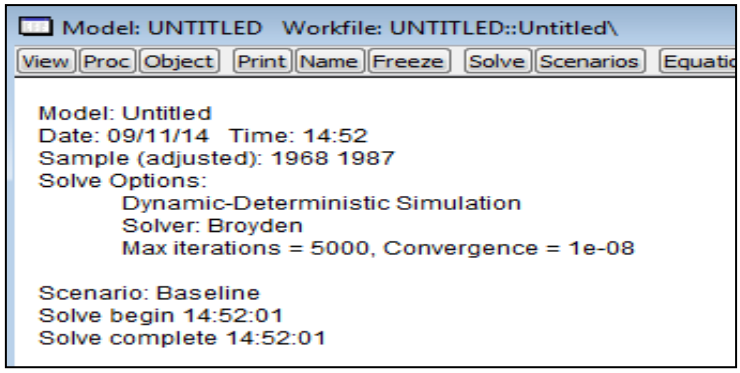


Cliquer sur l'onglet « **Solver** » pour préciser ce qui suit (***à décocher***) :



Ensuite, cliquer sur « **Ok** » pour obtenir la solution de base. En fait, les solutions de base sont des variables générées par EViews avec des indices « _o » ; dans notre cas, ces variables sont : **Ct_o** , **It_o** et **Yt_o**. L'on notera que le modèle est soluble (équations *juste-identifiées ou sur-identifiées*) s'il passe le test de convergence avec succès (*assurance pour la solution mathématique du modèle*). Donc, le message

suivant devra s'afficher (*il indique le nombre d'itérations, le type de simulation et l'algorithme utilisés, aussi les dates de début et de fin d'itérations/résolution*) :



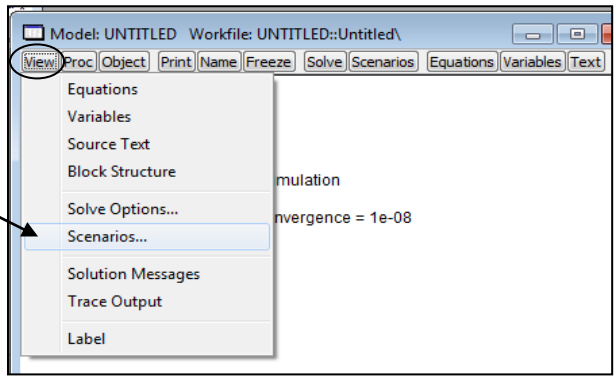
b) ***Elaboration du 1er Scénario et analyses : si « Gt augmente de 1%, ceteris paribus »***

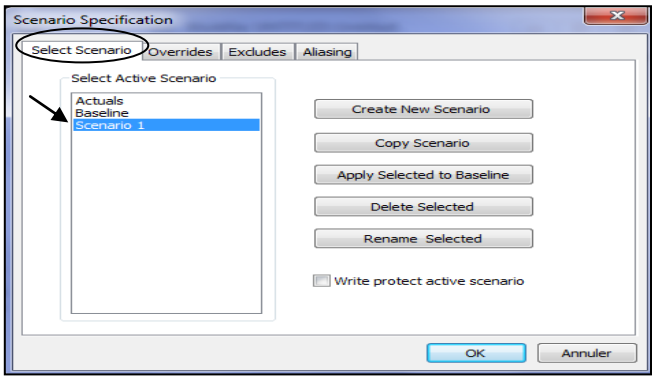
_____***Schéma*** : Selon le modèle, Gt→Yt→Ct et Gt→Yt→It

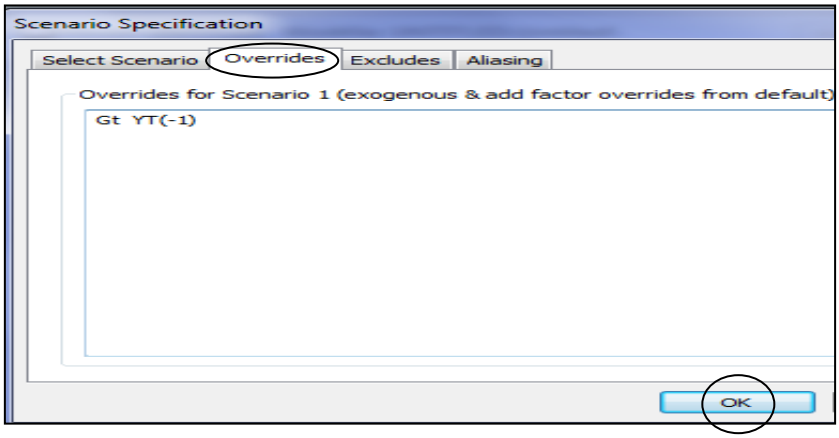
_____***Note*** : Bouger une variable exogène (Gt), ceteris paribus signifie ignorer les effets d'autres variables exogènes (Yt-1) ou bloquer les autres variables exogènes (Yt-1 considérée constante dont la dérivée est nulle).

_____***Etapes à suivre*** :

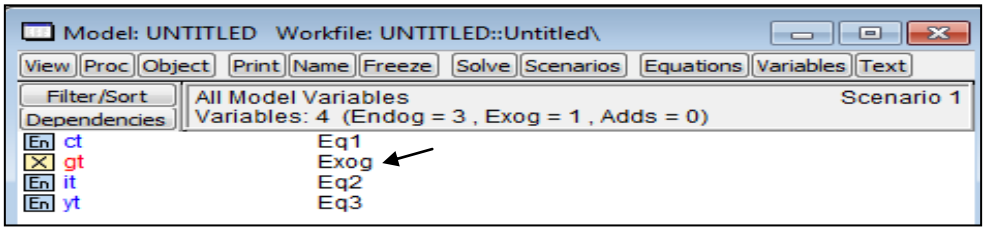
- Dans « **Boîte A** », procéder comme suit (suivre) : *View/Scenarios…→**Scenario 1**→dans « scenario Overrids », préciser les variables exogènes du modèle : Gt et Yt-1→cliquer sur Ok.* Les figures ci-dessous sont illustratives :







- Dans « **Boîte A** », cliquer sur l'onglet « **Variables** » pour vérifier que l'on ne s'est pas trompé dans la déclaration des variables exogènes. Les informations suivantes s'affichent :

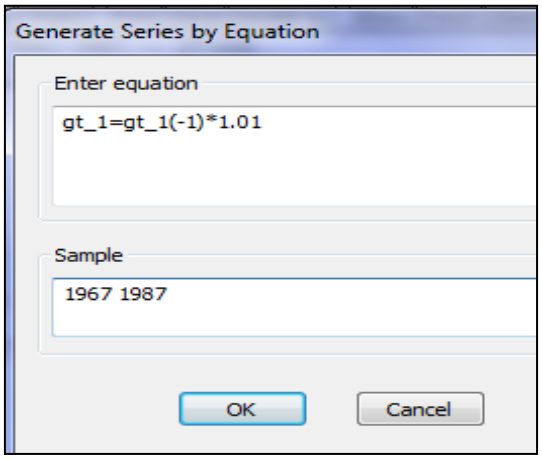


En rouge, les variables exogènes déclarées, et en bleu, les variables endogènes du modèle. Dans notre cas, « Gt » seulement apparaît comme exogène, alors que « Yt-1 » a été également déclarée exogène. *L'on notera ainsi que les variables décalées n'apparaissent pas dans la boîte A, à moins de les générer sous un autre nom à des fins de simulation.*

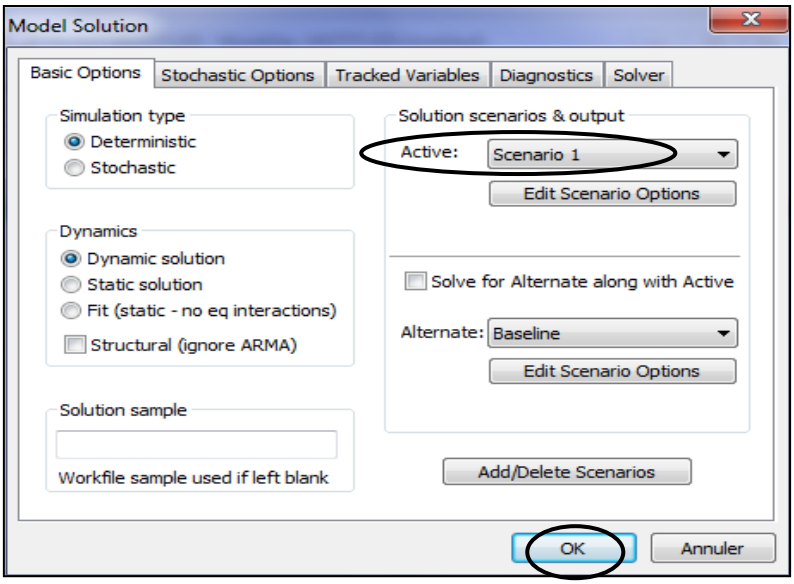
- ***Simulation (Gt ↑ 1% en 1987)*** :
  - La simulation est similaire à la prévision dans l'échantillon (*bloquer est nécessaire, mais l'on peut ouvrir/lever l'hypothèse ceteris paribus*) comme en dehors de l'échantillon (*blocage automatique, soit toujours bloquer*). *Ici, nous faisons une simulation dans l'échantillon (1987)* ;
  - L'on devra bouger « Gt » et bloquer « Yt-1 » pour voir la réaction du modèle (toutes les variables endogènes) ;
  - Deux possibilités pour simuler : rester sur la barre de commande ou aller dans : *Quik/Generate Series…* →Dans la boîte de dialogue qui s'affiche, saisir à tour de rôle (cliquer sur « ok » et reprendre le processus pour donner d'autres instructions), *avec « Gt_1 » traduisant la variable du scénario 1* :

gt_1=gt → sample : 1967 1987 → ok : *définir le 1er Scénario*
gt_1=gt_1(-1)*1.01 → sample : 1987 1987 → ok : *Gt ↑ 1% en 1987 (ouvrir « Gt »)*
ylag_1=yt(-1) → sample : 1967 1987 → ok : *définir le 1er Scénario (bloquer « Yt-1 »)*
ylag_1=@elem(ylag_1, 1986) → sample : 1987 1987 → ok : *bloquer « Yt-1 » (données 1986=1987 : Constance).*

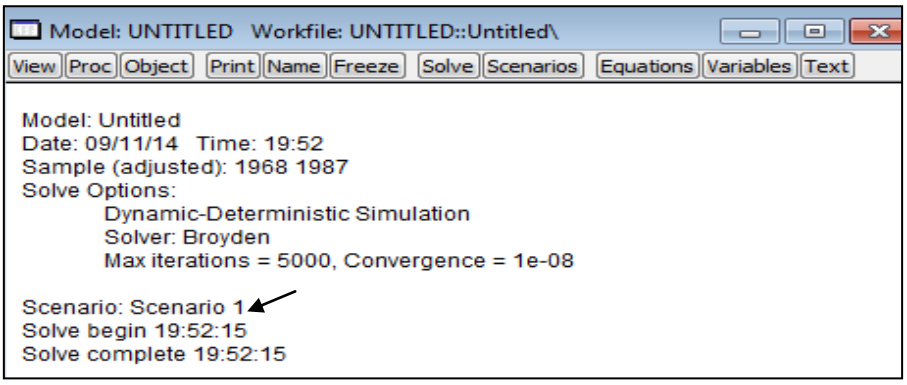
A titre illustratif, observons la boîte de dialogue ci-dessous :



- Revenir à la « Boîte A », cliquer sur l'onglet « **Solve** »→[Active : ***Scenario 1***], la boîte de dialogue ci-après complète la procédure[1] :



Résultats : Après avoir cliqué sur « Ok » (Cfr figure ci-dessus), le message suivant s'affiche (*indiquant que la simulation a réussi*)



Et nous pouvons observer ce qui suit (*Cfr données sur Workfile*) :

Période	Gt_1	Yt_1	It_1	Ct_1
1986	38.600	7.398.695	1.561.413	5.798.683
1987	38.986	9.798.528	2.080.087	7.679.455
%	1%	32.4%	33.2%	32.4%

Analyses :

[1] La simulation est différente du Scénario du fait que celui-ci cherche à saisir l'évolution d'une variable compatible avec une trajectoire imposée à d'autres variables.

c) ***Elaboration du 2ème Scénario et Analyses : « Yt-1 » augmente de « 2% », ceteris paribus***

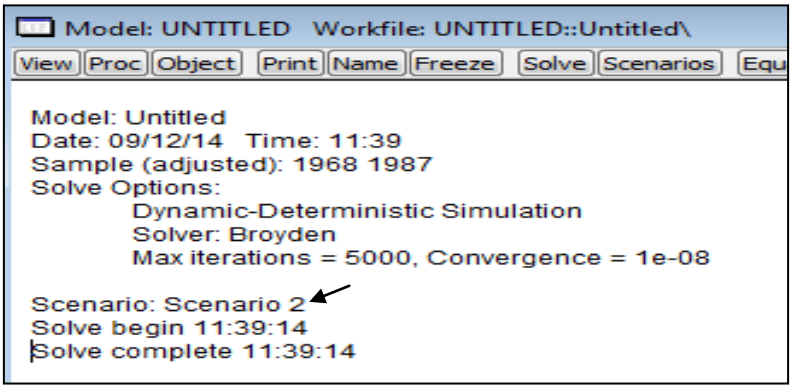
Schéma : Selon le modèle, Yt-1→It→Yt→Ct→Yt→It
Note : Cette fois ci, on va bouger Yt-1 et bloquer Gt.

Etapes à suivre :

- Dans « **Boîte A** », pour créer un nouveau scénario, procéder comme suit (suivre) : *View/Scenarios…* →*Create New Scenario*→*Sélectionner* « ***Scenario 2*** ».
- ***Simulation (Yt-1 ↑ 2% en 1987)*** :
  - *Ici, nous faisons une simulation dans l'échantillon (1987)* ;
  - L'on devra bouger « Yt-1 » et bloquer « Gt » pour voir la réaction du modèle (toutes les variables endogènes) ;
  - Suivre : *Quik/Generate Series…* →Dans la boîte de dialogue qui s'affiche, saisir à tour de rôle, *avec « YLAG_2 » traduisant la variable du scénario 2* :

ylag_2=yt(-1) → sample : 1967 1987 → ok : *définir le 2ème Scénario*
ylag_2=ylag_2(-1)*1.02 → sample : 1987 1987 → ok : *Yt-1 ↑ 2% en 1987*
gt_2=gt → sample : 1967 1987 → ok : *bloquer «Gt »*
gt_2=@elem(gt_2, 1986) → sample : 1987 1987 → ok *: bloquer «Gt » pour 1987.*

  - Revenir à la « Boîte A », cliquer sur l'onglet « **Solve** »→[Active : ***Scenario 2***]→Ok, le message suivant s'affiche (*indiquant que la simulation a réussi*) :



Et nous pouvons observer ce qui suit (*Cfr données sur Workfile*) :

Période	Ylag_2	Yt_2	It_2	Ct_2
1986	358.800	7.398.695	1.561.413	5.798.683
1987	365.976	9.621.355	1.995.352	7.540.603
%	2%	30.04%	27.79%	30.04%

Analyses :

d) ***Elaboration du 3ème Scénario et Analyses : « Gt» augmente de « 1% » et « Yt-1 » augmente en même temps de 1,5% (en 1987)***

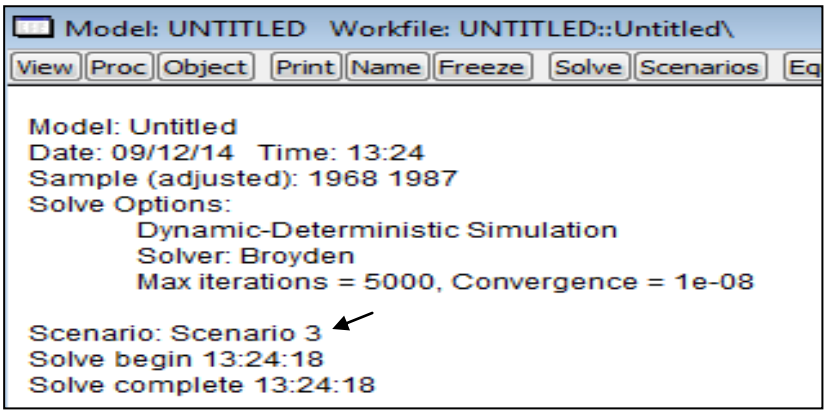
Schéma : Pour rappel, selon le modèle : Yt-1→It→Yt→Ct et Gt→Yt→It
Note : Cette fois ci, on va bouger Yt-1 et Gt en même temps.

Etapes à suivre :

- Dans « **Boîte A** », procéder comme suit (suivre) : *View/Scenarios…→Create New Scenario→Sélectionner « **Scenario 3** ».*
- ***Simulation (Yt-1 ↑ 1,5% et Gt 1% en 1987)*** :
  - *Ici, nous faisons une simulation dans l'échantillon (1987)* ;
  - L'on devra bouger, en même temps, « Yt-1 » et « Gt » pour voir la réaction du modèle (toutes les variables endogènes) ;
  - Suivre : *Quik/Generate Series…* →Dans la boîte de dialogue qui s'affiche, saisir à tour de rôle, *avec « YLAG_3 » et « Gt_3 » traduisant les variables du scénario 3* :

ylag_3=yt(-1) → sample : 1967 1987 → ok : *définir le 3ème Scénario sur « Yt-1 »*
ylag_3=ylag_3(-1)*1.015 → sample : 1987 1987 → ok : *Yt-1 ↑ 1,5% en 1987*
gt_3=gt → sample : 1967 1987 → ok : *définir le 3ème Scénario sur «Gt »*
gt_3= gt_3(-1)*1.01 → sample : 1987 1987 → ok *: Gt ↑ 1% en 1987.*

  - Revenir à la « Boîte A », cliquer sur l'onglet « **Solve** »→[Active : ***Scenario 3***]→Ok, le message suivant s'affiche (*indiquant que la simulation a réussi*) :



Et nous pouvons observer ce qui suit (*Cfr données sur Workfile*) :

Période	Gt_3	Ylag_3	Yt_3	It_3	Ct_3
1986	38.600	358.800	7.398.695	1.561.413	5.798.683
1987	38.986	364.182	9.621.355	1.995.352	7.540.603
%	1%	1.5%	30.04%	27.79%	30.04%

Analyses :

e) ***Elaboration du 4ème Scénario et Analyses : « Gt» baisse de « 3% », ceteris paribus, mais prévision/simulation hors échantillon (en 1988)***

Schéma : Pour rappel, selon le modèle : Gt→Yt→It et Gt→Yt→Ct

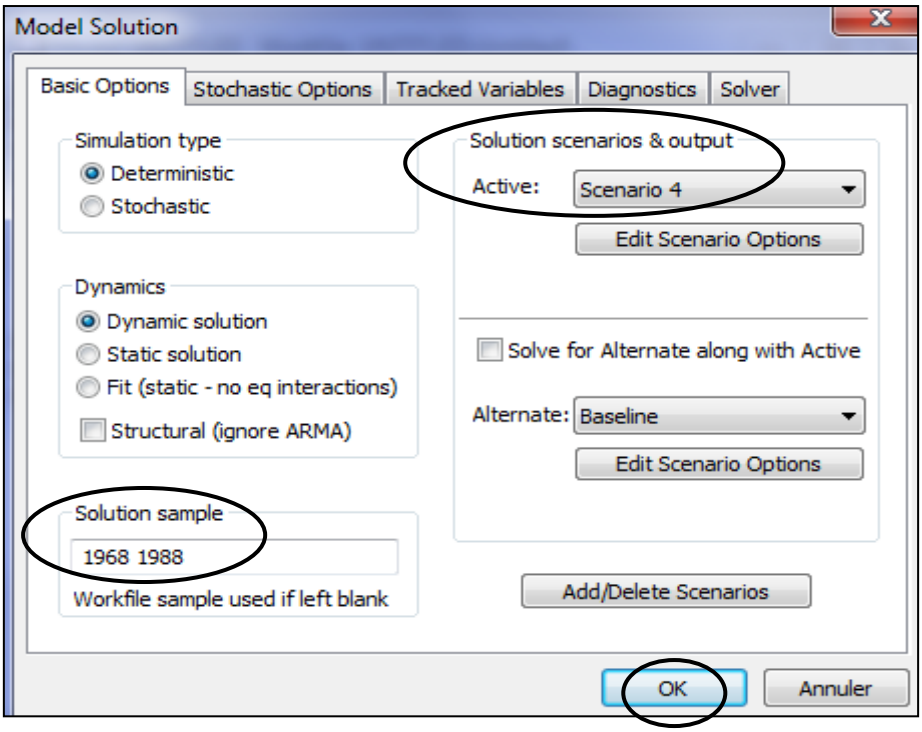
Note : on bouge « Gt », « Yt-1 » reste constante (bloquée).

Etapes à suivre :

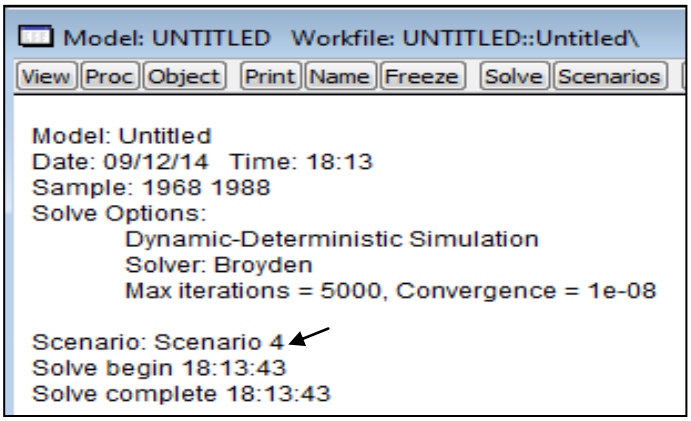
- Dans « **Boîte A** », procéder comme suit (suivre) : *View/Scenarios…→Create New Scenario→Sélectionner « **Scenario 4** ».*
- ***Simulation (Gt ↓ 3% en 1988, ceteris paribus)*** :
  - *Ici, nous faisons une simulation **hors** échantillon (1988)* ;
  - L'on devra bouger seulement « Gt » pour voir la réaction du modèle (toutes les variables endogènes) en dehors de l'échantillon ;
  - *Suivre* :
    - *Cfr Workfile : Proc→Structure/Resize Current Page…→dans « End date », indiquer « 1988 »→Ok→Yes* : Elargir l'horison temporel jusqu'à 1988 ;
    - *Cfr Barre des menus : Quik/Generate Series…* →Dans la boîte de dialogue qui s'affiche, saisir à tour de rôle, *avec « Gt_4 » traduisant la variable du scénario 4* :

ylag_4=yt(-1) → sample : 1967 **1988** → ok : *bloquer « Yt-1 »*
ylag_4=@elem(ylag_4, **1987**)→ sample : **1988 1988** → ok : *bloquer "Yt-1" en **1988***
gt_4=gt → sample : 1967 **1988** → ok : *définir le 4ème Scénario*
gt_4=gt_4(-1)*0.97 → sample : **1988 1988** → ok : *Gt ↓ 3% en **1988***

- Revenir à la « ***Boîte A*** », cliquer sur l'onglet « **Solve** »→[Active : ***Scenario 4***]→… la boîte de dialogue ci-dessous complète la procédure :



→ Cliquer sur « Ok », le message suivant s'affiche (*indiquant que la simulation a réussi*) :



Et nous pouvons observer ce qui suit (*Cfr données sur Workfile*) :

Période	Gt_4	Yt_4	It_4	Ct_4
1987	85.400	9.621.355	1.995.352	7.540.603
1988	82.838	12.616.090	2.645.645	9.887.602
%	-3%	31.13%	32.59%	31.12%

Analyses :

Exportation des « résultats/simulations/projections » d'EViews vers Excel :

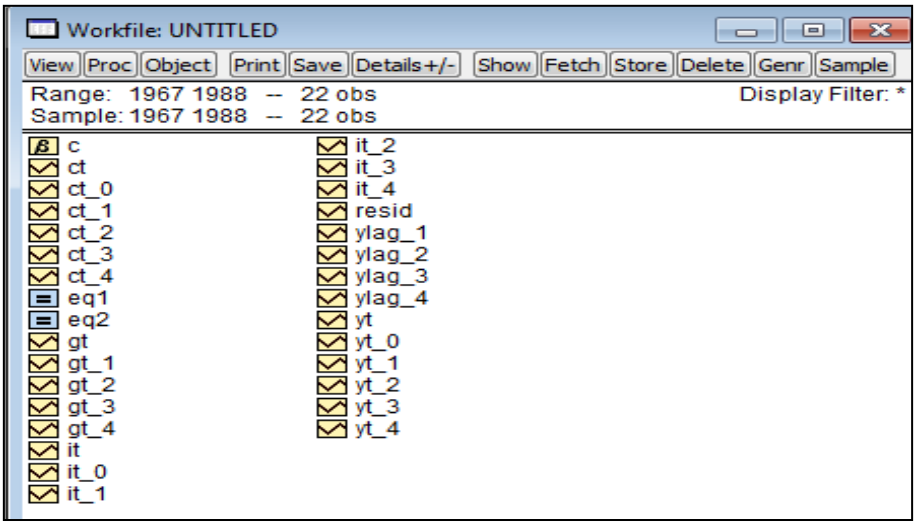
Dans EViews, sélectionner les variables à transférer et suivre : *File/Export/Write Text-Lotus-Excel*→Choisir le nom du fichier Excel « *simulation result* » et son emplacment, et préciser le type de fichier de sortie (« *Excel* »)→dans « *Series to export* », vérifier que toutes les variables sélectionnées apparaissent dans la boîte de dialogue qui s'affiche→Ok.

Mise à jour du modèle :

- Plusieurs raisons justifient la mise à jour d'un modèle macro-économétrique, entre autres :
  - Le renouvellement des données par *an/trimestre/mois* : cela implique ré-estimation avec l'ajout des nouvelles données ;
  - Les mutations économiques → l'augmentation des équations de comportement → prise en compte des chocs ou nouveaux enjeux économiques. L'on notera que, pour les modèles annuels par exemple, ils sont mis à jour au moins une fois par an.
- Etapes à suivre : dans EViews, après avoir introduit toutes les nouvelles données quantitatives et/ou qualitatives, suivre (*Cfr « Boîte A »*) : *Proc/Links/Update All Links-Recompile Model*. Après, résoudre/simuler/projeter les politiques économiques ou sociales.

****

Annexe : *Paysage relatif à notre Workfile*



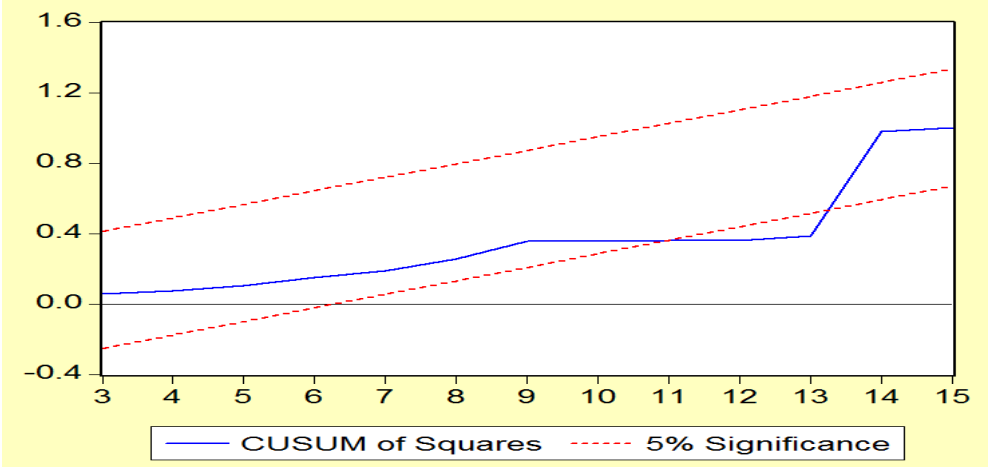
Annexe

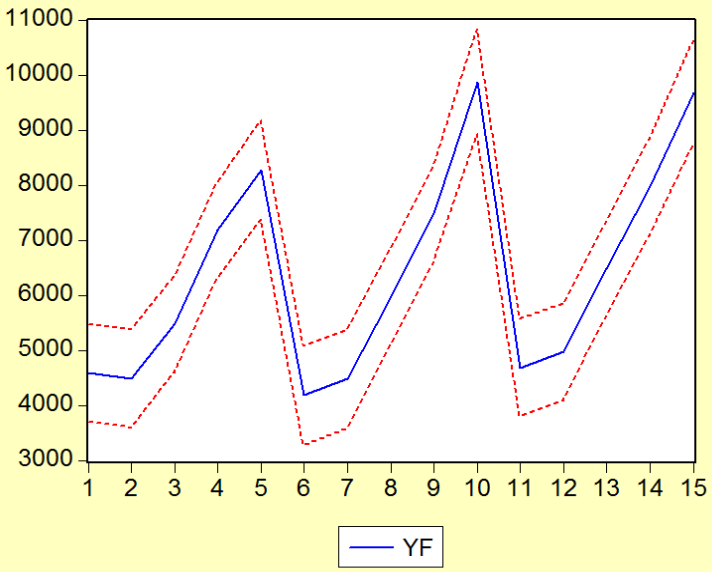
Modèle de départ (coût – production) estimé et problème posé :

Dependent Variable: Y
Method: Least Squares
Date: 10/27/13 Time: 19:08
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1983.075	276.5384	7.171065	0.0000
X	99.87273	5.778212	17.28437	0.0000

R-squared	0.958300	Mean dependent var	6384.133
Adjusted R-squared	0.955092	S.D. dependent var	1971.691
S.E. of regression	417.8304	Akaike info criterion	15.03159
Sum squared resid	2269569.	Schwarz criterion	15.12600
Log likelihood	-110.7370	F-statistic	298.7493
Durbin-Watson stat	2.327887	Prob(F-statistic)	0.000000





Forecast: YF
Actual: Y
Forecast sample: 1 15
Included observations: 15

Root Mean Squared Error	388.9789
Mean Absolute Error	316.6523
Mean Abs. Percent Error	5.080918
Theil Inequality Coefficient	0.029218
Bias Proportion	0.000000
Variance Proportion	0.010648
Covariance Proportion	0.989352

Transformation logarithmique (modèle estimé et correction du problème de stabilité)

Estimation Command:
==========
LS LY C LX

Estimation Equation:
==========
LY = C(1) + C(2)*LX

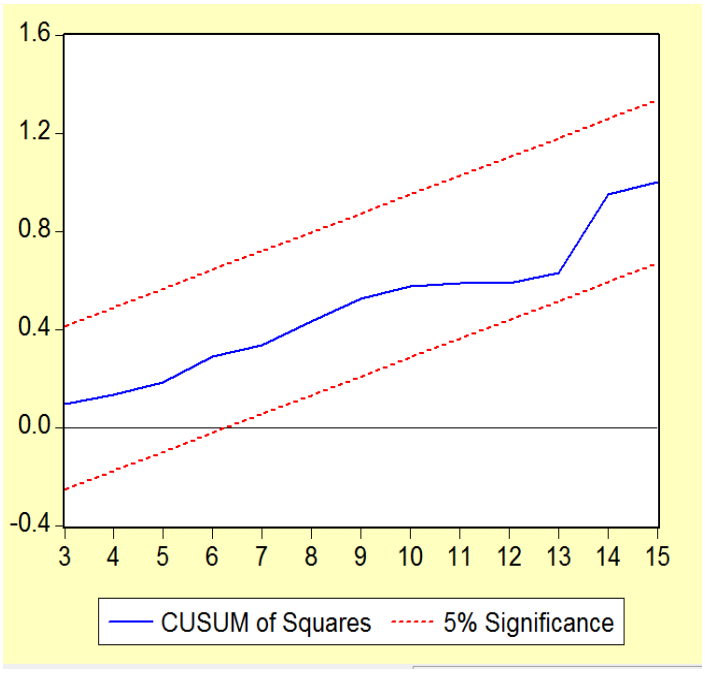
Substituted Coefficients:
==========
LY = 6.296353681 + 0.6555841585*LX

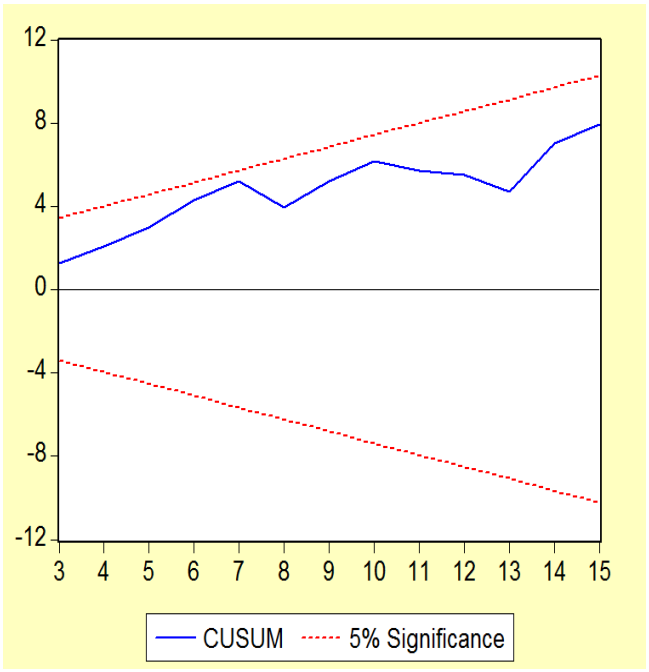
Dependent Variable: LY
Method: Least Squares
Date: 10/27/13 Time: 19:13
Sample: 1 15
Included observations: 15

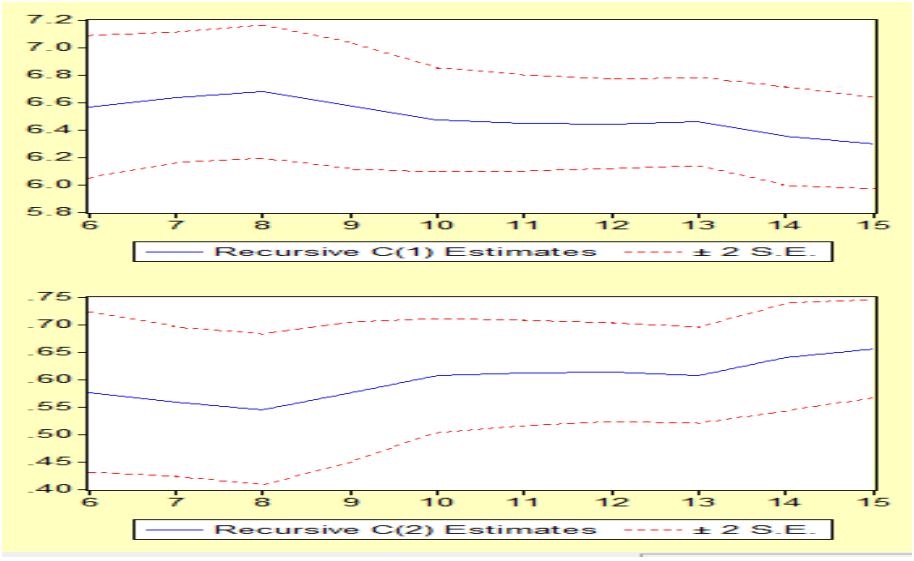
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.296354	0.164473	38.28190	0.0000
LX	0.655584	0.044208	14.82944	0.0000

R-squared	0.944185	Mean dependent var	8.719573
Adjusted R-squared	0.939892	S.D. dependent var	0.295522
S.E. of regression	0.072453	Akaike info criterion	-2.288184
Sum squared resid	0.068243	Schwarz criterion	-2.193777
Log likelihood	19.16138	F-statistic	219.9122
Durbin-Watson stat	2.160183	Prob(F-statistic)	0.000000

Stabilité de paramètres estimés dans l'échantillon (View/stability test/Recursive Estimate (OLS only…)/Recursive coefficient) :



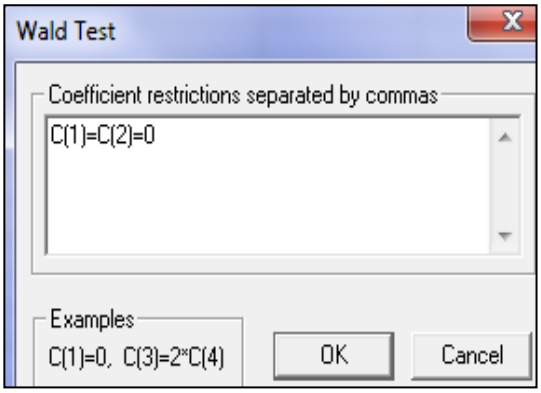




Matrice de variances-covariances

		C	LX
		C	LX
C	C	0.027051	-0.007224
LX	LX	-0.007224	0.001954

Test de restriction sur les paramètres (test de Fisher et celui de Wald) :



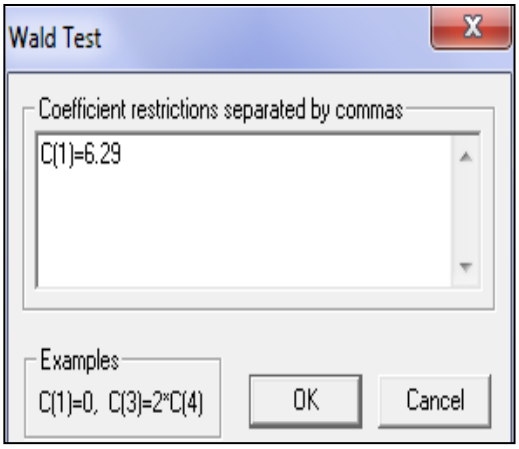
Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	108736.4	(2, 13)	0.0000
Chi-square	217472.8	2	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	6.296354	0.164473
C(2)	0.655584	0.044208

Restrictions are linear in coefficients.



Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	0.001492	(1, 13)	0.9698
Chi-square	0.001492	1	0.9692

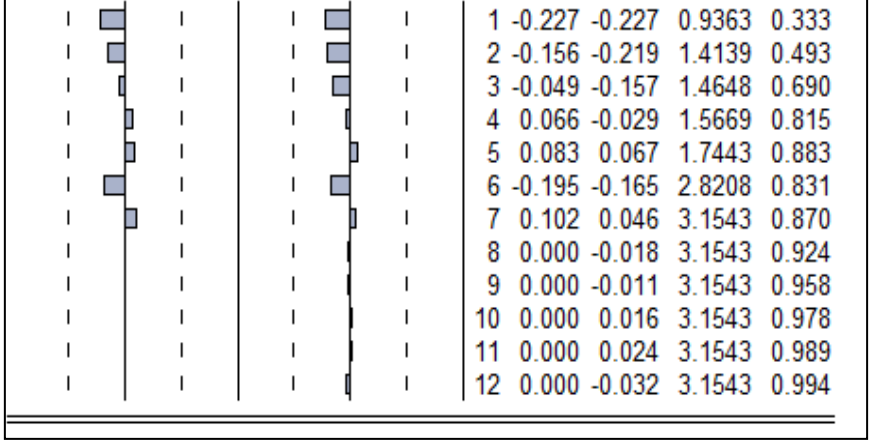
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-6.29 + C(1)	0.006354	0.164473

Restrictions are linear in coefficients.

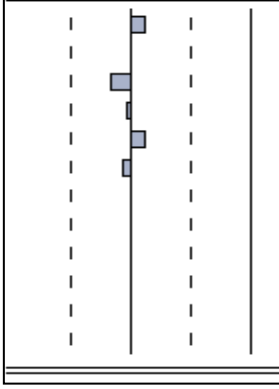
Corrélogramme :

Date: 10/27/13 Time: 19:17
Sample: 1 15
Included observations: 15



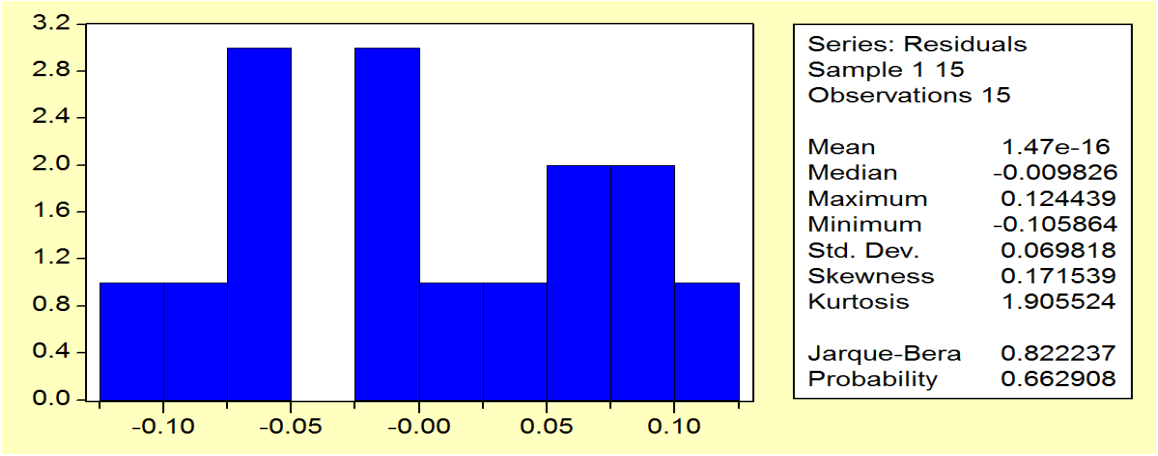
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.227	-0.227	0.9363	0.333
		2	-0.156	-0.219	1.4139	0.493
		3	-0.049	-0.157	1.4648	0.690
		4	0.066	-0.029	1.5669	0.815
		5	0.083	0.067	1.7443	0.883
		6	-0.195	-0.165	2.8208	0.831
		7	0.102	0.046	3.1543	0.870
		8	0.000	-0.018	3.1543	0.924
		9	0.000	-0.011	3.1543	0.958
		10	0.000	0.016	3.1543	0.978
		11	0.000	0.024	3.1543	0.989
		12	0.000	-0.032	3.1543	0.994

Date: 10/27/13 Time: 19:18
Sample: 1 15
Included observations: 15



	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1			0.118	0.118	0.2547	0.614
2			0.003	-0.012	0.2548	0.880
3			-0.172	-0.173	0.8830	0.830
4			-0.032	0.009	0.9070	0.924
5			0.125	0.136	1.3054	0.934
6			-0.062	-0.130	1.4157	0.965
7			-0.005	0.009	1.4164	0.985
8			0.000	0.057	1.4164	0.994
9			0.000	-0.040	1.4164	0.998
10			0.000	-0.022	1.4164	0.999
11			0.000	0.045	1.4164	1.000
12			0.000	-0.019	1.4164	1.000

Normalité



Autocorrélation :

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.886288	Probability	0.439671
Obs*R-squared	2.081698	Probability	0.353155

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 10/27/13 Time: 19:23
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.054091	0.170962	0.316393	0.7576
LX	-0.016299	0.046293	-0.352078	0.7314
RESID(-1)	-0.253924	0.301230	-0.842959	0.4172
RESID(-2)	-0.448059	0.361043	-1.241010	0.2404

R-squared	0.138780	Mean dependent var	1.47E-16
Adjusted R-squared	-0.096098	S.D. dependent var	0.069818
S.E. of regression	0.073095	Akaike info criterion	-2.170922
Sum squared resid	0.058772	Schwarz criterion	-1.982109
Log likelihood	20.28192	F-statistic	0.590859
Durbin-Watson stat	1.951619	Prob(F-statistic)	0.633709

ARCH Test:

F-statistic	0.026452	Probability	0.873509
Obs*R-squared	0.030792	Probability	0.860705

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/27/13 Time: 19:24
Sample (adjusted): 2 15
Included observations: 14 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004343	0.001875	2.316478	0.0390
RESID^2(-1)	0.047200	0.290213	0.162639	0.8735

R-squared	0.002199	Mean dependent var	0.004564
Adjusted R-squared	-0.080951	S.D. dependent var	0.004650
S.E. of regression	0.004835	Akaike info criterion	-7.694469
Sum squared resid	0.000280	Schwarz criterion	-7.603175
Log likelihood	55.86128	F-statistic	0.026452
Durbin-Watson stat	1.959328	Prob(F-statistic)	0.873509

White (no cross terms) :

White Heteroskedasticity Test:

F-statistic	0.381901	Probability	0.690569
Obs*R-squared	0.897619	Probability	0.638388

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/27/13 Time: 19:26
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.089660	0.122473	-0.732076	0.4782
LX	0.052233	0.066409	0.786534	0.4468
LX^2	-0.007142	0.008903	-0.802232	0.4380

R-squared	0.059841	Mean dependent var	0.004550
Adjusted R-squared	-0.096852	S.D. dependent var	0.004481
S.E. of regression	0.004693	Akaike info criterion	-7.708526
Sum squared resid	0.000264	Schwarz criterion	-7.566916
Log likelihood	60.81395	F-statistic	0.381901
Durbin-Watson stat	1.834697	Prob(F-statistic)	0.690569

(white avec cross terms) :

White Heteroskedasticity Test:

F-statistic	0.381901	Probability	0.690569
Obs*R-squared	0.897619	Probability	0.638388

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/27/13 Time: 19:26
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.089660	0.122473	-0.732076	0.4782
LX	0.052233	0.066409	0.786534	0.4468
LX^2	-0.007142	0.008903	-0.802232	0.4380

R-squared	0.059841	Mean dependent var	0.004550
Adjusted R-squared	-0.096852	S.D. dependent var	0.004481
S.E. of regression	0.004693	Akaike info criterion	-7.708526
Sum squared resid	0.000264	Schwarz criterion	-7.566916
Log likelihood	60.81395	F-statistic	0.381901
Durbin-Watson stat	1.834697	Prob(F-statistic)	0.690569

Bonne spécification (Ramsey-Reset) :

Ramsey RESET Test:

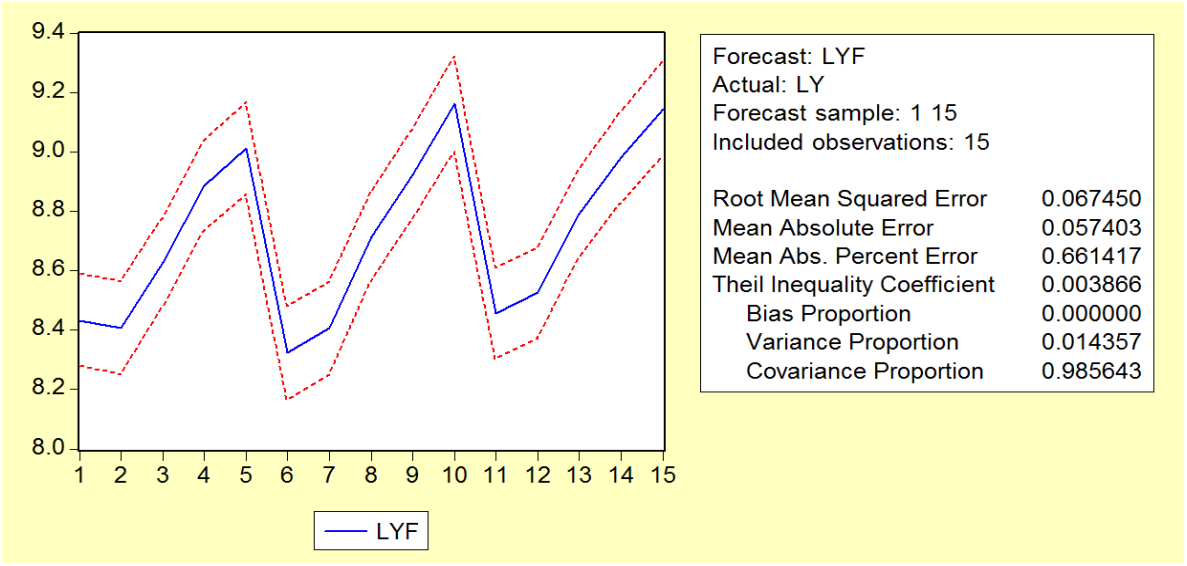
F-statistic	6.600463	Probability	0.013079
Log likelihood ratio	11.82743	Probability	0.002702

Test Equation:
Dependent Variable: LY
Method: Least Squares
Date: 10/27/13 Time: 19:30
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1640.250	792.9030	-2.068664	0.0629
LX	-322.1676	153.8089	-2.094597	0.0602
FITTED^2	55.55050	26.79803	2.072932	0.0625
FITTED^3	-2.087365	1.019841	-2.046755	0.0653

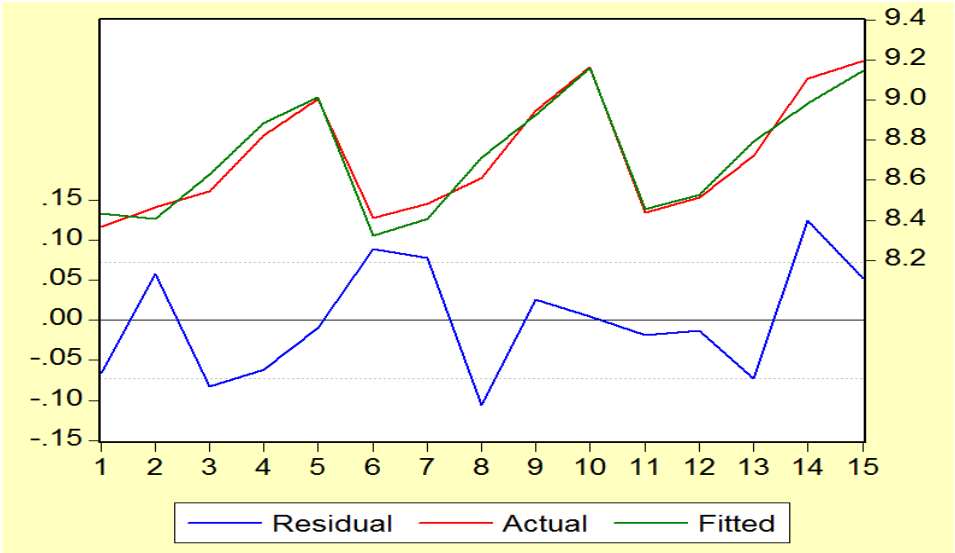
R-squared	0.974631	Mean dependent var	8.719573
Adjusted R-squared	0.967712	S.D. dependent var	0.295522
S.E. of regression	0.053102	Akaike info criterion	-2.810013
Sum squared resid	0.031018	Schwarz criterion	-2.621200
Log likelihood	25.07510	F-statistic	140.8639
Durbin-Watson stat	2.346071	Prob(F-statistic)	0.000000

Forecast :



Forecast: LYF
Actual: LY
Forecast sample: 1 15
Included observations: 15

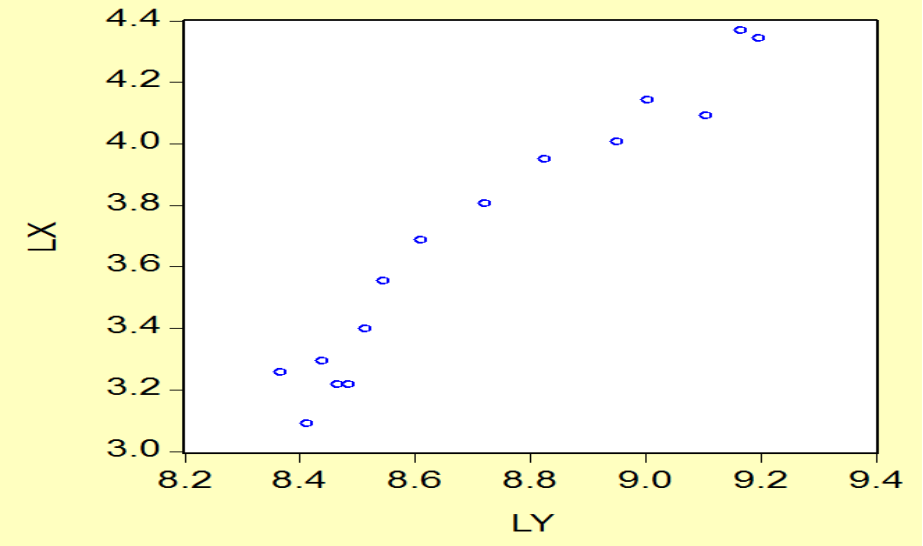
Root Mean Squared Error	0.067450
Mean Absolute Error	0.057403
Mean Abs. Percent Error	0.661417
Theil Inequality Coefficient	0.003866
Bias Proportion	0.000000
Variance Proportion	0.014357
Covariance Proportion	0.985643



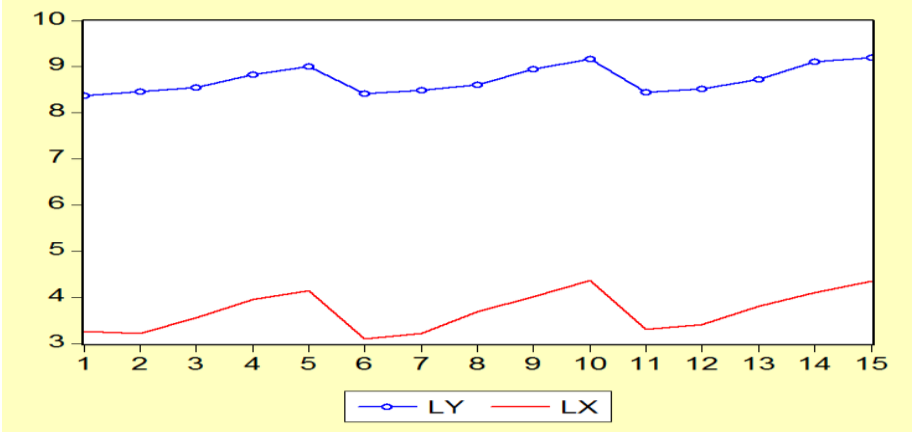
obs	Actual	Fitted	Residual	Residual Plot
obs	Actual	Fitted	Residual	Residual Plot
1	8.36637	8.43231	-0.06594	
2	8.46379	8.40660	0.05719	
3	8.54481	8.62718	-0.08238	
4	8.82468	8.88673	-0.06205	
5	9.00270	9.01253	-0.00983	
6	8.41183	8.32279	0.08904	
7	8.48467	8.40660	0.07807	
8	8.60886	8.71472	-0.10586	
9	8.94898	8.92350	0.02548	
10	9.16513	9.16089	0.00424	
11	8.43815	8.45705	-0.01890	
12	8.51319	8.52612	-0.01294	
13	8.71932	8.79194	-0.07262	
14	9.10498	8.98054	0.12444	
15	9.19614	9.14408	0.05206	

Graphiques :

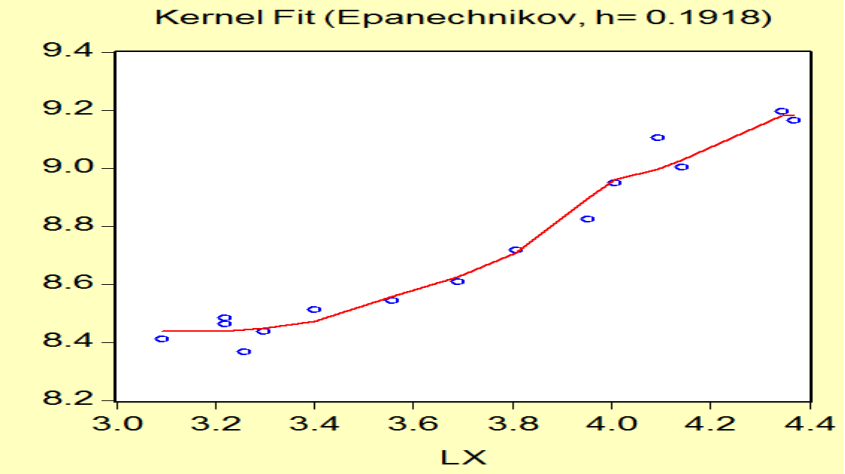
```
Scat LY LX
```



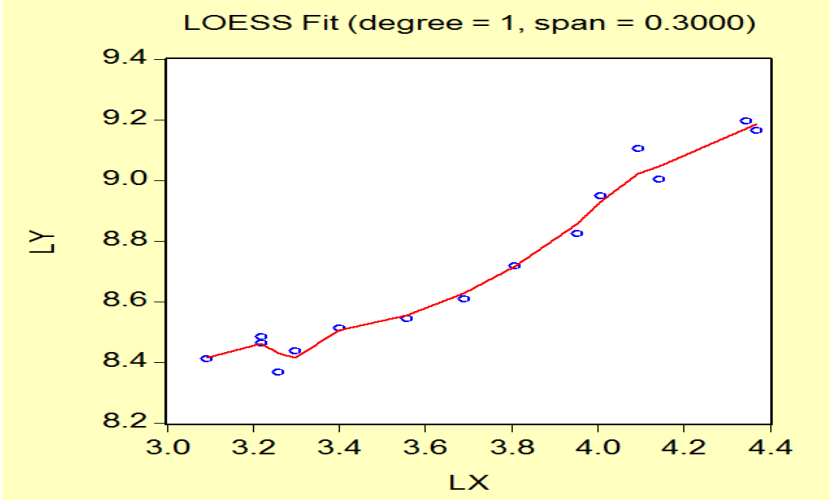
Sélectionner LY et LX/clic droit/open/view/Graph/Line :



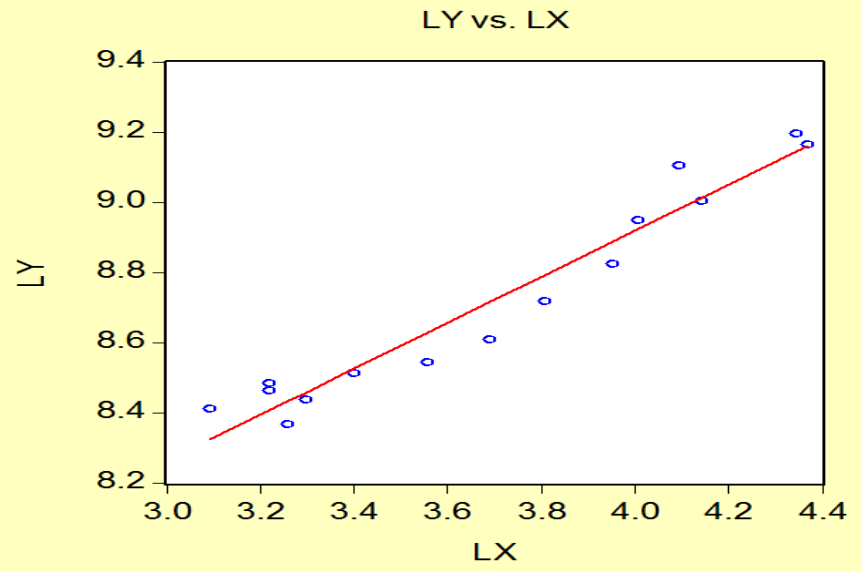
Sélectionner LY et LX/clic droit/open/view/Graph/Scatter/Scatter with Kernel Fit… :



Sélectionner LY et LX/clic droit/open/view/Graph/Scatter/Scatter with Nearest Neighbor Fit… :



Sélectionner LY et LX/clic droit/open/view/Graph/Scatter/Scatter with Regression :



Dans l'output, aller dans ***Forecast*** *pour sauver la variable dépendante prévue.*

Et tester :

$H_0: \beta_2 = 0$ ($|t_c| < |t_t|, prob > 5\%$) *: rejeter le modèle 5.2*
$H_1: \beta_2 \neq 0$ ($|t_c| > |t_t|, prob < 5\%$) *: retenir le modèle 5.2*

***

Références bibliographiques

Antoine Terracol (2008), « *Stata par la pratique : statistiques, graphiques et éléments de programmation par Eric Cahuzac et Christophe Bontemps* », the Stata Journal 8, Number 4, pp. 574-578.

Baltagi Badi H. (2005), « *Econometric Analysis of Panel Data* », 3è édition, JW edition, Angletere, 316 p.

Benchimol Jonathan, « *Formation EViews 7 : Introduction* », 95 p.

Bocquier Philippe (1996), « *L'analyse des enquêtes biographiques à l'aide du logiciel STATA* », Documents et Manuels du CEPED n°4, Paris, 224 p.

Bontemps Christophe (2002), « *Stata par la pratique – Partie II : les graphiques de Stata* », 20 p.

Bourbonnais R. (2015), « *Econométrie : cours et exercices corrigés* », 9è édition, éd. DUNOD, Paris, 392 p.

_____**(2009)**, « *Logiciel Eviews* », Université de Paris-Dauphine, 31 p.

Bourbonnais, R. et Terraza, M. (2016), « *Analyse des séries temporelles – Applications à l'économie et à la gestion : Cours et exercices corrigés* », éd. Dunod, 4è édition, Paris, 354 p.

Bozio Antoine (2005), « *Introduction au logiciel STATA* », Paris, 18 p.

Cadoret I. et al. (2009), « *Econométrie appliquée : Méthodes – Applications – Corrigés* », éd. de boeck, 2è édition, Bruxelles, 462 p.

Cadot Olivier (2012), « *Stata pour les nuls* », 65 p.

Casin Philippe (2009), « *Econométrie : Méthodes et applications avec EViews* », éd. Technip, Paris, 224 p.

Charpentier Arthur, « *Cours de séries temporelles : Théories et applications – Volume 2* », 141 p.

Christopher Baum F. (2001), « *Stata : the language of choice for time series analysis* », in The Stata Journal 1, number 1, pp. 1-16.

Couderc Nicolas, « *Econométrie appliquée avec Stata* », Université Paris 1 Panthéon-Sorbonne, 22 p.

Deniu C., Fiori G. et Mathis A. (2015), « *Sélection du nombre de retards dans un modèle VAR : Conséquences éventuelles du choix des critères* », in Economie et prévision, n° 106, 1992-5. Développements récents de la macro-économie, pp. 61-69. *(lien : http://www.persee.fr/doc/ecop_0249-4744_1992_num_106_5_5315).*

Desjardins Julie (2005), « *L'analyse de régression logistique* », Tutorial in Quantitative Methods for Psychology, vol. 1(1), pp. 35-41.

Doucoure Fodiye B. (2008), « *Méthodes économétriques : cours et travaux pratiques* », éd. ARIMA, 5è édition, Dakar, 511 p.

Goaied M. et Sassi S. (2012), « *Econométrie de données de Panel sous Stata* », 1ère édition, I.H.E.C/LEFA, 45 p.

Gosse, J-B. et Guillaumin, C. (2011), « *Christopher A. Sims et la représentation VAR* », 15 p.

Hurlin Christophe, « *Econométrie des variables qualitatives : Modèles à variable dépendante limitée (Modèles Tobit simples et Tobit Généralisés)* », 52 p.

_____ **(2003),** « *Econométrie des variables qualitatives : Modèles Dichotomiques Univariés (Modèles Probit, Logit et Semi-Paramétriques)* », 57 p.

_____ **(2003),** « *Econométrie des variables qualitatives : Modèles Multinomiaux (Modèles Logit Multinomiaux Ordonnés et non Ordonnés)* », 32 p.

_____, « *L'Econométrie des Données de Panel : Modèles Linéaires Simples* », 68 p.

_____ **(2007)**, « *Modèles ARCH-GARCH: Application à la VaR* », Université d'Orléans, 78 p.

Hurlin, C. et Mignon, V. (2006), « *Une synthèse des Tests de Cointégration sur Données de Panel* », 33 p.

I Gusti Ngurah A. (2009), « *Time Series Data Analysis Using Eviews* », édition John Wiley and Sons, 635 p.

Kenneth Simons L. (2013), « *Useful Stata Commands (for Stata version 12)* », 47 p.

Kintambu Mafuku E.G. (2004), « *Principes d'Econométrie* », Presses de l'Université Kongo, 4è édition, 285 p.

Kpodar Kangni (2007), « *Manuel d'initiation à Stata (version 8)* », CERDI, Clermont-Ferrand, 97 p.

Lubrano Michel (2008), « *Modélisation Multivariée et Cointégration* », 32 p.

_____ **(2008)**, « *Tests de Racine Unitaire* », 46 p.

Luyinduladio Menga E. (2009), « *Manuel d'initiation à EViews* », inédit, Septembre, 79 p.

Michée Sendula, « *Guide d'utilisation Stata 9* », inédit, 44p.

Mignon Valérie (2008), « *Econométrie : Théorie et Applications* », éd. ECONOMICA, 236 p.

Nicholas L., Hébert B-P et Laplante B. (2007), « *introduction à Stata* », 46 p.

Ouellet Estelle (2005), « *Guide d'économétrie appliquée pour Stata pour ECN 3950 et FAS 3900* », Université de Montréal.

Park Hun Myoung (2008), « *Univariate Analysis and Normality Test Using SAS, Stata, and SPSS* », Working Paper, The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University, 41 p.

Pellier K. (2007), « *Travaux Dirigés d'Econométrie – M1 : Guide d'utilisation d'EViews* », 14 p.

Pétry F. et Gélineau F., « *Guide pratique d'introduction à la régression en sciences sociales – Deuxième édition revue et augmentée* », 42 p.

Quantitative Micro Software/QMS (2004), « *EViews 7 Command and Programming Reference* », USA, Avril, 580 p.

Quantitative Micro Software/QMS (2004), « *EViews 7 Object Reference* », USA, Novembre, 764 p.

Rabe-Hesketh S. et Everitt B.S. (2004), « *A Handbook of Statistical Analyses using Stata* », 3è edition, éd. CHAPMAN and HALL/CRC, Londres, 304 p.

Ricardo Perez-Truglia (2009), « *Applied Econometrics using Stata* », Havard University, 170 p.

Robert Alan Y. (2007), « *Robust Regression Modeling with STATA – lecture notes* », 93 p.

_____ **(2007)**, « *Stata 10 (Time Series and Forecasting)* », in Journal of Statistical Software, volume 23, Software Review 1, december, 18 p.

Robert de Jong (2003), « *Eviews mini manual* », 6 p.

Robert Dixon, « *Simulation of the Klein–Goldberger Model using Eviews* », 6 p.

Rous Philippe, « *Modèles Estimés sur Données de Panel* », cours d'Econométrie des données de panel – Master « Economie et Finance », Université de Limoges, 76 p.

Tombola Muke C. (2012), « *Séminaire d'Economie Mathématique I-avec initiation aux logiciels EVIEWS, STATA et MATLAB/modules 1 & 2* », Laréq, Décembre, 37 p. (www.lareq.com).

Tsasa Vangu K. JP. (2012), « *Introduction à la programmation à l'aide du logiciel EViews – avec 11 programmes exécutables sur EViews : illustrations + commentaires* », Laréq, Décembre, 19 p. (www.lareq.com).

Vescovo Aude, « *Cours sur le logiciel Stata* », IRD-AFRISTAT, 40 p.
