COMMENTAIRES SUR LA COMMANDE LINREG

C. BABUSIAUX 10/03/2007

1 La commande LINREG

Prenons l'exemple cmus.prg et les données cmus.rat pour décortiquer un peu la commande LINREG (notée LIN car RATS dans toutes les commandes se contente des trois premières lettres des commandes). Pour plus de détails voir dans le HELP SEARCH et LINREG.

On va régresser CM en fonction d'une constante et des variables RD, SP, TCHO. Les données sont trimestrielles du premier trimestre 1949 au dernier trimestre 2002
cal 1949 1 4
all 2002:4
open data cmus.rat
data(for=rats) / cm pc tcho rd sp
com start = 1955:1
com end = 2002:4
lin cm start end residus
 # constant rd sp tcho
Sur la première ligne on écrit LIN puis la variable endogène suivie de / si on veut travailler sur tout l'échantillon ou de la valeur de début et de fin sinon; si on veut récupérer les résidus on met ensuite le nom que l'on souhaite donner au vecteur des résidus. Dans cet exemple on a des données de 1949 à 2002, et on utilise les données pour les MCO seulement de 1955:1 à 2002:4
Pour cela on a deux possibilités:

- soit faire comme ci-dessus et l'échantillon sera réduit à 1955-2002 seulement dans cette régression et pas dans la suite
- soit utiliser la commande SMPL , dans ce cas le travail se fera sur la période 1955 2002 tant que l'on aura pas changé de SMPL

  smpl start end

  lin cm / residus

  # constant rd sp tcho

- si on n'avait pas mis de SMPL et écrit la formule suivante

  lin cm / residus

  # constant rd sp tcho

  alors la régression sera faite sur toute la période 1949-2002.

Sur la seconde ligne on met un # suivi de la liste des variables explicatives en commençant en général par la constante notée CONSTANT par RATS qui construit ainsi automatiquement le vecteur unité.

1.1 Les options de LINREG

Je vous donne les plus utilisées (voir dans le HELP SEARCH et LINREG pour plus de détails).

- Si vous ne souhaitez pas voir apparaitre le résultat de la régression écrire LIN(NOPRINT)

  lin(noprint) cm / residus

  # constant rd sp tcho

- Si on souhaite la correction des degrés de liberté dans le cas de modèles autorégressifs ayant r retards dans l'endogène comme variables explicatives indiquer DCF= r

  lin(dcf=2) cm / residus

  # constant cm{1 to 2} rd sp tcho

- Si on souhaite utiliser la méthode des variables instrumentales on utilise INSTRUMENTS. On commence par donner la liste des instruments

  instruments constant rd{1 to 2} tcho{1 to 2}

  lin(inst) cm / residus

  # constant cm{1 to 2} rd sp tcho

2 Descriptions des sorties RATS

com start = 1955:1
com end = 2002:4
lin cm start end residus
 # constant rd sp tcho

```
Linear Regression - Estimation by Least Squares
Dependent Variable CM
Quarterly Data From 1955:01 To 2002:04
Usable Observations    192      Degrees of Freedom   188
Centered R**2     0.999209      R Bar **2   0.999196
Uncentered R**2   0.999865      T x R**2     191.974
Mean of Dependent Variable      3335.9619792
Std Error of Dependent Variable 1514.8307629
Standard Error of Estimate        42.9408363
Sum of Squared Residuals        346656.09939
Regression F(3,188)               79169.1015
Significance Level of F           0.00000000
Log Likelihood                    -992.30114
Durbin-Watson Statistic             0.626909

   Variable                     Coeff       Std Error      T-Stat     Signif
********************************************************************************
1.  Constant                 62.75344243  13.98955930      4.48573  0.00001263
2.  RD                        0.86985854   0.00432032    201.34138  0.00000000
3.  SP                        0.34947662   0.01982126     17.63140  0.00000000
4.  TCHO                     -9.36158468   2.61057508     -3.58602  0.00042796
```

- ligne 1: Rats précise la méthode utilisée ici les MCO
- ligne 2: La variable endogène ici CM
- ligne 3: Les données sont trimestrielles du premier trimestre 1955 au quatrième trimestre 2002
- ligne 4: on utilise n=192 observations sur cette période et comme on a k=4 variables explicatives le nombre de degrés de liberté n-k=188
- Le R^2 centré $R^2 = 1 - \frac{\sum e_t^2}{\sum (CM_t - \overline{CM})^2}$ =0,999209

  et le R^2 corrigé $\overline{R^2} = 1 - \frac{\sum e_t^2/(n-k)}{\sum (CM_t - \overline{CM})^2/(n-1)}$ =0.999196
- ligne 5 : Le R^2_{nc} non centré $R^2_{nc} = 1 - \frac{\sum e_t^2}{\sum (CM_t)^2} = 0.999865$, il est utilisé seulement lorsque le modèle ne contient pas de constante ce qui n'est pas le cas ici.

  et ce R^2_{nc} multiplié pas la taille n de l'échantillon nR^2_{nc} =191.974 ce résultat est utilisé dans un test d'autocorrélation.
- ligne 6 : la moyenne de la variable endogène ici $\overline{CM}$ =3335.9619792
- ligne 7 : l'écart-type de cette variable endogène $\sqrt{\frac{\sum (CM_t - CM)^2}{(n-1)}}$ =1514.8307629
- ligne 8 : la valeur de S tel que S^2 est la meilleure estimation de σ^2 soit s=$\sqrt{\frac{\sum e_t^2}{(n-k)}} = \sqrt{\frac{SCR}{(n-k)}} = 42.9408363$
- ligne 9 : la somme des carrés des résidus $\sum e_t^2 = SCR$ =346656.09939
- ligne 10 : la statistique F=79169.1015 de la régression test de Fisher testant la nullité de tous les coefficients des MCO autres que le terme constant (Voir ce test dans la partie STUDENT et FISHER)
- ligne 11 : le niveau de significativité de ce test Pv=0.00000000 ce qui indique que l'on décide ici très nettement H1 soit que l'un au moins des coefficients est significatif
- ligne 12 : la valeur de la fonction de vraisemblance sous l'hypothèse de Normalité des erreurs lv=-992.30114
- ligne 13 : valeur de la statistique de DURBIN et WATSON DW= $\frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2}$qui teste l'autocorrélation d'ordre 1 des erreurs