## Assignment 1

Regression Modeling

This assignment is based on the dairy farms data used in the class demonstrations. Though they are a panel of 247 farms observed in 6 years, for present purposes, we will treat them as a cross section. For illustration purposes, we will also use the farm means of the 6 years of data. The data set for this exercise is dairy.lpj.

We set up some preliminary definitions – these are shorthands that we will use later.

```
? Some preliminaries to set up the data
? These commands obtain the group (farm) means of the 4 inputs,
? then computes the logs of the group means.
SETPANEL ; Group = farm ; Pds = t $
NAMELIST ; (new) ; means=cowsbar,landbar,laborbar,feedbar $
NAMELIST ; factors = cows,land,labor,feed $
CREATE   ; means = Group Mean (factors,pds=t)$
CREATE   ; milkbar=Group Mean (milk,pds=t) $
CREATE   ; yb=log(milkbar)
         ; x1b=log(cowsbar)  ;x2b=log(landbar)
         ; x3b=log(laborbar) ;x4b=log(feedbar)$
NAMELIST ; cobbdgls = one,x1,x2,x3,x4 $
NAMELIST ; quadrtic = x11,x22,x33,x44,x12,x13,x14,x23,x24,x34 $
NAMELIST ; translog = cobbdgls,quadrtic $
NAMELIST ; time = year93,year94,year95,year96,year97 $ Omit 1998
```

1. Introduction to the data

Get familiar with the data set. There is a description of the variables in the project file at the top of the project window. On the desktop menu, select Window/Output.
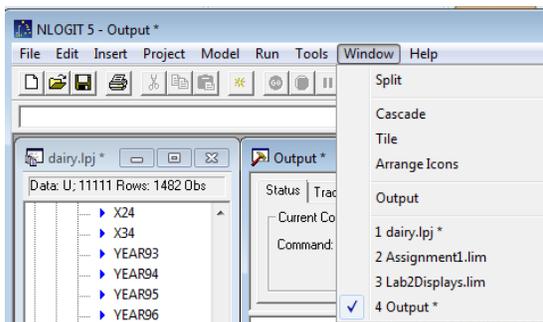
When the output window opens, scroll to the top to find the description

```
+--------------------------------------------------------------+
|Dairy Farm Data, N = 247, T = 6                               |
|Variables                                                     |
|FARM = Farm ID                                                |
|YEAR = year                                                   |
|Inputs                                                        |
|  COWS,  X1 = log of, deviations from means (logs)            |
|  LAND,  X2 = same                                            |
|  LABOR, X3 = same                                            |
|  FEED,  X4 = same                                            |
|  Translog terms, X11, X22, X33, X44,                         |
|                  X12, X13, X14,                              |
|                  X23, X24,                                   |
|                  X34                                         |
|             = squares and cross products                     |
|  YEAR93,....,YEAR98 = year dummy variables                   |
|Output                                                        |
|  MILK = farm output                                          |
|  YIT = log of MILK production                                |
+--------------------------------------------------------------+
```

Now, obtain a broad picture of the variables in the data set with

DSTAT ; Rhs = * $

2. Cobb – Douglas linear regression

Fit a Cobb-Douglas production function by least squares with the command

REGRESS   ; Lhs = yit ; Rhs = cobbdgls $

To obtain White's robust, heteroscedasticity corrected covariance matrix for the least squares estimates, it is only necessary to add **;Heteroscedasticity** or just **;Het** to the regression command. This will replace the conventional covariance estimator with the White estimator. It will also report a Breusch and Pagan LM statistic for testing for heteroscedasticity.

REGRESS   ; Lhs = yit ; Rhs = cobbdgls ; Het $

a. Reestimate the model with the White correction. Do the standard errors differ much from the OLS standard errors?
b. What is the result of the LM test for heteroscedasticity?
c. Reestimate the linear regression with a 'cluster correction.' How do the standard errors change when the panel data aspect is included in the estimates?

3. Least absolute deviations and quantile regression

In order to use least absolute deviations (LAD) estimator instead of OLS, the **REGRESS** command is changed by adding **;ALG = LAD**. (Omit the **;HET** parameter for this estimator.) In addition, the standard errors must be computed using bootstrapping. You must specify the number of bootstrap replications. Use **;NBT = the number**, for example **;NBT = 25**. Fit the Cobb-Douglas model using LAD instead of OLS, with at least 25 bootstraps. Do the results differ much from OLS?

REGRESS   ; Lhs = yit ; Rhs = cobbdgls; Alg = LAD ; NBT = 25 $

The theory of the LAD estimator is that in the model, $\beta'\mathbf{x}$ provides the median, i.e., the 50th percentile of the distribution of y|$\mathbf{x}$. It might be interesting to see how the production function behaves around other quantiles. That is, suppose we model the production function a bit less

parametrically as $y = \boldsymbol{\beta}(q)'\mathbf{x} + \varepsilon$, where q denotes a quantile of the distribution. Then, the regression above would be $y = \boldsymbol{\beta}(.5)'\mathbf{x} + \varepsilon$. You can try different quantiles, and see if the parameter estimates change by using a command such as

```
QREG  ;  LHS = YIT  ;  RHS  =  COBBDGLS  ;  Qnt = .25 $
```

With **;Qnt = .5**, you get the same results as **REGRESS** with **;Alg=LAD**. Try a few different quantiles, and see if the parameter estimates do change as q changes. You can collect the results in one table as shown below

```
QREG ; LHS = YIT ; RHS = COBBDGLS ; Qnt = .25 ; table = quant25$
QREG ; LHS = YIT ; RHS = COBBDGLS ; Qnt = .50 ; table = quant50$
QREG ; LHS = YIT ; RHS = COBBDGLS ; Qnt = .75 ; table = quant75$
MAKETABLE ; quant25,quant50,quant75 $
```

4. Hypothesis tests

We want to test for the presence of 'time' effects in the Cobb-Douglas model. The class example showed how to use an F test the translog model against the Cobb-Douglas.

a. Adapt that procedure to test for the joint significance of the 5 time dummy variables in the model

```
REGRESS ; Lhs = yit ; Rhs = time,cobbdgls $
CALC    ; ru = rsqrd ; dfn = col(time) ; dfd = n-kreg $
REGRESS ; Lhs = yit ; Rhs = cobbdgls $
CALC    ; rl = rsqrd ; list
        ; f = ((ru-rl)/dfn)/((1-ru)/dfd) ; fc=ftb(.95,dfn,dfd) $
```

What is the result of the test? Are the 5 dummy variables jointly significant?

b. The Wald test is another way to test the joint significance of a set of variables. You can do this with matrix algebra. The Wald statistic is

$$W = (\mathbf{b} - \boldsymbol{\beta}^0)' \, [\mathrm{Var}(\mathbf{b})]^{-1} \, (\mathbf{b} - \boldsymbol{\beta}^0)$$

where $\mathbf{b}$ is the estimate, $\boldsymbol{\beta}^0$ is the hypothesized value, and $\mathrm{Var}(\mathbf{b})$ is the estimated covariance matrix. In this case, $\mathbf{b}$ is the first 5 coefficients in the model and $\boldsymbol{\beta}^0$ is $\mathbf{0}$. The matrix commands you can use after fitting the model by OLS are

```
MATRIX ; bt=b(1:5) ; vt=varb(1:5,1:5) ; list ; w=bt'<vt>bt $
CALC ; list ; cc = ctb(.95,5) $
```

The result is a chi-squared statistic with 5 degrees of freedom. Is the result larger than the critical value? The last line of the example above shows how to compute the critical value from the chi-squared table. Carry out the test.

c. There is an easier way to do these calculations. The preceding is for illustration.

```
REGRESS ; Lhs = yit ; Rhs = time,cobbdgls ; Test: time $
```

5. Restriction on translog model

We now want to carry out the Wald test for the translog model against the Cobb-Douglas model, with time effects included. The following shows an advantage of using namelists as defined at the outset. Note that the quadratic terms are embedded in translog.

```
REGRESS ; Lhs = yit ; Rhs = translog,time ; Test: quadrtic $
```

6. Constant returns to scale

We now wish to test the hypothesis of a linear restriction on the model coefficients. We will test for constant returns to scale in the Cobb-Douglas model. In the Cobb-Douglas model,

```
REGRESS ; LHS = YIT ; RHS = One,X1,X2,X3,X4 $
```

the hypothesis is that the four input coefficients sum to 1. The easiest way to carry out the test is to let the program do it for you by computing the Wald (and F) statistic. Use

```
REGRESS ; LHS = YIT ; RHS = One,X1,X2,X3,X4
        ; Test:  x1 + x2 + x3 + x4 = 1 $
```

The regression results will contain an F statistic. Carry out the test. What do you conclude? Another way to do this test is to use the **WALD** command (which is very general – we'll use the simplest possible form).

```
WALD ; Fn1 = b_X1+b_X2 + b_X3 + b_X4 – 1 $
```

Carry out the test. What is the result? (Note, the **WALD** command is based on the previous regression. A more general form that is based on any parameter vector estimate and its covariance matrix, for this test, would be

```
WALD ; Start = B ; Var = VARB ; Labels = b0,b1,b2,b3,b4
     ; Fn1 = b1+b2+b3+b4 – 1 $
```

which will give the identical result. You can use any other functions of the parameters if you like.

Constant returns in the translog model involves several restrictions. The first order terms must sum to 1.0 and each row of the matrix of second order terms must sum to zero. Here is how to impose the restrictions and test the hypothesis at the same time.

```
REGRESS ; Lhs = yit ; Rhs = Translog ;
CLS: x1+x2+x3+x4 = 1,
x11+x12+x13+x14=0,
x12+x22+x23+x24=0,
x13+x23+x33+x34=0,
x14+x24+x34+x44=0 $
```

Carry out the test in a translog model that also has time effects. (It is only necessary to add **,time** to the **REGRESS** command.)

7. Nonnormality

We will now analyze the least squares residuals for evidence of nonnormality. We need first to compute them. There are two ways:

REGRESS ; Lhs = yit ; Rhs = cobbdgls ; Res = u $

or, after the **REGRESS** command, use **CREATE ; u = yit – cobbdgls'b $** (Note how the different parts of the program interact.) With the residuals in hand, first look at the distribution by estimating it with a kernel density estimator. Use

KERNEL ; Rhs = u ; normal $

Does it appear to be asymmetric? Now, we can use a chi-squared sort of statistic to 'test' for nonnormality. The test is based on the third and 4th moments of the residuals – they should be 0 and 3, respectively. The test statistic is

$$C = N \times [(m_3/s_u^3)^2 / 6 + (m_4/s_u^4 - 3)^2/24]$$

where N is the sample size, s_u is the standard deviation of the residuals and m_3 and m_4 are the third and fourth sample moments. (We can't use the name '**s**' because like '**B**' and '**VARB**' it is a program reserved name.) After obtaining the residuals as above, you can compute the parts with the following

```
CREATE ; u2 = u*u ; u3 = u2*u ; u4 = u2*u2 $
CALC ; su = sqr ( xbr ( u2 ))
     ;  m3 = xbr( u3 ) ; m4 = xbr ( u4) $
CALC ; List
     ; c = n *(( m3/su^3)^2/6 + (m4/su^4 – 3)^2 /24) $
```