THE UNIVERSITY OF ILLINOIS AT CHICAGO
ECON 534: Econometrics I
AUTUMN 2009

Prof. George Karras

Example 3: Non-Linear Least Squares and Maximum Likelihood Estimation

Consider again Mankiw, Romer, and Weil's (*QJE*, May 1992) neoclassical growth equation:

$$\ln(y_T) - \ln(y_0) = (1 - e^{-\lambda T})\left[\alpha(1-\alpha)^{-1}\ln(s) - \alpha(1-\alpha)^{-1}\ln(n) + \beta(1-\alpha)^{-1}\ln(h) - \ln(y_0)\right], \qquad (0)$$

where y is GDP per adult, T is the terminal year, 0 the initial year, λ is the rate of convergence, s is the fraction of income invested in physical capital, n is the population growth rate (plus the depreciation rate and the rate of technological growth), h is human capital, and α and β are the output elasticities with respect to physical capital and human capital, respectively. Note that the equation is *linear* in the variables $\ln(y_T)$-$\ln(y_0)$, $\ln(s)$, $\ln(n)$, $\ln(h)$, and $\ln(y_0)$, but highly *nonlinear* in the parameters λ, α, and β.

Because of the linearity of the equation in the variables, we have been able to estimate it in the linear form:

$$growth_i = \beta_0 + \beta_1 \ln y1960_i + \beta_2 \ln s_i + \beta_3 \ln pop_i + \beta_4 \ln school_i + \varepsilon_i, \qquad (1)$$

where the variables are $growth = \ln(y_{1985}) - \ln(y_{1960})$, $\ln y1960 = \ln(y_{1960})$, ln$s$ ≡ ln(s), lnpop ≡ ln(pop + 0.05), and ln$school$ ≡ ln($school$); the parameters are β_0, $\beta_1 = -(1 - e^{-\lambda T})$, $\beta_2 = (1 - e^{-\lambda T})\alpha/(1-\alpha)$, $\beta_3 = -(1 - e^{-\lambda T})\alpha/(1-\alpha)$, and $\beta_4 = (1 - e^{-\lambda T})\beta/(1-\alpha)$; and ε is the error term. But note that this method does not directly estimate the *structural* parameters λ, α, and β, or obtain their standard errors.

Here we will estimate the structural parameters λ, α, and β directly by Non-Linear Least Squares (NLLS) and Maximum Likelihood (ML). A significant additional advantage of these nonlinear approaches is that they will give us direct estimates of standard errors for the structural parameters. To implement the nonlinear estimation, add a constant and an error term to (0) and write it as

$$growth_i = \beta_0 - (1 - e^{-\lambda T})\ln y1960_i + (1 - e^{-\lambda T})\alpha(1-\alpha)^{-1}\ln s_i$$
$$- (1 - e^{-\lambda T})\alpha(1-\alpha)^{-1}\ln pop_i + (1 - e^{-\lambda T})\beta(1-\alpha)^{-1}\ln school_i + \varepsilon_i, \qquad (2)$$

where the variables are defined as in equation (1), but now the parameters to be estimated are β_0, λ, α, and β.

A. Input the data and construct the variables

```
allocate 98
*
* cross section data from Mankiw, Romer, Weil, QJE 1992 (Appendix)
data(unit=input,org=obs) / number y1960 y1985 growth pop iy school
1 2485 4371 4.8 2.6 24.1 4.5
2 1588 1171 0.8 2.1 5.8 1.8
   ...  etc.  ...
120 9523 12308 2.7 1.7 22.5 11.9
121 1781 2544 3.5 2.1 16.2 1.5
*
set y1960 = log(y1960)
set y1985 = log(y1985)
set growth = y1985 - y1960
set pop = log(pop/100.+0.05)
set iy = log(iy/100.)
set school = log(school/100.)
```

B. Run the Linear Model to obtain Initial Values

B1. Run OLS

```
linreg growth
# constant y1960 iy pop school

Dependent Variable GROWTH - Estimation by Least Squares
Usable Observations     98      Degrees of Freedom    93
Centered R**2     0.484306      R Bar **2   0.462126
Uncentered R**2   0.745475      T x R**2      73.057
Mean of Dependent Variable      0.4492766111
Std Error of Dependent Variable 0.4458053414
Standard Error of Estimate      0.3269532452
Sum of Squared Residuals        9.9415534807
Regression F(4,93)                   21.8349
Significance Level of F           0.00000000
Durbin-Watson Statistic             2.134934

   Variable                        Coeff       Std Error      T-Stat     Signif
*******************************************************************************
1.  Constant                    3.004870632  0.827868892      3.62965  0.00046387
2.  Y1960                      -0.286576570  0.061716598     -4.64343  0.00001123
3.  IY                          0.523736529  0.086847397      6.03054  0.00000003
4.  POP                        -0.504594920  0.288579024     -1.74855  0.08366852
5.  SCHOOL                      0.229479466  0.059533600      3.85462  0.00021318
```

B2. Compute and display the initial values

```
compute ib0 = %beta(1)
compute isigmasq = %seesq
* note: %beta(2) = -[1-exp(-lambda*T)]
compute ilambda = - log(%beta(2)+1.)/(1985.-1960.)
* note: %beta(3) = - %beta(2)*alpha/(1-alpha)
compute ialpha = %beta(3)/(%beta(3)-%beta(2))
* note: %beta(5) = - %beta(2)*beta/(1-alpha)
compute ibeta = - %beta(5)*(1.-ialpha)/%beta(2)
display    isigmasq    ilambda     ialpha     ibeta
           0.10690     0.01351    0.64634    0.28320
*
nlpar(subiterations=100)
```

C. Run Non-Linear Least Squares (NLLS)

C1. List the parameters and write the non-linear equation

```
nonlin b0 lambda alpha beta
*
frml equation = $
     b0 - (1.-exp(-lambda*(1985.-1960.)))*y1960 $
   + (1.-exp(-lambda*(1985.-1960.)))*alpha/(1.-alpha)*iy $
   - (1.-exp(-lambda*(1985.-1960.)))*alpha/(1.-alpha)*pop $
   + (1.-exp(-lambda*(1985.-1960.)))*beta/(1.-alpha)*school
```

C2. Set initial values equal to the values implied by the OLS estimation

```
compute b0 = ib0
compute lambda = ilambda
compute alpha = ialpha
compute beta = ibeta
```

C3. Estimate the equation by NLLS

```
nlls(frml=equation,iterations=100) growth

Dependent Variable GROWTH - Estimation by Nonlinear Least Squares
Iterations Taken     2
Usable Observations     98      Degrees of Freedom    94
Centered R**2     0.484285      R Bar **2   0.467826
Uncentered R**2   0.745465      T x R**2      73.056
Mean of Dependent Variable      0.4492766111
Std Error of Dependent Variable 0.4458053414
Standard Error of Estimate      0.3252161481
Sum of Squared Residuals        9.9419610413
Durbin-Watson Statistic             2.140439

   Variable                     Coeff       Std Error      T-Stat     Signif
*******************************************************************************
1.  B0                       2.9678112818 0.5671820197      5.23256  0.00000101
2.  LAMBDA                   0.0135709153 0.0032908809      4.12379  0.00008039
3.  ALPHA                    0.6445455868 0.0485153349     13.28540  0.00000000
4.  BETA                     0.2847445520 0.0775753697      3.67055  0.00040160
```

Note that NLLS produces standard errors for each of the estimated parameters. In fact the entire variance-covariance matrix of the parameter vector is estimated. This allows us to proceed with inference as usual.

D. Run Maximimum Likelihood (ML)

D1. List the parameters and write the log likelihood function (Note that the variance σ^2 is now to be jointly estimated with he rest of the parameters)

```
nonlin sigmasq b0 lambda alpha beta
*
frml resid = growth - $
  (  b0 - (1.-exp(-lambda*(1985.-1960.)))*y1960 $
   + (1.-exp(-lambda*(1985.-1960.)))*alpha/(1.-alpha)*iy $
   - (1.-exp(-lambda*(1985.-1960.)))*alpha/(1.-alpha)*pop $
   + (1.-exp(-lambda*(1985.-1960.)))*beta/(1.-alpha)*school  )
*
frml logl = $
     -.5*log(sigmasq) - .5*(1/sigmasq)*(resid**2)
```

D2. Set initial values equal to the values implied by the OLS estimation

```
compute sigmasq = isigmasq
compute b0 = ib0
compute lambda = ilambda
compute alpha = ialpha
compute beta = ibeta
```

D3. Estimate the equation by ML

```
maximize(method=bhhh,iterations=100) logl

Estimation by BHHH
Iterations Taken    10
Usable Observations     98      Degrees of Freedom    93
Function Value                      63.12195118

   Variable                     Coeff       Std Error      T-Stat     Signif
*******************************************************************************
1.  SIGMASQ                  0.1014280561 0.0136226552      7.44554  0.00000000
2.  B0                       2.9665559411 0.6541677030      4.53486  0.00000576
3.  LAMBDA                   0.0135658878 0.0041292501      3.28532  0.00101868
4.  ALPHA                    0.6446537726 0.0540997724     11.91602  0.00000000
5.  BETA                     0.2845499324 0.0694294595      4.09840  0.00004160
```

Once again, we have standard errors for each of the estimated parameters and the entire variance-covariance matrix of the parameter vector.

IMPORTANT NOTE: The specific instructions needed to implement NLLS and ML will differ from software to software, but the following three elements will always be there: (i) specification of the nonlinear equation (or likelihood function), (ii) setting of intitial values for the parameters to be estimated (otherwise 0 is assumed, which may be a very bad guess), and (iii) actual minimization (maximization) of the sum of squares (log likelihood).