



**ECONOMETRICS FOR RESEARCHERS, 2015  
(MODULE 1: MARCH 16-20, 2015)**

**NONLINEAR REGRESSION**

Afees A. Salisu

Contacts: Department of Economics, Federal University of Agriculture,  
Abeokuta & Centre for Econometric & Allied Research (CEAR), University  
of Ibadan. Email: [adebare1@yahoo.com](mailto:adebare1@yahoo.com); [aa.salisu@cear.org.ng](mailto:aa.salisu@cear.org.ng).  
Mobile: +234 (0)803 471 1769.

# Nonlinear Regression

Afees A. Salisu<sup>1</sup>

## Learning Outcomes

This lecture deals with the analysis of nonlinear regression models covering issues relating to specification, estimation and diagnostics. At the end of the lecture, participants should be able to:

- Specify nonlinear regression models
- Estimate nonlinear regression models using Stata
- Analyze the regression results

## How can we identify a nonlinear regression model?

Recall that in a linear regression model, the regression function is a linear function of the unknown parameters; however, in a nonlinear regression model, the regression function is expressed as a nonlinear function of the unknown parameters. In other words, a regression model can be either linear or nonlinear in the parameters. These distinctions are important for the choice of estimation technique(s). How? A model that is linear in the parameters can be estimated with Linear Least Squares technique (i.e. the Ordinary Least Squares (OLS)) while a model that is nonlinear in the parameters must be estimated with nonlinear least squares (the procedure is described in subsequent sections). The next concern certainly would be: How do we identify a model that is linear / nonlinear in the parameters?

Suppose that we have the regression specification:

$$y_i = f(x_i, \beta) + \varepsilon_i$$

Where  $f$  is a general function of the explanatory variables  $x_i$  and the parameters  $\beta$ . The model is said to be linear in the parameters if the derivatives of  $f$  with respect to the parameters are not expressed as functions of  $\beta$ ; otherwise, it is nonlinear. That is, if the derivatives are functions of  $\beta$ , we say the model is nonlinear in the parameters.

---

<sup>1</sup> Contacts: Department of Economics, Federal University of Agriculture, Abeokuta & Centre for Econometric & Allied Research (CEAR), University of Ibadan. Email: [adebare1@yahoo.com](mailto:adebare1@yahoo.com); [aa.salisu@cear.org.ng](mailto:aa.salisu@cear.org.ng). Mobile: +234 (0)803 471 1769.

In a broader perspective, we can categorize a regression model in econometrics into four:

1. Linearity in the variables
2. Linearity in the parameters
3. Nonlinearity in the variables
4. Nonlinearity in the parameters

Out of these four categories, only the nonlinearity in the parameters is considered to be nonlinear. *Why?* The table below describes these categories:

**Table 1: LRM vs. NLRM**

Linear in the Variables?	Linear in the Parameters?	
	YES	NO
YES	LRM	NLRM
NO	LRM	NLRM

Note: LRM = Linear Regression Model

NLRM = Nonlinear Regression Model

Let us consider the following examples:

$$y_t = \beta_1 + \beta_2 x_t \tag{1}$$

$$y_t = \beta_1 + \beta_2 x_t^2 \tag{2}$$

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_t^2 \tag{3}$$

$$y_t = e^{\beta_1 + \beta_2 x_t} \tag{4}$$

$$y_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3} \tag{5}$$

$$y_t = \beta_1 e^{-\beta_2 x_t} \tag{6}$$

$$y_t = \frac{1}{(\beta_1 + \beta_2 x_t)^{\beta_3}} \tag{7}$$

$$y_t = \frac{1}{\beta_1 + \beta_2 x_t + \beta_3 x_t^2} \tag{8}$$

The unknown parameters to be estimated are  $\beta_1$ ,  $\beta_2$  and  $\beta_3$ .

From the examples above, equations (1)-(3) are linear since the derivatives of  $y_t$  with respect to the parameters cannot be expressed as functions of the parameters. Therefore,

these equations can be estimated with the OLS technique (assuming that all the other assumptions of OLS are not violated). However, equations (4)-(7) are nonlinear as the derivatives of  $y_t$  with respect to the parameters are dependent on the parameters themselves. In this case, the nonlinear least squares approach must be used to estimate the equations.

### **Linearizable Nonlinear Models**

Actually, in some instances, it may be possible to transform a nonlinear regression model using appropriate transformations of the dependent variable, explanatory variable(s), the parameters, or any combination of these, such that the transformed function becomes linear in the unknown parameters. If the transformed model satisfies the assumptions of Classical Linear Regression Models, then the OLS technique can be applied to estimate the model.

As an example, consider equations (4), (5) & (6) above. By taking the natural logarithm of both sides of the model, we get:

$$\ln(y_t) = \ln(\beta_1) + \beta_2 \ln(L_t) + \beta_3 \ln(K_t) \quad (9)$$

Or

$$y_t^* = \beta_1^* + \beta_2 L_t^* + \beta_3 K_t^*$$

where  $y_t^* = \ln(y_t)$ ;  $\beta_1^* = \ln(\beta_1)$   $L_t^* = \ln(L_t)$ ; &  $K_t^* = \ln(K_t)$

$$\ln(y_t) = \beta_1 + \beta_2 x_t \quad (10)$$

Or

$$y_t^* = \beta_1 + \beta_2 x_t$$

where  $y_t^* = \ln(y_t)$

$$\ln(y_t) = \ln(\beta_1) - \beta_2 x_t \quad (11)$$

Or

$$y_t^* = \beta_1^* + \beta_2^* x_t$$

Where  $y_t^* = \ln(y_t)$ ;  $\beta_1^* = \ln(\beta_1)$ ; &  $\beta_2^* = -\beta_2$ .

Equations (9), (10) & (11) are the linearized functions of equations (4), (5) & (6) respectively and they are considered to be approximately linear. Thus, the transformed

functions can be estimated using OLS. In essence, we can use the underlying procedure for a linear regression analysis to make approximate inferences about the parameters  $\beta_1^*$  and  $\beta_2^*$  which will in turn lead to inferences about  $\beta_1$  and  $\beta_2$ , the parameters of interest in the original model.

With the exception of nonlinearity in the unknown parameters, the nonlinear regression shares the same assumptions with the linear regression. In essence, if the researcher is confident that the transformed equations satisfy assumptions for straight line regression (at least approximately), then appropriate inferences can be drawn about the transformed and original estimates. A number of linearizable models including their linear representations are given in the table below.

**Table 2: Examples of Linearizable Functions**

Original Regression Function	Linearizing Transformation	Condition(s) for Transformation of $y_t / x_t$
$y_t = \frac{1}{\beta_1 - \beta_2 x_t}$	$\ln(y_t) = \beta_1 - \beta_2 x_t$ $y_t^* = \beta_1^* + \beta_2^* x_t$ Where $y_t^* = \ln(y_t)$ ; $\beta_1^* = \beta_1$ ; $\beta_2^* = -\beta_2$	$y_t > 0$
$y_t = \beta_1 x_t^{\beta_2}$	$\ln(y_t) = \ln(\beta_1) + \beta_2 \ln(x_t)$ $y_t^* = \beta_1^* + \beta_2^* x_t^*$ where $y_t^* = \ln(y_t)$ ; $x_t^* = \ln(x_t)$ ; $\beta_1^* = \ln(\beta_1)$ ; $\beta_2^* = \beta_2$	$y_t > 0$ and $x_t > 0$
$y_t = \beta_1 \beta_2^{x_t}$	$\ln(y_t) = \ln(\beta_1) + x_t \ln(\beta_2)$ $y_t^* = \beta_1^* + \beta_2^* x_t$ where $y_t^* = \ln(y_t)$ ; $\beta_1^* = \ln(\beta_1)$ ; $\beta_2^* = \ln(\beta_2)$	$y_t > 0$
$y_t = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x_t)}}$	$\ln\left(\frac{y_t}{1 - y_t}\right) = \beta_1 + \beta_2 x_t$ $y_t^* = \beta_1^* + \beta_2^* x_t$ where $y_t^* = \ln\left(\frac{y_t}{1 - y_t}\right)$ ; $\beta_1^* = \beta_1$ ; $\beta_2^* = \beta_2$	$0 < y_t < 1$

In a situation where the researcher is not confident that the assumptions of a linear regression model hold (even approximately) for the transformed equations, it may be necessary to follow the estimation procedure for nonlinear regressions. However, the transformed equations may still be estimated and the estimates obtained can be used as starting values for the nonlinear regression analysis.

The next concern now is: How can we estimate the unknown parameters of a nonlinear regression model? This question is answered in sections that follow.

### **Estimation of Nonlinear Regression Parameters**

Usually, it is assumed that the form of the nonlinear regression function is known except that the parameters are unknown. There are different nonlinear models in the literature and it is the responsibility of the researcher/modeller to identify the appropriate model based on the working knowledge about the subject matter being analyzed. In any case, there are relevant diagnostics that can be used to evaluate the performance of the chosen model.

A popular method for estimating the unknown parameters in a nonlinear regression function is the method of least squares. However, unlike the linear regression, the estimation of parameters in the nonlinear case usually requires the use of *iterative methods* which can be implemented in any of the commonly used statistical software packages like Stata, Eviews and R.

For the purpose of this workshop, we shall be using Stata to demonstrate the estimation of nonlinear models. Like any other statistical package, Stata requires that a set of starting values or initial guesses for all the parameters must be supplied in addition to the data. *Why ?* It is often helpful if the starting values are close to the actual least squares estimates. However, we may not have such initial guesses. Sometimes, we can obtain good initial estimates or at least the correct signs based on theoretical considerations or by plotting the sample data.

#### **Study Problem:**

- ✚ *Determining the Elasticity of Substitution (EOS) between Capital & Labour at the Firm Level*

We may wish to fit the CES production function to compute the EOS:

$$\ln Q_i = \beta_0 - \frac{1}{\rho} \ln \{ \delta K_i^{-\rho} + (1 - \delta) L_i^{-\rho} \} + \varepsilon_i \quad (12)$$

Where  $\ln Q_i$  is the log of output for firm  $i$ ,  $K_i$  and  $L_i$  are firm  $i$ 's capital and labor usage, respectively; and  $\varepsilon_i$  is a regression error term. Recall, we have to provide a set of initial values for all the parameters before we proceed to estimation. Since  $\rho$  appears in the denominator of a fraction, zero is not a feasible initial value; for a CES production function,  $\rho = 1$  may be a reasonable choice. Setting  $\delta = 0.5$  implies that labor and capital have equal impacts on output, which may also be a reasonable choice for an initial value.

### Box 1

#### Stata Commands for Estimation:

use <http://www.stata-press.com/data/r10/production.dta>

```
nl (lnoutput={b0} - 1/{rho=1}*ln({delta=0.5}*capital^(-1*{rho})) + (1 - {delta})*labor^(-1*{rho})))
```

**Note the following about the Stata commands for nonlinear regression<sup>2</sup>:**

- *nl* is used to fit the nonlinear regression function.
- The *nl* command line involves the use of a *substitutable expression*. In other words, the function enters directly on the command line.
- The variable names specified in the model must correspond with the data file.
- The parameter names must be provided but we may decide to use the conventional parameter names as specified in the original function.
- Note that in the substitutable expression, the parameters to be estimated with their initial values are enclosed in braces - { }
- The parameter name is followed by an equal sign and the initial value
- The entire equation is enclosed in parentheses - ( )

---

<sup>2</sup> Note that the Stata commands provided here are intended for the average users. The advanced commands may be explored by the non-average users.

- Because the parameters are enclosed in braces, *nl* knows that they are parameters in the model.
- *nl* will initialize the parameters to zero by default (i.e. if the initial values are not supplied). See `b0` above.
- It is important to point out here that many nonlinear functions cannot be evaluated when some parameters are zero, so in those cases, specifying alternative initial values is crucial.
- Where a parameter appears in the nonlinear function multiple times (see the CES function above), it is not necessary to repeat the initial value for that parameter. In other words, the initial value should be supplied once to save time. However, it should not be supplied at all if the initial value is set to zero.

Let us write the stata commands for the estimation of the following nonlinear functions:

1.  $y_i = \beta_0 \left(1 - e^{-\beta_1 x_i}\right) + \varepsilon_i$
2.  $y_i = \beta_0 \left(1 - e^{-(\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i})}\right) + \varepsilon_i$

**Answer:**

1. `nl (y = {b0=1}*(1-exp(-1*{b1}*x)))`

## Note that `b0` cannot be initialized to zero. Why?

2. `nl (y={b0=1}*(1-exp(-1*({b1}*x1+{b2}*x2+{b3}*x3))))`

## Note that `b0` cannot be initialized to zero. Why?

Since the nonlinear function in 2 contains a linear combination of variables, the Stata command can also be expressed as:

`nl (y = {b0=1}*(1 - exp(-1*{xb: x1 x2 x3})))`

`xb` is the name given to the linear combination and *nl* names the corresponding parameters as `xb_x1` , `xb_x2` , and `xb_x3`, in the estimation output and initialize them to zero.

Let us go back to our CES function. The results obtained after implementing the command are provided below:

(obs = 100)

Iteration 0: residual SS = **29.38631**  
 Iteration 1: residual SS = **29.36637**  
 Iteration 2: residual SS = **29.36583**  
 Iteration 3: residual SS = **29.36581**  
 Iteration 4: residual SS = **29.36581**  
 Iteration 5: residual SS = **29.36581**  
 Iteration 6: residual SS = **29.36581**  
 Iteration 7: residual SS = **29.36581**

Source	SS	df	MS			
Model	<b>91.1449924</b>	<b>2</b>	<b>45.5724962</b>	Number of obs =	<b>100</b>	
Residual	<b>29.3658055</b>	<b>97</b>	<b>.302740263</b>	R-squared =	<b>0.7563</b>	
				Adj R-squared =	<b>0.7513</b>	
				Root MSE =	<b>.5502184</b>	
Total	<b>120.510798</b>	<b>99</b>	<b>1.21728079</b>	Res. dev. =	<b>161.2538</b>	

  

Input	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/b0	<b>3.792158</b>	<b>.099682</b>	<b>38.04</b>	<b>0.000</b>	<b>3.594316</b>	<b>3.989999</b>
/rho	<b>1.386993</b>	<b>.472584</b>	<b>2.93</b>	<b>0.004</b>	<b>.4490443</b>	<b>2.324941</b>
/delta	<b>.4823616</b>	<b>.0519791</b>	<b>9.28</b>	<b>0.000</b>	<b>.3791975</b>	<b>.5855258</b>

Parameter b0 taken as constant term in model & ANOVA table

### Interpretation of Results

It is important to emphasize here that it is the responsibility of the researcher/analyst to:

- (1) identify the appropriate nonlinear function to use for analysis
- (2) understand the theoretical underpinnings for the specified function
- (3) use the information in (2) to interpret the estimation results.

In essence, our interpretation of the results is specific to the CES function and may not be generalized. For instance,  $\delta$  and  $(1-\delta)$  represent the parameter shares of capital and labor respectively and are used to capture the impacts of these input factors on output. As shown in the results, delta is slightly below 0.5. Can we still assume that the two input factors have the same share? This can be verified by conducting a formal test (see Box 2). The result shows that the two input factors have the same share in the output. In essence, the estimated input share is not statistically different from 0.5. This is an indication that both labour and capital play a very significant role in the production process.

### Box 2

**Stata command:**

```
testnl _b[/delta]=0.5
```

Or

```
testnl _coef[/delta]=0.5
```

**Result:**

```
(1) [delta]_cons = .5  
     F( 1, 97) = 0.12
```

In line with the spirit of this model, we can also compute the Elasticity of Substitution (EOS) and the formula is given as:  $\sigma = 1/(1 + \rho)$ . The EOS expresses the relative change in the factor proportion due to the relative change in the marginal rate of technical substitution when output is held constant. In other words, it represents how factor income shares change as the ratio of the factors varies.

We can use the *nlcom* command in Stata to compute the EOS. The advantage of using this command over other alternative approaches is that it computes the EOS and also determines its significance. Please check the box below:

### Box 3

**Stata command:**

```
nlcom (1/(1 + _b[/rho]))
```

**Result:**

```
_nl_1: 1/(1 + _b[/rho])
```

Inoutput	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_nl_1	.4189372	.0829424	5.05	0.000	.2543194	.583555

### Decision Criteria for the EOS:

- If the computed EOS is less than 1, it implies that the inputs are complements
- If the computed EOS is greater than 1, then the inputs are considered as substitutes

### The Significance of EOS for the Choice of Production Function

- If EOS is not different from 1; Cobb-Douglas Production Function is appropriate
- If EOS is not different from 0, Leontief Production Function is preferable
- Otherwise (i.e, if the EOS is neither 0 nor 1), the CES production function is considered to give the best fit.

The result shows that the EOS is less than 1; thus, validating the complementarity between the two input factors. Nonetheless, we can still test whether the computed EOS is statistically different from 0 and 1. Recall that the EOS for Cobb-Douglas is 1 and therefore, if the EOS is not different from 1 then it may be more appropriate to fit the latter. See Box 4 below.

#### Box 4

##### Stata Command:

```
nlcom (1/(1 + _b[/rho])), post
```

```
testnl _b[_nl_1]=1
```

##### Result:

```
_nl_1: 1/(1 + _b[/rho])
```

lnoutput	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_nl_1	.4189372	.0829424	5.05	0.000	.2543194 .583555

```
. test _b[_nl_1]=1
```

```
( 1)  _nl_1 = 1
```

```
F( 1, 97) = 49.08  
Prob > F = 0.0000
```

The result rejects the Cobb-Douglas production function. We also verify the plausibility of another competing model which is the Leontief Production Function. It assumes that the EOS is zero (0). The test is presented in Box 5.

#### **Box 5**

##### **Stata Command:**

```
testnl _b[_nl_1]=0
```

##### **Result:**

```
( 1) _nl_1 = 0
```

```
F( 1, 97) = 25.51  
Prob > F = 0.0000
```

Overall, the results strongly support the choice of CES production function and consequently reject the Cobb-Douglas and Leontief production functions for modelling output at the firm level.

#### **Class Exercise**

- ✚ Fit the Cobb-Douglas production function below using the same dataset and interpret your results:

$$\ln Q_i = \beta_0 + \beta_1 \ln L_i + \beta_2 K_i + \varepsilon_i$$

#### **Diagnostics and Inferences**

Apart from the theoretical justification for the choice of CES production function, there are model selection criteria that can be used when confronted with related competing models. These criteria are AIC and BIC and the model with the least values of these criteria is preferable. This is demonstrated in Box 6 below.

### Box 6

#### Stata command

Recall the estimation command for CES (since we are performing the test on the model). However, if the model being considered is the most recently fitted model, then recalling is not necessary.

```
nl (lnoutput = {b0} - 1/{rho=1}*ln({delta=0.5}*capital^(-1*{rho}) + (1 - {delta})*labor^(-1*{rho})))
```

#### AIC & SIC Command:

```
estat ic
```

#### Results:

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	100	.	-80.62689	3	167.2538	175.0693

Note: N=Obs used in calculating BIC; see [\[R\] BIC note](#)

You may want to compare the AIC and BIC values of CES model with that of Cobb-Douglas. See Box 7 below:

### Box 7

#### Stata commands:

Recall the estimation command for Cobb-Douglas:

```
nl (lnoutput={b0=1}+{b1=0.5}*ln(capital)+{b2=0.5}*ln(labor))
```

Then the AIC & BIC Command:

```
estat ic
```

```
. estat ic
```

Model	obs	ll(null)	ll(model)	df	AIC	BIC
.	100	.	-102.6954	3	211.3909	219.2064

Note: N=obs used in calculating BIC; see [\[R\] BIC note](#)

From Boxes 6 and 7, we can see that the SIC and AIC values of CES are lower than Cobb-Douglas. Therefore, the CES is expected to give a better fit. This further corroborates the theoretical justification for CES. In addition, the root mean square error (rmse) reported also strengthens the choice of CES. See the regression outputs for CES and Cobb-Douglas.

Another important inference relates to the behaviour of nonlinear predictions after estimation. This involves generating the fitted values for the estimated CES function which can be compared with the actual values. The idea is to inspect the predictive ability of the function. See the box below.

#### **Box 8**

##### **Stata commands:**

*Ensure that the model under examination is the most recently fitted model in stata. Otherwise, we recall the regression.*

```
predictnl double lnoutputyhat = predict()
```

*Note that the result is saved as lnoutputyhat.*

To view it, type:

```
tab lnoutputyhat
```

We can compare the summary statistics of both the fitted & actual:

```
summarize lnoutput lnoutputyhat
```

The results show that the predictive ability of the model is quite high judging by the root mean square error as well as the summary statistics. The adjusted R-Square here is interpreted the same way as the linear regression model. In other words, the adjusted coefficient of determination measures the explanatory power of the model. On the basis of the estimation results of the CES function, we find that labor and capital account for about 75 percent of the total variations in output.

- Participants/ Readers are advised to practise with other examples of nonlinear functions using the estimation procedure highlighted above. We hope to update

this presentation as soon as possible to include more advanced and recent nonlinear models