ES1004 Econometrics by Example

Lecture 2: Functional Forms in Regression

Dr. Hany Abdel-Latif

Swansea University, UK

Gujarati textbook, second edition

May 7th 2016

Linearity I

- an equation is linear in the variables if plotting the function in terms of X and Y generates a straight line

$$Y = \beta_0 + \beta_1 X + u \qquad \text{linear in variables}$$

$$Y = \beta_0 + \beta_1 X^2 + u \qquad \text{not linear in variables}$$

Linearity II

- an equation is linear in the coefficients only if the coefficients appear in their simplest form i.e.,
  - not raised to any powers (other than one)
  - not multiplied or divided by other coefficients
  - do not include some sort of function (like logs or exponents)

$Y = \beta_0 + \beta_1 X + u$ linear in coefficients

$Y = \beta_0 + \beta_1 X^2 + u$ linear in coefficients

$Y = \beta_0 + X^{\beta_1}$ not linear in coefficients

OLS Estimation

- OLS method is restricted to models that are linear in the parameters

$Y = \beta_0 + \beta_1 X^2 + u$ can be estimated by OLS

$Y = \beta_0 + X^{\beta_1}$ cannot be estimated by OLS

- models that are nonlinear in parameters can be estimated using nonlinear least squares
  - an iterative procedure which searches for the parameter value(s) which minimise the RSS of the model

مجتمع الاقتصاد
THE ECONOMIC SOCIETY

Cobb-Douglas Production Function I

$$Q_i = \beta_1 L_i^{\beta_2} K_i^{\beta_3} \quad (1)$$

- can be transformed into a linear model by taking natural logs of both sides

$$\ln Q_i = \ln \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i \quad (2)$$

- adding the error term u_i, we obtain the following LRM

$$\ln Q_i = \ln \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + u_i \quad (3)$$

- eq. 3 is known as double-log, log-log or constant elasticity model
  - because both the regressand and regressors are in the log from

Cobb-Douglas Production Function II

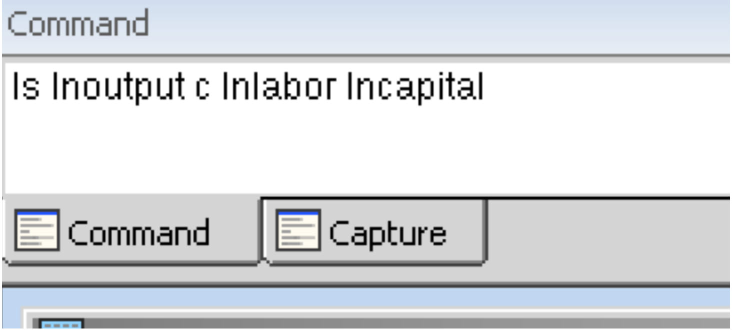
$$\ln Q_i = \ln \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + u_i$$

- the slope coefficients can be interpreted as partial elasticities
  - holding other variables constant
- returns to scale of CD function
  - if $(\beta_2 + \beta_3) = 1 \rightarrow$ constant returns to scale
  - if $(\beta_2 + \beta_3) > 1 \rightarrow$ increasing returns to scale
  - if $(\beta_2 + \beta_3) < 1 \rightarrow$ decreasing returns to scale

Example: CD function for USA

- table 2.1 cross section data for 51 states for 2005
  - output [value added, thousands of dollars]
  - labor [worker hours, in thousands]
  - capital [capital expenditure, in thousands of dollars]

Example: Estimate in EViews



Example: EViews Output

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNOUTPUT
Method: Least Squares
Date: 05/06/16 Time: 16:21
Sample: 1 51
Included observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.887600	0.396228	9.811519	0.0000
LNLABOR	0.468332	0.098926	4.734170	0.0000
LNCAPITAL	0.521279	0.096887	5.380281	0.0000

R-squared	0.964175	Mean dependent var	16.94139
Adjusted R-squared	0.962683	S.D. dependent var	1.380870
S.E. of regression	0.266752	Akaike info criterion	0.252027
Sum squared resid	3.415517	Schwarz criterion	0.365664
Log likelihood	-3.426697	Hannan-Quinn criter.	0.295451
F-statistic	645.9317	Durbin-Watson stat	1.946387
Prob(F-statistic)	0.000000		

Example: Inference & Interpretation

- hypothesis testing & goodness of fit
  - all regression coefficients (i.e., elasticities) are individually statistically highly significant (quite low p values)
  - according to $F - statistic$ collectively both factors inputs [labour and capital] are statistically significant
  - quite hight R^2 [unusual for cross-section data!]
- if we increase labour input by 1%, on average, output goes up by about 0.47% [holding the capital input constant]
- if we increase the capital input by 1%, on average, the output increases by about 0.52% [holding the labour input constant]
- $\beta_2 + \beta_3 = 0.9896 \rightarrow$ constant returns to scale in 2005

Growth Models I

- the rate of growth of real gdp

$$RGDP_t = RGDP_{1960}(1 + r)^t \qquad (4)$$

- can be transformed into a linear model by taking natural logs of both sides

$$\ln RGDP_t = \ln RGDP_{1960} + t \ln(1 + r) \qquad (5)$$

- let $\beta_1 = RGDP_{1960}$, $\beta_2 = \ln(1 + r)$, this can be written as

[see next slide]

Growth Models II

- this can be written as

$$\ln RGDP_t = \beta_1 + \beta_2 t \tag{6}$$

- adding the error term u_t, we obtain

$$\ln RGDP_t = \beta_1 + \beta_2 t + u_t \tag{7}$$

- note that the regressor is "time", which takes values of 1,2,. . . , T
- called a semilog or log-lin model

Growth Models III

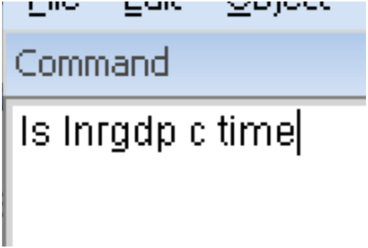
$$\ln RGDP_t = \beta_1 + \beta_2 t + u_t$$

$$\beta_2 = \frac{\textit{relative change in regressand}}{\textit{absolute change in regressor}} \quad (8)$$

- β_2
  - multiply it by 100 to compute the percentage change, or the growth rate
  - known as the semi-elasticity of the regressand with respect to the regressor [or an instantaneous growth rate]

Example: Growth Rate Real GDP

- table 2.5 USA real gdp [adjusted for inflation] 1960-2007



Example: EViews Output

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNRGDP
Method: Least Squares
Date: 05/06/16 Time: 23:21
Sample: 1 48
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.875662	0.009759	807.0064	0.0000
TIME	0.031490	0.000347	90.81649	0.0000

R-squared	0.994454	Mean dependent var	8.647156
Adjusted R-squared	0.994333	S.D. dependent var	0.442081
S.E. of regression	0.033280	Akaike info criterion	-3.926967
Sum squared resid	0.050947	Schwarz criterion	-3.849001
Log likelihood	96.24722	Hannan-Quinn criter.	-3.897504
F-statistic	8247.634	Durbin-Watson stat	0.347739
Prob(F-statistic)	0.000000		

Example: Interpretation

- over the period of 1960-2007, the USA's real GDP had been increasing at the rate of 3.15% per year
- the growth rate is statistically significant
- what is the interpretation of the intercept?
  - take anti-log (7.8756) = 2632.27 which is the estimated value of real GDP in 1960

Regressors in Log Form

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \tag{9}$$

$$\beta_2 = \frac{absolute\ change\ in\ Y}{change\ in\ ln\ X} = \frac{\Delta Y}{\Delta X / X} \tag{10}$$

- a change in the log of a number is a relative change, or percentage change, after multiplying by 100
  - β_2 is the absolute change in Y responding to a percentage [or relative] change in X
  - if X increases by 100%, predicted Y increases by B_2 units

Example: Engel Expenditure Functions

- the share of expenditure on food decreases as total expenditure increases
  - table 2.8 data on food consumed at home `Exfood` and total household expenditure `Expend`
  - both in dollars for 869 US households in 1995
  - regress the share of food expenditure `sfdho` on the log of total expenditure `lnexpend`

Example: EViews output

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SFDHO
Method: Least Squares
Date: 05/07/16 Time: 13:31
Sample: 1 869
Included observations: 869

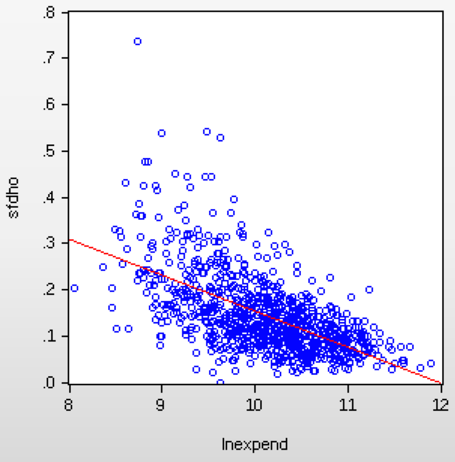
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.930387	0.036367	25.58359	0.0000
LNEXPEND	-0.077737	0.003591	-21.64823	0.0000

R-squared	0.350876	Mean dependent var	0.144736
Adjusted R-squared	0.350127	S.D. dependent var	0.085283
S.E. of regression	0.068750	Akaike info criterion	-2.514368
Sum squared resid	4.097984	Schwarz criterion	-2.503396
Log likelihood	1094.493	Hannan-Quinn criter.	-2.510170
F-statistic	468.6456	Durbin-Watson stat	1.968386
Prob(F-statistic)	0.000000		

Example: Interpretation

- estimated coefficients are individually highly statistically significant
- if total expenditure increases by 1%, on average, the share of expenditure on food goes down by about 0.0008 units
  - divide the slope coefficient by 100
  - supporting engel hypothesis
- or if total expenditure increases by 100%, on average, the share of expenditure on food goes down by about 0.8 units

What Data Tell



Inverse Model

$$Y_i = \beta_1 + \beta_2(\frac{1}{X_i}) + u_i \quad (11)$$

- note that
  - as X increases indefinitely, the term $\beta_2(\frac{1}{X_i})$ approaches zero and Y approaches the limiting or asymptotic value B_1
  - the slope is

$$\frac{dY}{dX} = -\beta_2(\frac{1}{X^2})$$

- if β_2 is positive, the slope is negative throughout
- if β_2 is is negative the slope is positive throughout

مجتمع الاقتصاد
THE ECONOMIC SOCIETY

Example: Food Expenditure Revisited

$$sfdho = \beta_1 + \beta_2 \frac{1}{expend_i} + u_i \qquad (12)$$

-

Example: EViews Output

Dependent Variable: SFDHO
Method: Least Squares
Date: 05/07/16 Time: 16:29
Sample: 1 869
Included observations: 869

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.077263	0.004012	19.25950	0.0000
EXPEND_REC	1331.338	63.95713	20.81610	0.0000

R-squared	0.333236	Mean dependent var	0.144736
Adjusted R-squared	0.332467	S.D. dependent var	0.085283
S.E. of regression	0.069678	Akaike info criterion	-2.487556
Sum squared resid	4.209346	Schwarz criterion	-2.476584
Log likelihood	1082.843	Hannan-Quinn criter.	-2.483357
F-statistic	433.3100	Durbin-Watson stat	1.997990
Prob(F-statistic)	0.000000		

Example: Interpretation

- both regression coefficients are statistically highly significant
- the intercept
  - if total expenditure increases indefinitely, the share of food in total expenditure will eventually settle down to about 8%
- slope coefficient β_2
  - positive suggesting that the rate of change of sfdho with respect to total expenditure will be negative throughout

Quadratic Function

- the following regression predicts GDP is an example of a quadratic function, or more generally, a second-degree polynomial in the variable `time`

$$RGDP_t = A_1 + A_2 time + A_3 time^2 + u_t \tag{13}$$

- the slope is nonlinear and equal to

$$\frac{dRGDP}{time} = A_2 + 2A_3 time$$

MODEL	FORM	SLOPE	ELASTICITY
		$(\frac{dY}{dX})$	$\frac{dY}{dX}\cdot\frac{X}{Y}$
Linear	$Y = B_1 + B_2 X$	B_2	$B_2(\frac{X}{Y})$
Log-linear	$\ln Y = B_1 + \ln X$	$B_2(\frac{Y}{X})$	B_2
Log-lin	$\ln Y = B_1 + B_2 X$	$B_2(Y)$	$B_2(X)$
Lin-log	$Y = B_1 + B_2 \ln X$	$B_2(\frac{1}{X})$	$B_2(\frac{1}{Y})$
Reciprocal	$Y = B_1 + B_2(\frac{1}{X})$	$-B_2(\frac{1}{X^2})$	$-B_2(\frac{1}{XY})$

مجتمع الاقتصاد
THE ECONOMIC SOCIETY