Lecture 7

FUNCTIONAL FORMS OF REGRESSION MODELS

Gujarati D., Basic Econometrics, (2004)
Chapter 6, p. 175 – 201

Functional forms of regression models

We will discuss the following regression models:

1. The log – log model
2. Semilog models: lin-log and log-lin models
3. Reciprocal models

- The special features of each model , when they are appropriate, and how they are estimated

The LOG – LOG model

Consider the following model:

$$Y_i = \beta_1 X^{\beta_2} e^{u_i} \qquad (6.5.1)$$

which may be expressed alternatively as

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i \qquad (6.5.2)$$

where ln=natural log (i.e. Log to the base e, and where e = *2.718)*

To obtain model which is linear in the parameters, denote α = lnβ_1

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \qquad (6.5.3)$$

The LOG – LOG model

If the assumptions of the classical linear regression model are fulfilled, the parameters of (6.5.3) can be estimated by the OLS method by letting

$$Y_i^* = \alpha + \beta_2 X_i^* + u_i, \text{ where}$$
$$Y_i^* = \ln Y_i \quad \text{and} \quad X_i^* = \ln X_i$$

The OLS estimators $\hat{\alpha}$ and $\hat{\beta}_2$ obtained will be best linear unbiased estimators of α and β_2 respectively.

The LOG – LOG model

$$\ln \hat{Y}_i = \hat{\alpha} + \hat{\beta}_2 \ln X_i \,, \qquad \hat{u}_i = \ln Y_i - \ln \hat{Y}_i$$

One attractive feature of the log-log model, which has made it popular in applied work, is that the slope coefficient β_2 measures elasticity of Y with respect to X:

β_2 shows the percentage change in Y for a given (small) percentage change in X.

The LOG – LOG model

If Y represents the quantity of commodity demanded and X its unit price, β_2 measures the price elasticity of demand, a parameter of considerable economic interest.

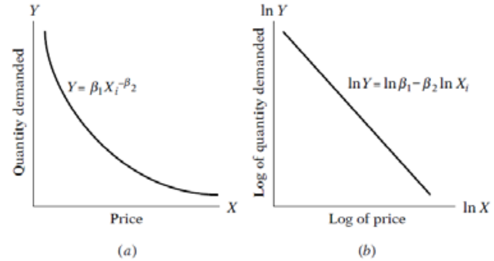
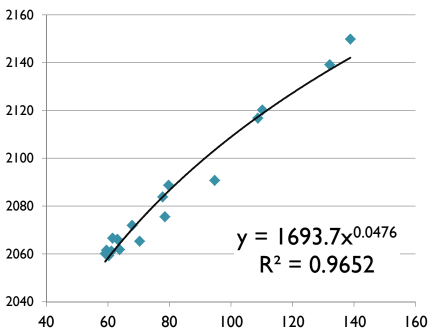


FIGURE 6.3 Constant-elasticity model.

Example

	Purchase price of wheat (lats per t)	Production of wheat (thsd t)
1995	70.34	2065.34
1996	94.72	2090.72
1997	78.57	2075.57
1998	63.84	2061.84
1999	60.3	2059.3
2000	61.19	2061.19
2001	59.14	2060.14
2002	59.58	2061.58
2003	63.1	2066.1
2004	67.91	2071.91
2005	61.56	2066.56
2006	77.83	2083.83
2007	132.11	2139.11
2008	108.76	2116.76
2009	79.75	2088.75
2010	110.21	2120.21
2011	138.82	2149.82



Increase in purchase price of wheat by 1% leads to about 0.0479% change in production of whaet.
Data source: www.csb.gov.lv

LOG -LIN model

- PRF in case of log-lin model:

$$Y_i = \beta_1 e^{\beta_2 X_i} e^{u_i}$$

where β_1, β_2 – population regression coefficients, u_i – residual term.

- For sample data: $\hat{Y}_i = \hat{\beta}_1 e^{\hat{\beta}_2 X_i} e^{\hat{u}_i}$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ – OLS estimates for unknown population parameters, $\hat{u}_i$ – estimate for residual term

- Take log from both sides:

$$\ln Y_i = \ln\left(\beta_1 e^{\beta_2 X_i} e^{u_i}\right) = \ln \beta_1 + \beta_2 X_i + u_i$$

- β_2 shows: if value of X increases by 1 unit, Y mean increases by β_2 *100%

LOG-LIN model

How to measure the Growth rate: application of The LOG-LIN model

Economists, businesspeople, and governments are often interested in finding out the rate of growth of certain economic variables, such as population, GDP, money supply, employment, productivity, and trade deficit.

Let have values $Y_{0;} Y_{1;} \ldots, Y_{t,}$ where t - time

Suppose that we want to estimate compound (i.e., over time) rate of growth of Y

How to measure the Growth rate

- r is the compound (i.e., over time) rate of growth of Y:

$$Y_t = Y_0 (1+r)^t,$$

where Y_0 is the initial value of Y.

- Taking log from both sides:

$$\ln Y_t = \ln\left(Y_0 (1+r)^t\right) = \ln Y_0 + t \cdot \ln(1+r)$$

Denote: $\ln Y_0 = \ln \beta_1$ un $\beta_2 = \ln(1+r)$, so we obtain log-lin model

- β_2 shows: $\beta_2 = \ln(1+r) = d(\ln Y)/dX = (1/Y)\ dY/Y/dt = (dY/Y)/dt$

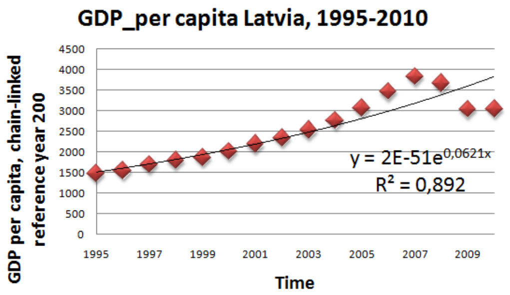
  or $\beta_2 \approx (\Delta Y/Y)/\Delta t$
- $\Delta Y/Y \approx \beta_2 \Delta t$ and $100 \cdot \beta_2$ shows mean change (percentage) over the given time period
- Or 100 times β_2 gives the growth rate in Y.

Example

GROSS DOMESTIC PRODUCT_ Latvia

Chain-linked reference year 2000

	Per capita, lats
1995	1472
1996	1543
1997	1688
1998	1786
1999	1859
2000	2002
2001	2180
2002	2337
2003	2519
2004	2753
2005	3061
2006	3454
2007	3819
2008	3657
2009	3028
2010	3039



Example: interpretation

- The coefficient of the trend variable in the growth model β_2 gives the instantaneous (at a point of time) rate of growth and not the compound (over a period of time) rate of growth.

- But the latter can be easily found from model by taking antilog of the estimated β_2 and subtracting 1 from it and multiplying the difference by 100.
- Thus, for our illustrative example:

$$\hat{\beta}_2 = 0{,}0621$$

therefore

$$\exp(0{,}0621) - 1 = 0{,}0641$$

Thus, in our example the compound rate of growth on GDP per capita in Latvia was about 6.41 % per year.

LIN-LOG model

PRF is following:

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

Sample regression function is:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \ln X_i \,, \quad \hat{u}_i = Y_i - \hat{Y}_i$$

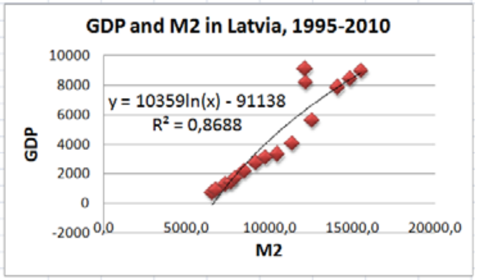
Slope coefficient:

$\beta_2 = dY/d(\ln X) = dY/(dX/X)$ or $\beta_2 \approx \Delta Y/(\Delta X/X)$

Thus: $\Delta Y \approx \beta_2 (\Delta X/X)$ and $0{,}01 \cdot \beta_2$ shows change in Y value, if X changes by 1%

Example

	GDP	M2
1995	6541,8	722
1996	6778,2	887
1997	7343,6	1286
1998	7695,7	1411
1999	7946,1	1705
2000	8495,6	2176
2001	9179,4	2737
2002	9773,6	3083
2003	10476,4	3301
2004	11385,3	4035
2005	12592,4	5610
2006	14132,9	7821
2007	15543,0	8963
2008	14883,4	8420
2009	12211,1	8171
2010	12169,2	9085



Regression slope coefficient is $\hat{\beta}_2 = 10359$

Interpretation: Increase in M2 by 1 percent, on average, leads to about 103.57 milj. of euro increase in GDP.

Reciprocal model

Models of following type are known as reciprocal models:

$$Y_i = \beta_1 + \beta_2 (1/X_i) + u_i$$

Features:

- As X increases indefinitely, the term $\beta_2 (1/X_i)$ approaches zero and Y approaches the limiting or *asymptotic* value β_1
- Therefore: reciprocal models have built in them an asymptote or limit value that the dependent variable will take when the value of X increases indefinitely.

Reciprocal model

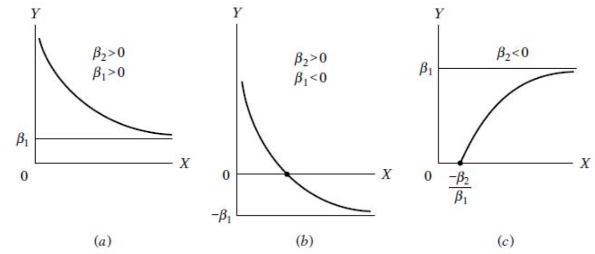


FIGURE 6.6 The reciprocal model: $Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$.

Reciprocal model

The Most popular application of reciprocal model is The Philips curve.

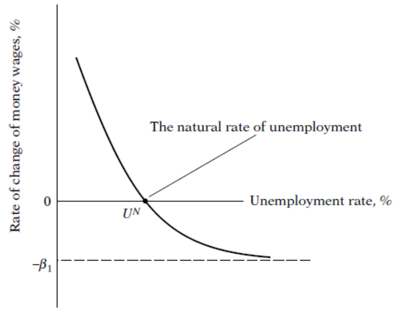
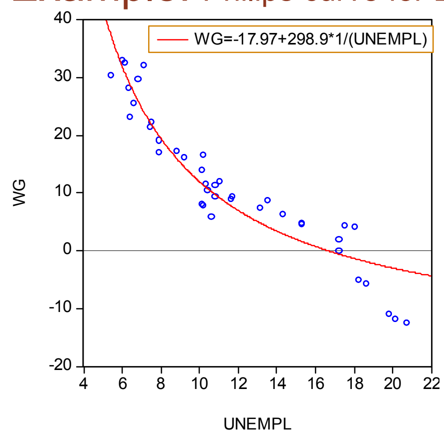


FIGURE 6.8 The Phillips curve.

Example: Philips curve for Latvia



Choice of Functional Form

TABLE 6.6

Model	Equation	Slope $\left(=\frac{dY}{dX}\right)$	Elasticity $\left(=\frac{dY}{dX}\frac{X}{Y}\right)$
Linear	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2\left(\frac{X}{Y}\right)^*$
Log–linear	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{Y}{X}\right)$	β_2
Log–lin	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2\,(Y)$	$\beta_2\,(X)^*$
Lin–log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{1}{Y}\right)^*$
Reciprocal	$Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$	$-\beta_2\left(\frac{1}{X^2}\right)$	$-\beta_2\left(\frac{1}{XY}\right)^*$
Log reciprocal	$\ln Y = \beta_1 - \beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{Y}{X^2}\right)$	$\beta_2\left(\frac{1}{X}\right)^*$

Note: * indicates that the elasticity is variable, depending on the value taken by X or Y or both. When no X and Y values are specified, in practice, very often these elasticities are measured at the mean values of these variables, namely, $\bar{X}$ and $\bar{Y}$.

Choice of Functional Form

The choice of a particular functional form may be comparatively easy in the two-variables case, because we can plot the variables and get rough idea about appropriate model.

Some guidelines:

1. The underlying theory.
2. Change in units vs. %
3. The coefficients of the model chosen should satisfy certain a priori expectations.
4. Sometimes more than one model may fit a given set of data reasonably well. Then we can use R^2 value. *But make sure that in comparing two R^2 values the dependent variable, or regresand, of the two models is the same; the regressor(s) can take any form.*
5. What is of greater importance is the theoretical underpinning of the chosen model, the signs of the estimated coefficients and their statistical significance. If a model is good on these criteria, a model with a lower R^2 may be quite acceptable.

Summary

Model	Interpretation of $\hat{\beta}_2$
$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$	When X↑1 unit, Y mean ↑by $\hat{\beta}_2$ units
$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \ln X_i$	When X↑1 %, Y mean ↑ by 0.01* $\hat{\beta}_2$ units
$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \dfrac{1}{X_i}$	Has no direct interpretation
$\widehat{\ln Y_i} = \hat{\beta}_1 + \hat{\beta}_2 X_i$	When X↑1 unit, Y mean ↑by 100* $\hat{\beta}_2$%
$\widehat{\ln Y_i} = \hat{\beta}_1 + \hat{\beta}_2 \ln X_i$	When X↑1 %, Y mean ↑by $\hat{\beta}_2$ %

Comparing two R^2

$$\hat{Y}_t = 2.6911 - 0.4795 X_t \quad (7.8.8)$$
$$\text{se} = (0.1216) \quad (0.1140) \qquad RSS = 0.1491;\ r^2 = 0.6628$$

$$\widehat{\ln Y_t} = 0.7774 - 0.2530 \ln X_t \quad (7.8.9)$$
$$\text{se} = (0.0152) \quad (0.0494) \qquad RSS = 0.0226;\ r^2 = 0.7448$$

TABLE 7.2 RAW DATA FOR COMPARING TWO R^2 VALUES

Year	Y_t (1)	$\hat{Y}_t$ (2)	$\widehat{\ln Y_t}$ (3)	Antilog of $\widehat{\ln Y_t}$ (4)	$\ln Y_t$ (5)	$\ln(\hat{Y}_t)$ (6)
1970	2.57	2.321887	0.843555	2.324616	0.943906	0.842380
1971	2.50	2.336272	0.853611	2.348111	0.916291	0.848557
1972	2.35	2.345863	0.860544	2.364447	0.854415	0.852653
1973	2.30	2.341068	0.857054	2.356209	0.832909	0.850607
1974	2.25	2.326682	0.846863	2.332318	0.810930	0.844443
1975	2.20	2.331477	0.850214	2.340149	0.788457	0.846502
1976	2.11	2.173233	0.757943	2.133882	0.746688	0.776216
1977	1.94	1.823176	0.627279	1.872508	0.662688	0.600580
1978	1.97	2.024579	0.694089	2.001884	0.678034	0.705362
1979	2.06	2.115689	0.731282	2.077742	0.722706	0.749381
1980	2.02	2.130075	0.737688	2.091096	0.703098	0.756157

Notes: Column (1): Actual Y values from Table 7.1
Column (2): Estimated Y values from the linear model (7.8.8)
Column (3): Estimated log Y values from the double-log model (7.8.9)
Column (4): Antilog of values in column (3)
Column (5): Log values of Y in column (1)
Column (6): Log values of $\hat{Y}_t$ in column (2)

Comparable R value we can obtain = CORREL(Y; antilog $\widehat{\ln Y_t}$)^2

Comparing two R^2

R^2 values for two models are comparables only if dependent variable is the same for both models.

LOG- … model:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\widehat{\ln Y} - \overline{\ln Y}\right)^2}{\sum \left(\ln Y - \overline{\ln Y}\right)^2}$$

LIN- … model:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y} - \bar{Y}\right)^2}{\sum \left(Y - \bar{Y}\right)^2}$$

These values are not directly comparable. If we want to compare, we have to make additional calculations.