

Variable Transformations & Functional Forms

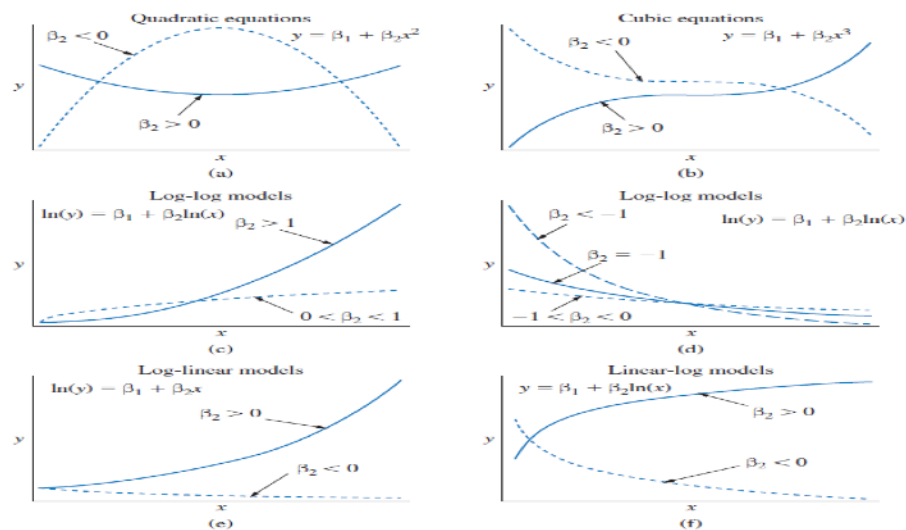


FIGURE 4.5 Alternative functional forms.

Linear Equation

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$\beta_1 = \frac{dY}{dX}$$

the impact of one unit change in X on Y

Double Log (log-log) equation

$$\ln Y_t = \beta_0 + \beta_1 \ln X_t + u_t$$

$$Y_t^* = \ln Y_t \quad X_t^* = \ln X_t$$

$$Y_t^* = \beta_0 + \beta_1 X_t^* + u_t$$

$$\beta_1 = \frac{dY^*}{dX^*}$$

or

$$\beta_1 = \frac{d \ln Y_t}{d \ln X_t} \rightarrow \beta_1 = \frac{\frac{d Y_t}{Y_t}}{\frac{d X_t}{X_t}}$$

β_1 = proportional Δ in Y / Proportional Δ in X

(multiplying both sides by 100)

$$\beta_1 = \frac{\% \Delta \text{ in } Y}{\% \Delta \text{ in } X} \rightarrow \text{the elasticity of } Y \text{ with respect to } X$$

Semi-log equations : log-lin model

$$\ln Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$\beta_1 = \frac{d \ln Y_t}{d X_t} \rightarrow \beta_1 = \frac{\frac{d \ln Y_t}{d Y_t}}{\frac{d \ln X_t}{d X_t}}$$

$$\beta_1 = \frac{1}{Y_t} \cdot \frac{d Y_t}{d X_t}$$

$$\beta_1 = \frac{\frac{d Y_t}{Y_t}}{d X_t} \quad \frac{\text{proportional } \Delta \text{ in } Y}{\text{absolute } \Delta \text{ in } X} \quad \xrightarrow{\text{multiple by } 100}$$

$$100\beta_1 = 100 \left(\frac{d Y_t / Y_t}{d X_t} \right) = \frac{\% \Delta \text{ in } Y}{\text{absolute } \Delta \text{ in } X}$$

$$\beta_1 = \frac{dY_t}{d \ln X_t}$$

$$\beta_1 = \frac{\text{absolute } \Delta \text{ in } Y}{\text{proportional } \Delta \text{ in } X} \rightarrow \frac{\beta_1}{100} = \frac{\text{absolute } \Delta \text{ in } Y}{\text{percentage } \Delta \text{ in } X}$$

example

$$\ln PH_t = 9.23 + 0.30 RN_t$$

PH: price of house

RN: number of rooms

when RN ↑ by 1, PH increase by $(100 \times 0.30) \% = 30\%$

Model Comparisons

$$(1) Y_t = a_0 + a_1 X_t + U_t$$

$$(2) \ln Y_t = b_0 + b_1 X_t + U_t$$

$$(3) \ln Y_t = c_0 + c_1 \ln X_t + U_t$$

$$(4) Y_t = d_0 + d_1 \ln X_t + U_t$$

$$(5) Y_t = e_0 + e_1 X_t^2 + U_t$$

assume that all the models are data-acceptable and theory-consistent

→ Using R^2

the specification of the dependent variable must be same

Models 1-4-5 & Models 2-3 can be compared using R^2 but not 1-2

⇒ Using SSR

SSR may change with scaling of units of measurement

Consider

$$(1) Y_t = a_0 + a_1 X_t + U_{1t}$$

$$(2) \ln Y_t = b_0 + b_1 X_t + U_{2t}$$

$$(3) Y_t^* = a_0 + a_1 X_t + U_{3t}$$

$$(4) \ln Y_t^* = b_0 + b_1 X_t + U_{4t}$$

• where $Y_t^* = m Y_t$ and m is a constant

Log-Lin Models

$$(2) \ln Y_t = b_0 + b_1 X_t + U_{2t}$$

$$(4) \ln Y_t^* = b_0 + b_1 X_t + U_{4t}$$

$$\sum \hat{U}_{2t}^2 = \sum (\ln Y_t - \ln \hat{Y}_t)^2$$

$$\sum \hat{U}_{4t}^2 = \sum (\ln Y_t^* - \ln \hat{Y}_t^*)^2$$

$$= \sum (\ln m Y_t - \ln m \hat{Y}_t)^2$$

$$= \sum (\ln m + \ln Y_t - \ln m - \ln \hat{Y}_t)^2$$

$$= \sum (\ln Y_t - \ln \hat{Y}_t)^2$$

$$= \sum \hat{U}_{2t}^2$$

$$\boxed{\sum \hat{U}_{2t}^2 = \sum \hat{U}_{4t}^2}$$

SSR does NOT change with scaling. The same case applies to the log-log models

Lin - Lin Models

$$(1) Y_t = a_0 + a_1 X_t + U_{1t}$$

$$(3) Y_t^* = a_0 + a_1 X_t + U_{3t}$$

$$\sum \hat{U}_{1t}^2 = \sum (Y_t - \hat{Y}_t)^2$$

$$\begin{aligned}\sum \hat{U}_{3t}^2 &= \sum m Y_t - m \hat{Y}_t)^2 \\ &= \sum m^2 (Y_t - \hat{Y}_t)^2 \\ &= m^2 \sum \hat{U}_{1t}^2\end{aligned}$$

* if the scale of measurement changes by m , then $\sum \hat{U}_t^2$ changes by m^2

SSR changes with scaling. The same case applies to the lin-log models

Lin - Lin & Lin-log Models: A solution

Consider

$$(1) Y_t = b_0 + b_1 X_t + U_t$$

$$(2) Y_t^* = c_0 + c_1 X_t + U_t^*$$

$$\text{where } Y_t^* = m Y_t$$

$$\sum \hat{U}_t^{*2} = m^2 \sum U_t^2$$

$$\text{Estimate (1)} \Rightarrow SSR_1 = S_L$$

$$\text{Estimate (2)} \Rightarrow SSR_2; S_L^A = \frac{S_L}{\bar{Y}_a^2}$$

$$\text{compute } Q = \frac{S_L^A}{S_L}$$

Scale Y_t by $\bar{Y}_G$ (geometric mean of Y)

$$\text{where } \bar{Y}_G = e^{\ln Y}$$

Estimate:

$$Y_t^S = a_0 + a_1 X_t + u_{1t}$$

$$\ln Y_t = c_0 + c_1 X_t + v_{2t}$$

$$\text{where } Y_t^S = \frac{Y_t}{\bar{Y}_G}$$

compare the SSR's

then whether the SSR's are equal by using **MVD test**

Mackinnon - Davidson - White (MDW) Test

$$\text{compute } Q_{MDW} = \frac{S_L^A}{S_L}$$

$$H_0: S_L = S_L^A$$

$$Q_{MDW} \sim \frac{I}{2} | \ln Q_{MDW} | \sim \chi^2_{(1)}$$

reject H_0 if $Q_{MDW} > \chi^2_{(1)}$

choose the model with smaller SSR

4- A researcher wishes to estimate the parameters of the following model using OLS

$$\text{Model 1} \Rightarrow Y_t = bX_t + u_t$$

however, the assistant confuses the dependent and independent variables and instead estimates the following model using OLS

$$\text{Model 2} \Rightarrow X_t = aY_t + u_t \text{ and finds}$$

$$\hat{X}_t = 12.17 \hat{Y}_t \quad R^2 = 0.993 \quad SSR = 92.7 \quad t_a = 35.38 \quad T = 10$$

a) using findings from model 2, estimate b and its standard error

$$R_2^2 = \frac{\sum \hat{X}_t^2}{\sum X_t^2} \quad \hat{X}_t = \hat{a} Y_t \text{ and } \hat{X}_t = X_t - \hat{u}_t$$

$$\begin{aligned} \sum \hat{X}_t^2 &= \sum \hat{X}_t \hat{X}_t = \sum \hat{a} Y_t (X_t - \hat{u}_t) \\ &= \sum \hat{a} X_t Y_t - \sum \hat{a} Y_t \hat{u}_t \\ &= \hat{a} \sum X_t Y_t - \underbrace{\hat{a} \sum Y_t \hat{u}_t}_{=0 \text{ from FOC's}} \end{aligned}$$

$$\sum \hat{X}_t^2 = \hat{a} \sum X_t Y_t \quad \hat{a} \rightarrow \hat{b}$$

$$R_2^2 = \frac{\sum \hat{X}_t^2}{\sum X_t^2} = \frac{\hat{a} \sum X_t Y_t}{\sum X_t^2}$$

$$R_2^2 = \hat{a} \hat{b} \Rightarrow 0.993 = (12.17) (\hat{b})$$

$$\boxed{\hat{b} = 0.08}$$

for standard error, note that

$t^2 = F$ for single restriction

$$\Rightarrow \left(\frac{\hat{b} - 0}{se(\hat{b})} \right)^2 = \frac{R_1^2}{1 - R_1^2} \cdot \frac{dof}{P}$$

$$\text{where } R^2_1 = \frac{\sum \hat{y}_t^2}{\sum y_t^2} = \frac{\hat{b} \sum y_t x_t}{\sum y_t^2} = \hat{a} \hat{b} = R^2_2$$

$$\left(\frac{0.082}{se(\hat{b})} \right)^2 = \frac{0.993}{1-0.993} \cdot \frac{\frac{7-1}{P}}{7} \quad se(\hat{b}) = 0.0228$$

b) test if b is significantly different from zero

$$t_{\hat{b}} = \frac{\hat{b} - 0}{se(\hat{b})} = \frac{0.082 - 0}{0.0228} = 3.59 > t_{0.025, 9} = 2.262$$

$\Rightarrow b$ is significantly different from zero