ACE 564
Spring 2006

Lecture 1

Extensions of the Simple Linear Regression Model I: Choosing the Functional Form

by
Professor Scott H. Irwin

Readings:

Griffiths, Hill and Judge. "Questions Relative to the Algebraic Form of the Household Expenditure-Income Relationship," Section 8.3 and "Other Economic-Statistical Models," Section 8.5 in *Learning and Practicing Econometrics*

Mirer. "Functions and Graphs," "Logarithms" and "The Coefficients in Logarithmic Models" in *Economic Statistics and Econometrics* (provided by instructor)

Functional Form

Some background is useful before discussing specifics of functional form

A regression model essentially makes two statements

- Economic statement about how the variables in the equation are related

- Statistical statement about the distribution of population disturbance term

Functional form is concerned with the economic statement embedded in a regression model

To this point, we have always assumed a linear economic model

Brings us to the first "what if" question regarding the simple linear regression model:

What if the relationship between y_t and x_t is not linear?

Economic theory often implies a non-linear relationship between variables

Fortunately, the simple linear regression model is much more flexible than it may appear at first

The reason is that linear in the simple linear regression model, as we learned last semester, refers to linear in parameters not linear in variables

$\Rightarrow$ parameters may not be multiplied together, divided, squared, cubed, etc.

$\Rightarrow$ variables may be transformed as desired so long as transformation preserves the model assumptions

$\Rightarrow$ In other words, a major objective of choosing a functional form, or transforming the variables, is to create a model in which the error term has the following properties

$$E(e_t) = 0$$

$$\text{var}(e_t) = \sigma^2$$

$$\text{cov}(e_t, e_s) = 0$$

$$e_t \sim N(0, \sigma^2)$$

If these assumptions hold then the least squares estimators have good statistical properties and we can use the procedures for statistical inference that we developed earlier

We will make use of two variable transformations to represent an amazing number of "shapes"

- Natural logarithm
- Reciprocal

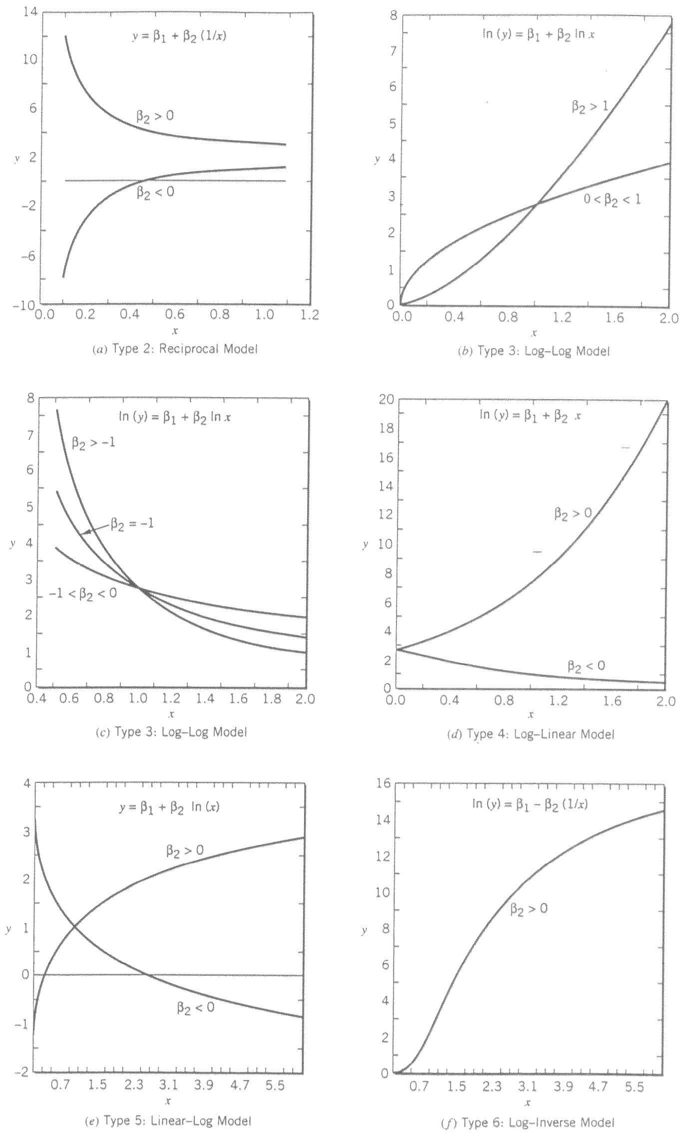

FIGURE 6.3 Alternative functional forms

Hill, C., W. Griffiths, and G. Judge. *Undergraduate Econometrics.* John Wiley & Sons, Inc., New York, NY 1997.

Economic Theory and Functional Form: An Example

Researcher is interested in estimating a production function for corn

Consider the simplified production function

$$Q = f(F)$$

where Q is corn output and F is nitrogen fertilizer

Estimation requires a specific economic model, such as the linear production function,

$$Q = \beta_1 + \beta_2 F$$

where

$$\frac{dQ}{dF} = \beta_2 = \text{marginal physical product of fertilizer}$$

This makes an important economic statement

- Marginal physical product (MPP) is constant with respect to the level of fertilizer use
- Does this make sense from a production economics standpoint?

Consider an alternative functional form known as the Cobb-Douglass function

$$Q = \beta_1 F^{\beta_2}$$

where

$$\frac{dQ}{dF} = \beta_1 \beta_2 F^{\beta_2 - 1} = \frac{\beta_1 \beta_2}{F^{1-\beta_2}}$$

Assuming β_1 positive and β_2 is between zero and one, the economic statement about MPP made by this functional form is

Illustrates the central role of economic theory in determining functional form

Linear Functional Form

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

Marginal effect

$$\frac{dy_t}{dx_t} = \beta_2$$

Elasticity

$$E_{yx} = \frac{dy_t}{dx_t}\frac{x_t}{y_t} = \beta_2 \frac{x_t}{y_t}$$

Linear specification makes two important economic statements

- Marginal effect of x on y is constant
- Elasticity of y with respect to x changes along the function

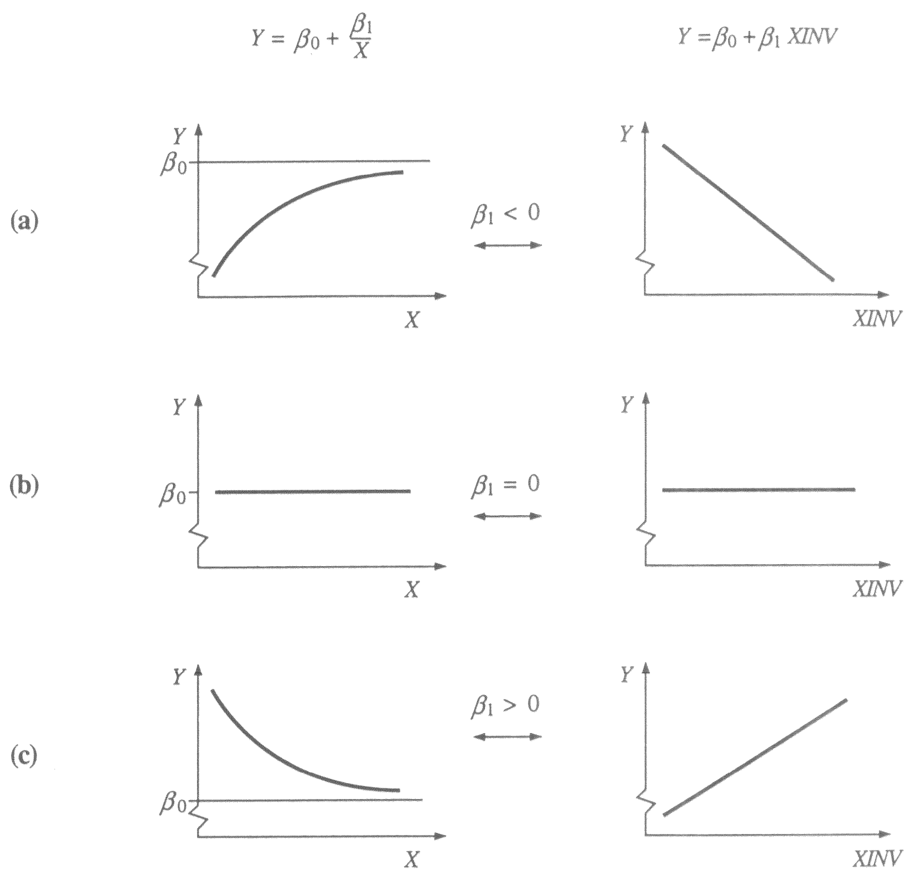

FIGURE 6.2 The geometry of the reciprocal relation depends on the sign of β_1. Except when $\beta_1 = 0$, Y is a nonlinear function of X; as X increases, Y increases ($\beta_1 < 0$) or decreases ($\beta_1 > 0$) and approaches the asymptotic limit β_0. Although Y is a nonlinear function of X, it is a linear function of the inverse of X—which is denoted by *XINV*.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

Reciprocal Functional Form

$$y_t = \beta_1 + \beta_2 \frac{1}{x_t} + e_t$$

or,

$$y_t = \beta_1 + \beta_2 x_t^{-1} + e_t$$

or,

$$y_t = \beta_1 + \beta_2 w_t + e_t$$

where $w_t = x_t^{-1}$

Direct effect

$$\frac{dy_t}{d\left(\frac{1}{x_t}\right)} = \frac{dy_t}{dw_t} = \beta_2$$

Marginal effect

$$\frac{dy_t}{dx_t} = \frac{dy_t}{d\left(\frac{1}{x_t}\right)} \frac{d\left(\frac{1}{x_t}\right)}{dx_t} = -\beta_2 \frac{1}{x_t^2}$$

Elasticity

$$E_{yx} = \frac{dy_t}{dx_t}\frac{x_t}{y_t} = -\beta_2 \frac{1}{x_t^2}\frac{x_t}{y_t} = -\beta_2 \frac{1}{x_t y_t}$$

Reciprocal specification makes three important economic statements

- Non-linear relationship between y and x
- Minimum or maximum levels for y
- Marginal effects are non-constant

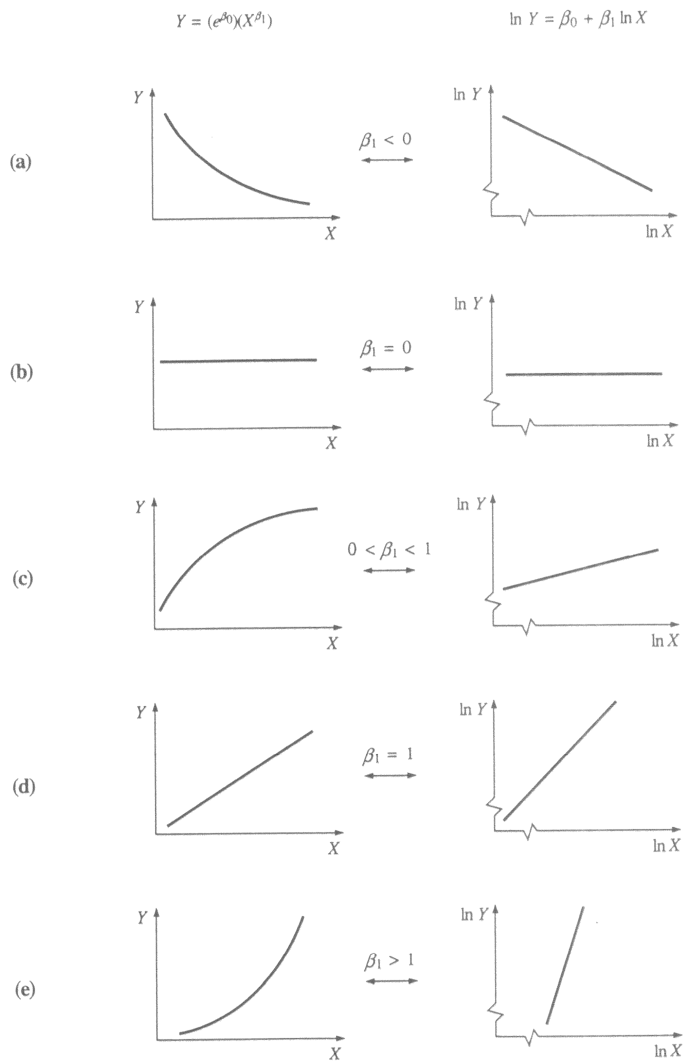

FIGURE 6.3 The geometry of the log-linear relation depends on the sign of β_1. When Y decreases as X increases [case (a), with $\beta_1 < 0$], it is concave upward. When Y increases with X (with $\beta_1 > 0$), the concavity may be upward or downward, depending on the magnitude of β_1. Although Y is a nonlinear function of X, $\ln Y$ is a linear function of $\ln X$; the slope of that line is the same β_1 as in the original formulation. The parameter β_1 is the elasticity of Y with respect to X.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

Log-Log Functional Form

Start with the nonlinear exponential model,

$$y_t = \exp(\beta_1) x_t^{\beta_2} \exp(e_t)$$

where $\exp(e_t)$ is the exponential of e_t

By taking the natural logarithms of both sides, this non-linear regression model can be transformed into an intrinsically linear regression model

$$\ln y_t = \beta_1 + \beta_2 \ln x_t + e_t$$

or,

$$z_t = \beta_1 + \beta_2 w_t + e_t$$

where $z_t = \ln y_t$, $w_t = \ln x_t$ and *ln* is the natural logarithm

This is called a log-log or double log specification

Direct effect

$$\frac{d\ln y_t}{d\ln x_t} = \frac{dz_t}{dw_t} = \beta_2$$

Marginal effect

$$\frac{dy_t}{dx_t} = ?$$

In order to find the answer we need a basic result from the differentiation of logarithms,

$$\frac{d\ln y_t}{dy_t} = \frac{1}{y_t} \qquad \frac{d\ln x_t}{dx_t} = \frac{1}{x_t}$$

or,

$$dy_t = \left(d\ln y_t\right) y_t \quad dx_t = \left(d\ln x_t\right) x_t$$

Substituting,

$$\frac{dy_t}{dx_t} = \frac{d\ln y_t}{d\ln x_t}\frac{y_t}{x_t} = \beta_2 \frac{y_t}{x_t}$$

Elasticity

$$E_{yx} = \frac{dy_t}{dx_t}\frac{x_t}{y_t} = \frac{d\ln y_t}{d\ln x_t}\frac{y_t}{x_t}\frac{x_t}{y_t} = \beta_2 \frac{y_t}{x_t}\frac{x_t}{y_t} = \beta_2$$

***Key Result*: Slope in this regression is the elasticity!**

> β_2 ***indicates the percent change in y for a one – percent increase in x***

Example

$$\beta_2 = -1.21$$

Indicates that y decreases 1.21 percent for a one percent increase in x

- No need to multiply β_2 by 100 to interpret

- We can simply think of β_2 as the ratio of the percentage change in y relative to the percentage change in x

Log-log specification makes two important economic statements

- Non-linear relationship between y and x
- Constant elasticity

Easy to understand, estimate, and interpret

As a result, quite popular functional form in applied economic research

Cautions with double-log functional form

- x and y cannot take on negative values
- Need to be careful in projecting changes in y based on estimated slopes
- Direct effect is not a global measure of change, but is a local measure of change

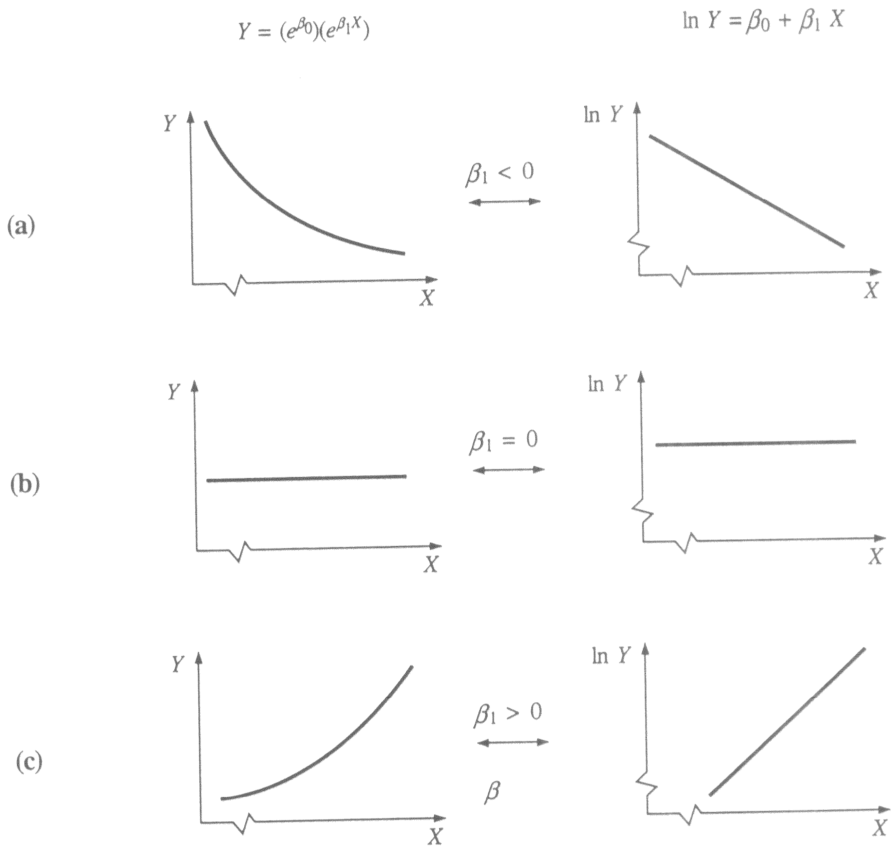

FIGURE 6.4 The geometry of the semilog relation depends on the sign of β_1. Except when $\beta_1 = 0$, Y is a nonlinear function of X. As X increases, Y decreases ($\beta_1 < 0$) or increases ($\beta_1 > 0$), and the relation is concave upward in both cases. Although Y is a nonlinear function of X, ln Y is a linear function of X; the slope of that line is the same β_1 as in the original formulation. The parameter β_1 can be interpreted as the proportional change in Y that results from a unit change in X.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

Log-Linear Functional Form

Start with the following "mixed" non-linear model,

$$y_t = \exp(\beta_1 + \beta_2 x_t + e_t)$$

By taking logarithms of both sides, this non-linear regression model can be transformed into an intrinsically linear regression model,

$$\ln y_t = \beta_1 + \beta_2 x_t + e_t$$

or,

$$z_t = \beta_1 + \beta_2 x_t + e_t$$

where $z_t = \ln y_t$

Direct Effect

$$\frac{d \ln y_t}{dx_t} = \frac{dz_t}{dx_t} = \beta_2$$

To interpret this slope coefficient we need to once again use our earlier result about the differential of a logarithm,

$$\frac{d\ln y_t}{dy_t} = \frac{1}{y_t} \quad \text{or} \quad d\ln y_t = \frac{dy_t}{y_t}$$

Substituting,

$$\beta_2 = \frac{d\ln y_t}{dx_t} = \frac{dy_t / y_t}{dx_t} \approx \frac{\Delta y_t / y_t}{\Delta x_t}$$

Slope in this regression indicates the proportional change in y for a one-unit increase in x

or,

Slope multiplied by 100 in this regression indicates the percent change in y for a one-unit increase in x

Example

$$\beta_2 = 0.09 = \frac{0.09}{1}$$

Indicates that *y* increases proportionally by 0.09 for a one unit increase in *x*

$$\beta_2 \cdot 100 = 9 = 0.09 \cdot 100 = \frac{0.09 \cdot 100}{1}$$

Indicates that *y* increases 9 percent for a one unit increase in *x*

Marginal Effect

$$\frac{dy_t}{dx_t} = \frac{d(\exp(\beta_1 + \beta_2 x_t + e_t))}{dx_t} = \beta_2 \exp(\beta_1 + \beta_2 x_t + e_t)$$
$$= \beta_2 y_t$$

Elasticity

$$E_{yx} = \frac{dy_t}{dx_t}\frac{x_t}{y_t} = \beta_2 y_t \frac{x_t}{y_t} = \beta_2 x_t$$

Log-lin specification makes two important economic statements

- Non-linear relationship between y and x
- Given unit increase in x will result in same percentage change in y

The log-lin regression model is widely used to measure the growth rate in economic time series variables

- Dependent variable is log of economic variable
- Independent variable is time index $(1,\ldots,T)$
- Slope coefficient measures percent change in dependent variable for a one-unit change in time (e.g. month, quarter, year)
- Gujarati has an excellent discussion of this application (pp. 169-171)

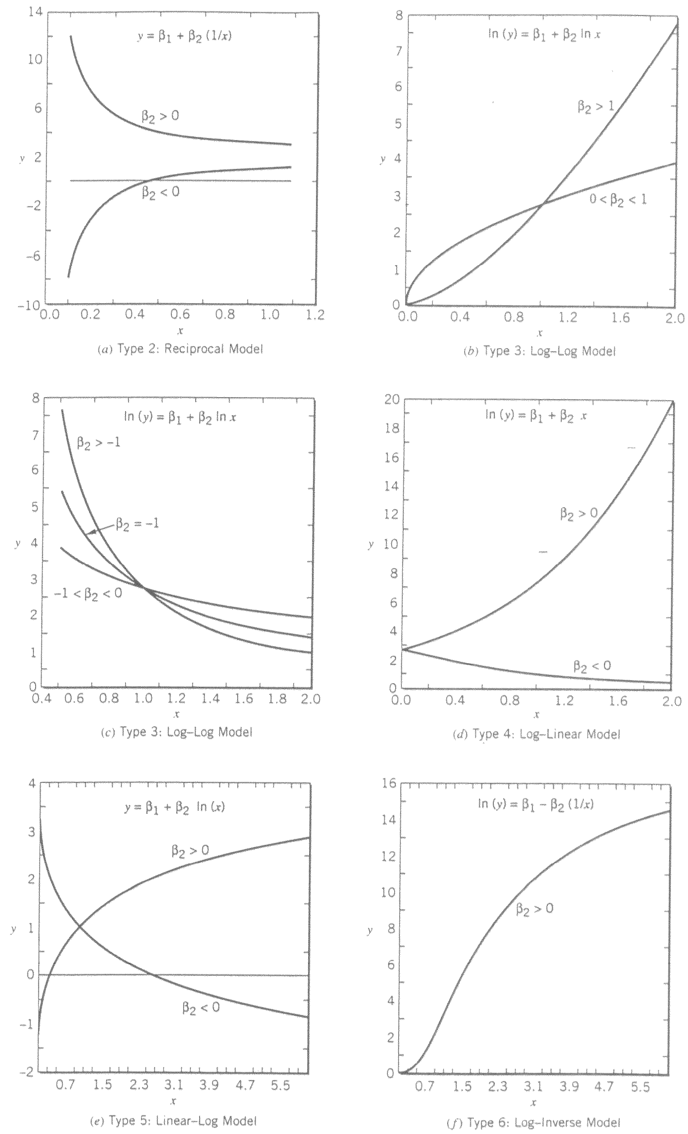

FIGURE 6.3 Alternative functional forms

Hill, C., W. Griffiths, and G. Judge. *Undergraduate Econometrics.* John Wiley & Sons, Inc., New York, NY 1997.

Lin-Log Functional Form

Start with the following "mixed" non-linear model,

$$\exp(y_t) = \exp(\beta_1) x_t^{\beta_2} \exp(e_t)$$

By taking logarithms of both sides, this non-linear regression model can be transformed into an intrinsically linear regression model,

$$y_t = \beta_1 + \beta_2 \ln x_t + e_t$$

or,

$$y_t = \beta_1 + \beta_2 w_t + e_t$$

where $w_t = \ln x_t$

Direct Effect

$$\frac{dy_t}{d\ln x_t} = \frac{dy_t}{dw_t} = \beta_2$$

To interpret this slope coefficient we need to use our earlier result about the differential of a logarithm,

$$\frac{d\ln x_t}{dx_t} = \frac{1}{x_t} \quad \text{or} \quad d\ln x_t = \frac{dx_t}{x_t}$$

Substituting,

$$\beta_2 = \frac{dy_t}{d\ln x_t} = \frac{dy_t}{dx_t / x_t} \approx \frac{\Delta y_t}{\Delta x_t / x_t}$$

Slope in this regression indicates the unit change in y for a one-unit proportional increase in x

or,

Slope divided by 100 in this regression indicates the unit change in y for a one-percent increase in x

Example

$$\beta_2 = 1200 = \frac{12}{0.01}$$

Indicates that y increases 1200 units for a one proportional "unit" increase in x

$$\frac{\beta_2}{100} = \frac{1200}{100} = 12 = \frac{12}{(0.01)100}$$

Indicates that y increases 12 units for a one percent increase in x

Marginal Effect

$$\frac{dy_t}{dx_t} = \frac{dy_t}{d\ln x_t}\frac{d\ln x_t}{dx_t} = \beta_2 \frac{1}{x_t}$$

Elasticity

$$E_{yx} = \frac{dy_t}{dx_t}\frac{x_t}{y_t} = \frac{dy_t}{d\ln x_t}\frac{d\ln x_t}{dx_t}\frac{x_t}{y_t} = \beta_2 \frac{1}{x_t}\frac{x_t}{y_t} = \beta_2 \frac{1}{y_t}$$

Lin-log specification makes two important economic statements

- Non-linear relationship between *y* and *x*
- Given percent increase in *x* will result in same absolute change in *y*

Also can give similar results as reciprocal function but does not have asymptote

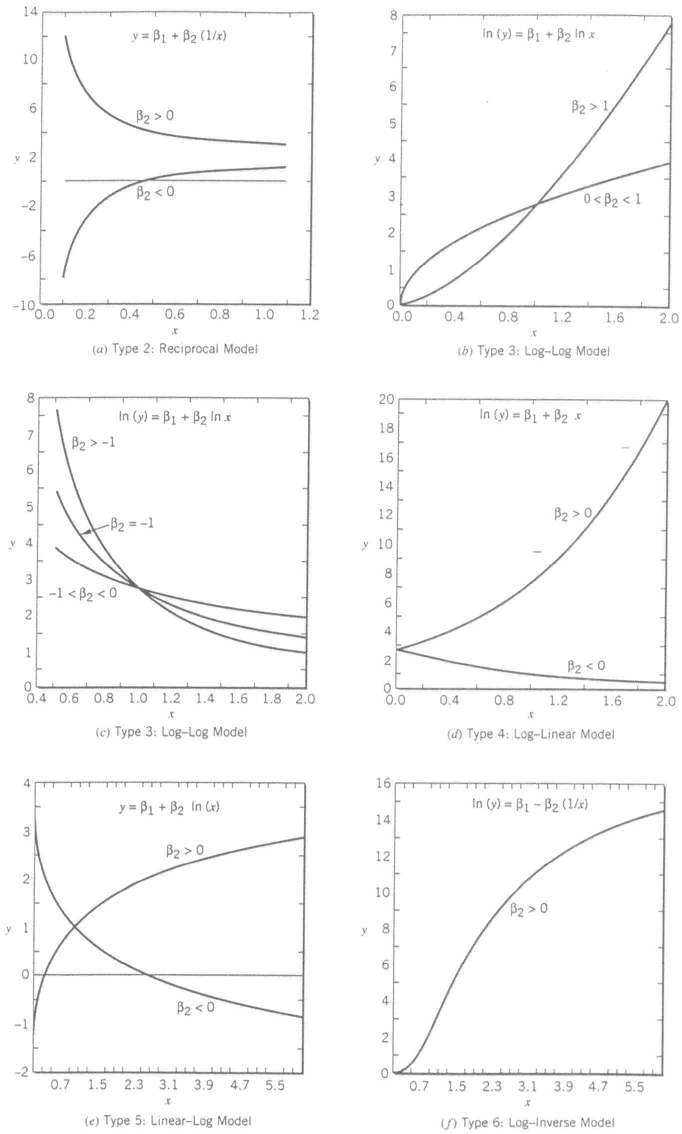

FIGURE 6.3 Alternative functional forms

Hill, C., W. Griffiths, and G. Judge. *Undergraduate Econometrics.* John Wiley & Sons, Inc., New York, NY 1997.

Log-Inverse Functional Form

Start with the following "mixed" non-linear model,

$$y_t = \exp(\beta_1 - \beta_2 \frac{1}{x_t} + e_t)$$

By taking logarithms of both sides, this non-linear regression model can be transformed into an intrinsically linear regression model,

$$\ln y_t = \beta_1 - \beta_2 \frac{1}{x_t} + e_t$$

or,

$$z_t = \beta_1 - \beta_2 w_t + e_t$$

where $z_t = \ln y_t$, $w_t = x_t^{-1}$

Direct Effect

$$\frac{d\ln y_t}{d\left(1/x_t\right)} = \frac{dz_t}{dw_t} = -\beta_2$$

Marginal Effect

$$\frac{dy_t}{dx_t} = \frac{d(\exp(\beta_1 - \beta_2 \frac{1}{x_t} + e_t))}{dx_t} = \beta_2 \frac{1}{x_t^2} \exp(\beta_1 - \beta_2 \frac{1}{x_t} + e_t)$$

$$= \beta_2 \frac{y_t}{x_t^2}$$

Elasticity

$$E_{yx} = \frac{dy_t}{dx_t}\frac{x_t}{y_t} = \beta_2 \frac{y_t}{x_t^2}\frac{x_t}{y_t} = \beta_2 \frac{1}{x_t}$$

Log-inverse specification makes several important economic statements

- Often used to model sales response to advertising
- Non-linear relationship between y and x
- S-shaped implies increasing slope at low levels of x and decreasing slope with high levels of x
- Inflection point at $\beta_2 / 2$
- Asymptotic to $\exp(\beta_2)$

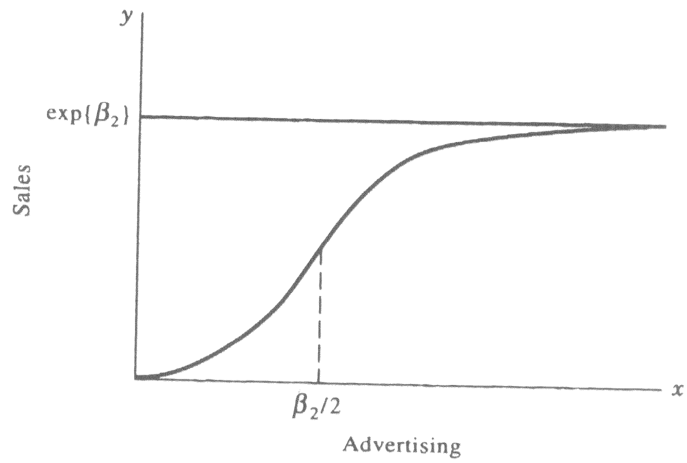

Figure 8.3 Log-inverse function sometimes used to model sales response to advertising.

Griffiths, W.E., R.C. Hill and G.C. Judge. *Learning and Practicing Econometrics*. John Wiley & Sons, Inc., New York, NY, 1993.

Table 8.3 **Some Conventional Functional Forms**

Type	Nonlinear Form	Statistical Model (linear and additive form)	Impact at Margin $\frac{dy_t}{dx_t}$	Elasticity $\frac{dy_t}{dx_t}\frac{x_t}{y_t}$
Linear		$y_t = \beta_1 + x_t\beta_2 + e_t$	β_2	$\beta_2 \frac{x_t}{y_t}$
Reciprocal		$y_t = \beta_1 + \frac{1}{x_t}\beta_2 + e_t$	$-\beta_2 \frac{1}{x_t^2}$	$-\beta_2 \frac{1}{x_t y_t}$
Log-log	$y_t = \alpha_1 x_t^{\beta_2} \exp\{e_t\}$	$\ln y_t = \beta_1 + \beta_2 \ln x_t + e_t$	$\beta_2 \frac{y_t}{x_t}$	β_2
Log-linear (exponential)	$y_t = \exp\{\beta_1 + x_t\beta_2 + e_t\}$	$\ln y_t = \beta_1 + x_t\beta_2 + e_t$	$\beta_2 y_t$	$\beta_2 x_t$
Linear-log (semilog)	$\exp\{y_t\} = \exp\{\beta_1 + e_t\}x_t^{\beta_2}$	$y_t = \beta_1 + \beta_2 \ln x_t + e_t$	$\beta_2 \frac{1}{x_t}$	$\beta_2 \frac{1}{y_t}$
Log-inverse	$y_t = \exp\left\{\beta_1 - \frac{1}{x_t}\beta_2 + e_t\right\}$	$\ln y_t = \beta_1 - \frac{1}{x_t}\beta_2 + e_t$	$\beta_2 \frac{y_t}{x_t^2}$	$\beta_2 \frac{1}{x_t}$

Griffiths, W.E., R.C. Hill and G.C. Judge. *Learning and Practicing Econometrics*. John Wiley & Sons, Inc., New York, NY, 1993.

Choice of Functional Forms

Contrary to the impression gained in many textbooks, there are no hard and fast rules to determine the appropriate functional form for a regression model

- This can lead to no end of frustration for applied researchers (and no shortage of arguments!)

- You must be prepared to defend your choice based on some "reasonable" criteria and evidence

Griffiths, Hill and Judge (1993) suggest two crucial points regarding this issue

- We must in some way make use of non-sample information in the form of economic principles and other logic, as well as the information in our sample of data

- Blind mechanical application of one particular criterion, such as R^2, is not a satisfactory strategy. All criterion are subject to the problem of "data mining"

Practical Suggestions

- Consider first the implications of relevant economic theory

  ⇒Carefully think through the marginal effect implied by theory, if any, compared to the marginal effect for a particular functional form

- Consider any patterns in the data

  ⇒Plotting the data may be especially useful

- Test alternatives and see if results of interest are sensitive to choice of functional form

  ⇒Choose a functional form that is sufficiently flexible to "fit" the data while at the same time preserving the underlying statistical assumptions

An Example: The Food Expenditure Problem

From the array of shapes in the figure on page (1-23), two possible choices appear to be consistent with economic theory: the reciprocal and lin-log models

The reciprocal model is

$$y_t = \beta_1 + \beta_2 \frac{1}{x_t} + e_t$$

and for the food expenditure model we might assume that $\beta_1 > 0$ and $\beta_2 < 0$

With these assumptions, then as income increases, household consumption of food increases at a decreasing rate and reaches an upper bound equal to β_1

> $\Rightarrow$ An absolute upper bound on food expenditure may or may not be realistic

Another property under these assumptions is that when $x < -\beta_2 / \beta_1$ the model predicts expenditure on food to be negative

$\Rightarrow$ This is obviously unrealistic and implies this functional form is inappropriate for small values of x

As noted earlier, when choosing a functional form one practical guideline is to consider how the dependent variable changes with the independent variable

In the reciprocal model, the slope of the relationship between y and x (marginal effect) is

$$\frac{dy_t}{dx_t} = -\beta_2 \frac{1}{x_t^2}$$

If the parameter $\beta_2 < 0$ then there is a positive relationship between food expenditure and income

In addition, as income increases the slope declines

$\Rightarrow$ The "marginal propensity to spend on food" diminishes, as economic theory predicts

An alternative to the reciprocal model is the lin-log model

$$y_t = \beta_1 + \beta_2 \ln x_t + e_t$$

and for the food expenditure model we might assume that $\beta_1 > 0$ and $\beta_2 > 0$

$\Rightarrow$ Since x must be greater than zero, it is not possible for food expenditure to be negative

In the lin-log model, the slope of the relationship between y and x (marginal effect) is

$$\frac{dy_t}{dx_t} = \beta_2 \frac{1}{x_t}$$

If the parameter $\beta_2 > 0$ then there is a positive relationship between food expenditure and income

In addition, as income increases the slope declines

$\Rightarrow$ The "marginal propensity to spend on food" diminishes, as economic theory predicts

The elasticity of the relationship between y and x is

$$E_{yx} = \frac{dy_t}{dx_t}\frac{x_t}{y_t} = \beta_2 \frac{1}{y_t}$$

If the parameter $\beta_2 > 0$ then there is a positive relationship between percentage increases in income and percentage increases in food expenditure

In addition, the greater the amount of food expenditure y the smaller the elasticity, β_2 / y_t

Summary: A review of the mathematical properties of the two models suggests the lin-log functional form is the most appropriate for the food expenditure problem

Final Thoughts

- Functional form is one of the most important choices faced by a researcher

- We will have more to say about this issue in the lecture on specification errors

- Pay particular attention to analysis of the estimated residuals

  ⇒Examining plots of estimated residuals is an especially useful device to uncover inadequacies in any functional form

  ⇒Patterns in estimated residuals will play an important role in other forms of model evaluation we will cover later this semester