Bahan Kuliah Ekonomterika I

Spesifikasi Model: Pemilihan Bentuk Fungsional

Imam Awaluddin

Jurusan Ekonomi Pembangunan

Fakultas Ekonomi dan Bisnis

Universitas Lampung

2020

Spesifikasi Model

- **Salah satu asumsi klasik adalah model regresi yang digunakan dalam analisis telah dispesifikasikan dengan "benar".**
- Jika tidak dispesifikasikan dengan "benar" maka terjadi bias spesifikasi model.
- Namun pencarian spesifikasi model yang "benar" seperti mencari Cawan Suci (the Holy Grail).

Bias-bias Kesalahan Spesifikasi

- Menghilangkan variabel yang relevan (omitted variable(s)) → **underfitting a model.**
- Menyertakan variabel yang tidak relevan (irrelevant variabel(s)) → **overfitting a model**.
- **Bentuk fungsi yang salah**
- Kesalahan pengukuran

Model Regresi Tanpa Intersep

- Model regresi dari titik origin

$$Y_i = \beta_2 X_i + u_i$$

- Contoh: model CAPM

$$ER_i - r_f = \beta_i (ER_m - r_f)$$

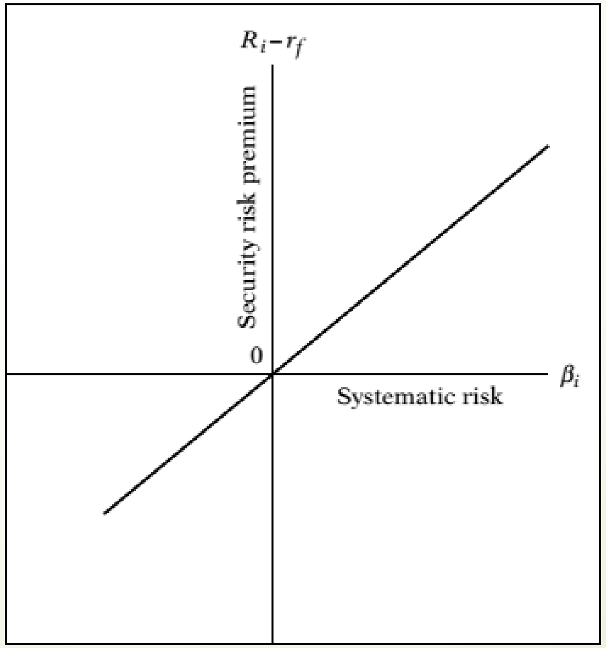
- dimana ER_i = ekspektasi tingkat penerimaan saham ke-i, ER_m = ekspektasi penerimaan dalam pasar porotofolio (misal tk pengembalian IHSG), r_f = tingkat pengembalian asset bebas resiko (misal tingkat bunga SUN), β_i = koefisien beta yang merupakan pengukuran resiko sistematik, yaitu resiko yang tidak dapat dihilangkan melalui diversifikasi.

Model CAPM

- Model empiris:

$$R_i - r_f = \alpha + \beta\,(R_i - r_f) + u_i \quad \rightarrow \quad \text{market model}$$

$$R_i - r_f = \beta\,(R_i - r_f) + u_i \quad \rightarrow \quad \text{market model jika CAPM tercapai}$$



Bentuk Fungsional dari Model Regresi

- Dalam asumsi klasik → Linier dalam parameter, dapat linier atau nonlinier dalam variabel.
- Bentuk fungsional regresi
  - Linier: $Y_t = \beta_0 + \beta_1 X_t + u_t$
  - Dobel log: $\ln Y_t = \beta_0 + \ln \beta_1 X_t + u_t$
  - Log-lin: $\ln Y_t = \beta_0 + \beta_1 X_t + u_t$
  - Lin-log: $Y_t = \beta_0 + \beta_1 \ln X_t + u_t$
  - Reciprocal: $Y_t = \beta_0 + \beta_1 \left(\frac{1}{X_t} \right) + u_t$
  - Polinomial: $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + .. + \beta_k X_t^k + u_t$

Contoh Penerapan: Fungsi Produksi Cobb-Douglas

- Misalkan fungsi produksi Cobb-Douglas

$$Y_i = \beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2} e^{u_i}$$

Dimana Y = output, X_1 = input tenaga kerja, X_2 = input modal, u = faktor gangguan stokastik, dan e = dasar dari logaritma natural

→ Model tersebut adalah nonlinier.

Transformasi model

- Jika model tsb ditransformasikan dalam logaritma natural →bentuk dobel log natural

$$\ln Y_i = \ln \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + u_i$$

$$\ln Y_i = \beta_0^* + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + u_i$$

$$\text{dimana} \quad \beta_0^* = \ln \beta_1$$

- β_1 adalah elastisitas output (parsial) thd tenaga kerja, dan β_2 adalah elastisitas output (parsial) thd modal

Contoh

- Dengan menggunakan data 51 negara bagian dalam sektor industri di AS pada tahun 2005 yang ada dalam Tabel 7.3 buku Gujarati & Porter.
- Y = Output yang diukur dengan tambahan nilai (dalam ribuan $)
- L = input tenaga kerja yang diukur dengan waktu kerja (dalam ribuan jam)
- K = input kapital diukur dengan pengeluaran kapital (dalam ribuan $)

Estimasi dengan bentuk linier

$$Y_i = \beta_0 + \beta_1 L_i + \beta_2 K_i + u_i$$

Dependent Variable: Y
Method: Least Squares
Date: 05/17/14 Time: 22:24
Sample: 1 51
Included observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	233621.6	1250364.	0.186843	0.8526
L	47.98736	7.058245	6.798766	0.0000
K	9.951891	0.978116	10.17455	0.0000

R-squared	0.981065	Mean dependent var	43217548
Adjusted R-squared	0.980276	S.D. dependent var	44863661
S.E. of regression	6300694.	Akaike info criterion	34.20724
Sum squared resid	1.91E+15	Schwarz criterion	34.32088
Log likelihood	-869.2846	Hannan-Quinn criter.	34.25066
F-statistic	1243.514	Durbin-Watson stat	1.684519
Prob(F-statistic)	0.000000		

$$\hat{Y}_i = 233621{,}6 + 47{,}98736 L_i + 9{,}951891 K_i$$

Interpretasi model linier

$$\hat{Y}_i = 233621,6 + 47,98736 L_i + 9,951891 K_i$$

- Konstanta sebesar 233621,6 menunjukkan bahwa jika jumlah jam kerja dan pengeluaran modal keduanya nol, output akan ada sebesar 233621,6 ribu dolar.

$$\beta_1 = \frac{\Delta Y}{\Delta L}$$

- Koefisien regresi parsial variabel L sebesar 47,98736 menunjukkan bahwa setiap kenaikan seribu jam kerja, output akan meningkat sebesar 47,98736 ribu dolar dengan modal adalah tetap.

$$\beta_2 = \frac{\Delta Y}{\Delta K}$$

- Koefisien regresi parsial variabel K sebesar 9,95189 menunjukkan bahwa setiap kenaikan seribu dolar pengeluaran modal, output akan meningkat sebesar 9,95189 ribu dolar dengan jam kerja adalah tetap.

Estimasi dengan bentuk double log

$$\ln Y_i = \ln \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + u_i$$

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 05/17/14   Time: 22:25
Sample: 1 51
Included observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.887600	0.396228	9.811514	0.0000
LOG(L)	0.468332	0.098926	4.734170	0.0000
LOG(K)	0.521279	0.096887	5.380274	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	0.252028
Sum squared resid	3.415520	Schwarz criterion	0.365665
Log likelihood	-3.426721	Hannan-Quinn criter.	0.295452
F-statistic	645.9311	Durbin-Watson stat	1.946387
Prob(F-statistic)	0.000000		

$$\ln \hat{Y}_i = 3,8876 + 0,4683 \ln L_i + 0,5213 \ln K_i$$

Interpretasi model double log

$$\ln \hat{Y}_i = 3{,}8876 + 0{,}4683 \ln L_i + 0{,}5213 \ln K_i$$

$$\beta_0^* = \ln \beta_0$$

- Konstanta sebesar 3,8876 tidak boleh langsung diinterpretasikan, namun harus dicari anti logaritma naturalnya (anti ln). Oleh karena itu konstantanya adalah anti ln 3,8876 yaitu sebesar 48,7936 menunjukkan bahwa jika jumlah jam kerja dan pengeluaran modal keduanya nol, output akan ada sebesar 48,7936 ribu dolar.

Interpretasi model double log

$$\ln \hat{Y}_i = 3,8876 + 0,4683 \ln L_i + 0,5213 \ln K_i$$

$$\beta_1 = \frac{\Delta \ln Y}{\Delta \ln K}$$

- Koefisien regresi parsial untuk model double log dapat secara langsung diinterpretasikan sebagai elastisitas.
- Koefisien regresi parsial untuk variabel lnL merupakan elastisitas output parsial terhadap tenaga kerja sebesar 0,4683 menunjukkan bahwa setiap kenaikan satu persen jam kerja, output akan meningkat sebesar 0,4683 persen dengan modal adalah tetap.

Interpretasi model double log

$$\ln \hat{Y}_i = 3{,}8876 + 0{,}4683 \ln L_i + 0{,}5213 \ln K_i$$

$$\beta_2 = \frac{\Delta \ln Y}{\Delta \ln K}$$

- Koefisien regresi parsial untuk variabel lnK merupakan elastisitas output parsial terhadap modal sebesar 0,5213 menunjukkan bahwa setiap kenaikan satu persen pengeluaran modal, output akan meningkat sebesar 0,5213 persen dengan jam kerja adalah tetap.
- Untuk fungsi produksi, penjumlahan ($\beta_1 + \beta_2$) menggambarkan *returns to scale*; yaitu respon output yang disebabkan oleh perubahan proporsional dari input. Dari hasil estimasi di atas ($\beta_1 + \beta_2$) adalah 0,4683 + 0,5213 = 0,9896 ≡ 1, atau *constant returns to scale*.

Model Log-Lin

- Pada umumnya untuk mengukur tingkat pertumbuhan.

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

$$\ln Y_t = \ln Y_0 + t \ln(1+r)$$

- Dengan menganggap

$$\beta_1 = \ln Y_0$$

$$\beta_2 = \ln(1+r)$$

- Sehingga dengan menambahkan faktor gangguan

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

Model Log-Lin

$$\ln Y_t = \ln Y_0 + t \ln(1 + r)$$

$$\ln Y_t = \beta_1 + t\beta_2$$

$$\beta_2 = \frac{d \ln Y_t}{dt} = \frac{\frac{dY_t}{Y_t}}{dt}$$

$$\beta_2 \times 100\% = \frac{\%\ \text{perubahan}\ Y}{\text{setiap 1 unit waktu}}$$

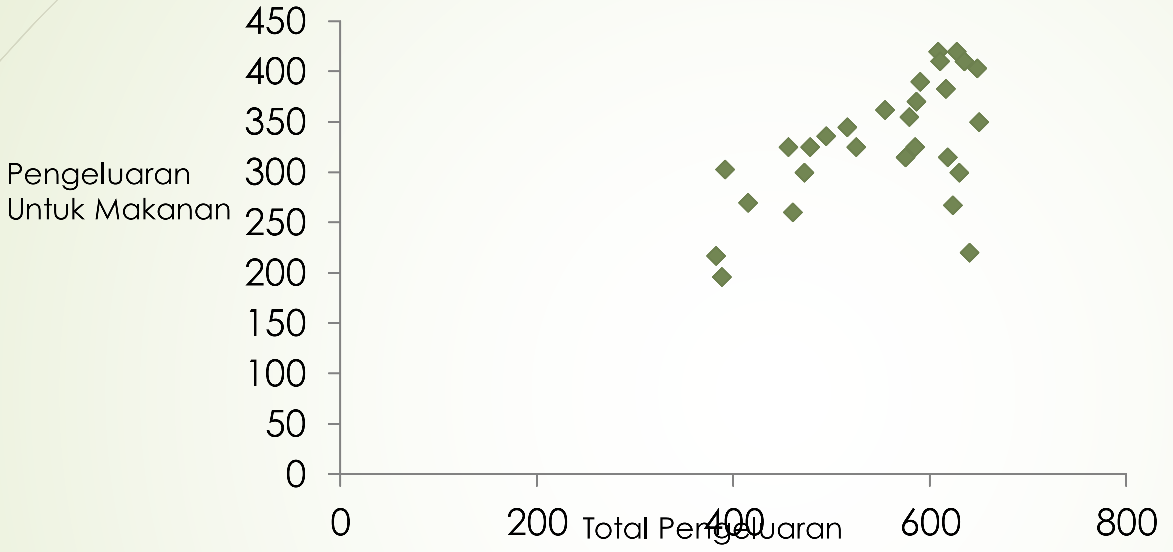
Laju pertumbuhan

- Model pertumbuhan

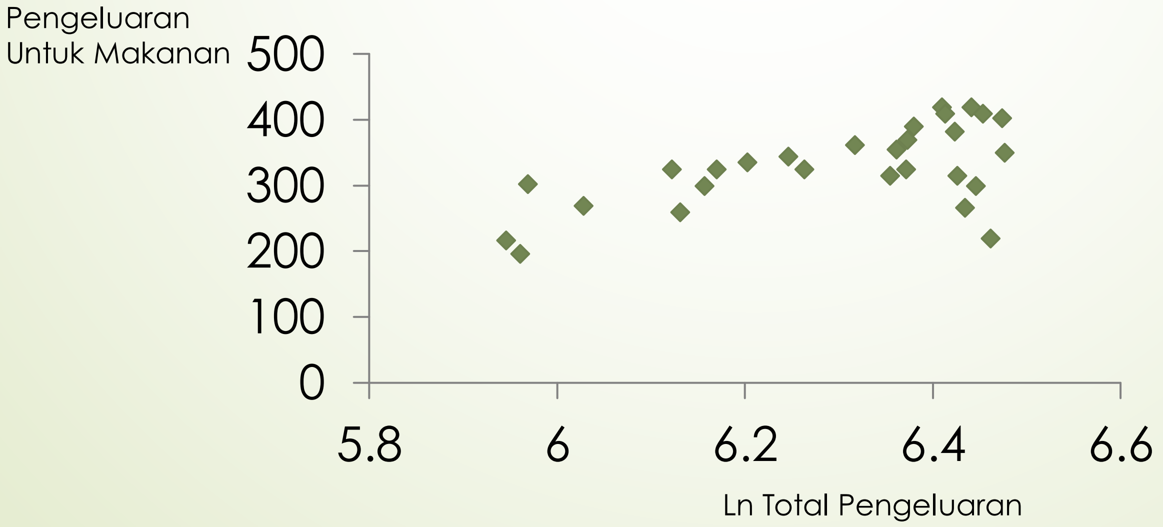
Model Lin-Log→ Engel Expenditure Model

- Hubungan antara pengeluaran untuk makanan dan total pengeluaran.
- Pengeluaran untuk makanan tergantung dari total pengeluaran.
- Engel Expenditure:
  - Total pengeluaran meningkat secara geometrik
  - Total pengeluaran untuk makanan meningkat secara aritmatik
- Data pengeluaran untuk makanan vs total pengeluaran pada 28 daerah di India
- Linier model:
  - Pengeluaran untuk makanan= f (Total Pengeluaran)
- Lin Log model:
  - Pengeluaran untuk makanan= f(ln Total Pengeluaran)

Total Pengeluaran (X) vs Pengeluaran untuk Makanan (Y)



Linier model



Lin-Log model

Interpretasi Model Lin-Log

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

$$\beta_2 = \frac{dY}{d \ln X} = \frac{dY}{\frac{dX}{X}}$$

$$dY = \beta_2 \frac{dX}{X} = \frac{\beta_2}{100} \boxed{\frac{dX}{X} 100\%}$$

$$\frac{\beta_2}{100} = \frac{1 \text{ unit perubahan } Y}{1\% \text{ perubahan } X}$$

1% perubahan *X*

Model Reciprocal

$$Y_t = \beta_0 + \beta_1\left(\frac{1}{X_t}\right) + u_t$$

- Model ini merupakan model nonlinier dalam variabel X karena variabel ini merupakan model terbalik atau secara *reciprocal*, namun linier dalam parameter (β_0 dan β_1).
- Jika X meningkat dalam waktu yang tak terbatas, nilai $\beta_1(1/X)$ akan mendekati nol dan Y mendekati nilai batas atau nilai asimtotik β_0.

Model Polinomial

- Dapat melalui pemisalan
- Tidak dapat secara langsung diterjemahkan pengaruhnya.
- Dapat dicari elastisitas rata-ratanya dengan cara menurunkannya.

Pemilihan Model

- Didasarkan pada model teoritis
- Dilihat dari hasil estimasi
  - "Goodness of fit" → koefisien determinasi (r^2), pilih yang paling tinggi.
  - Tanda dan siknifikansi koefisien regresi → tanda sesuai teori dan yang siknifikan.
  - Kriteria informasi → Akaike, pilih yang paling kecil.

Pemilihan bentuk fungsional berdasarkan perbandingan nilai R^2

- Perbandingan dua nilai R^2 boleh dilakukan pada:
  - Dua atau beberapa model dengan variabel dependen (Y) dengan bentuk fungsional yang sama
  - Ukuran sampel yang sama
  - Bentuk fungsional variabel independen boleh berbeda

- Semakin tinggi R^2 tidak berarti semakin baik modelnya
- Yang utama dalam pemilihan model
  - Kesesuaian tanda dari estimator parameter dengan teori ekonomi yang mendasari
  - Signifikansi parameter secara statistik

Pemilihan bentuk fungsional berdasarkan perbandingan nilai R^2 (lanjutan)

- Peneliti harus lebih memperhatikan hubungan logis/teoritis dari variabel independen terhadap variabel dependen
- Jika estimator parameter signifikan, dengan tanda sesuai dengan teori:
  - Model tetap dianggap baik walaupun R^2 kecil.

Materi Tambahan dari Buku Studenmund

- Slide-slide berikut materi dari buku Studenmund: Using Econometrics

The Use and Interpretation of the Constant Term

- An estimate of β_0 has at least **three components**:
  1. the true β_0
  2. the constant impact of any specification errors (an omitted variable, for example)
  3. the mean of ε for the correctly specified equation (if not equal to zero)
- Unfortunately, these components can't be distinguished from one another because we can observe only β_0, the sum of the three components
- As a result of this, we usually **don't interpret** the constant term
- On the other hand, we should **not suppress** the constant term, either, as illustrated by Figure 7.1

© 2011 Pearson Addison-Wesley. All rights reserved.

Figure 7.1 The Harmful Effect of Suppressing the Constant Term

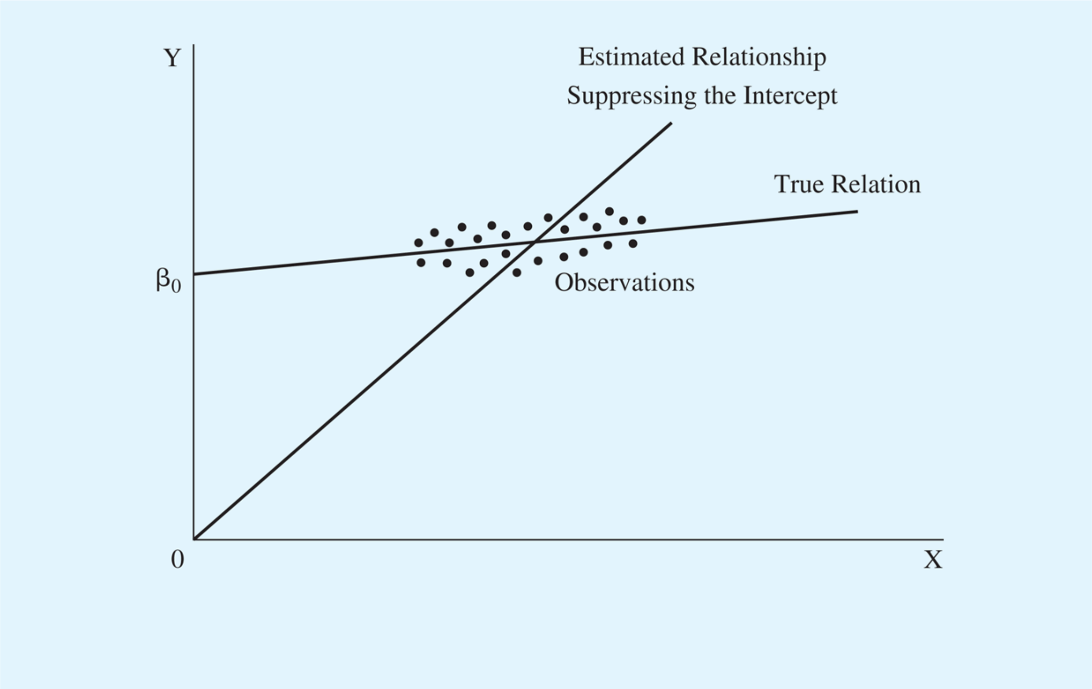


Figure 7.1 The Harmful Effect of Suppressing the Constant Term

If the constant (or intercept) term is suppressed, the estimated regression will go through the origin. Such an effect potentially biases the $\hat{\beta}$s and inflates their *t*-scores. In this particular example, the true slope is close to zero in the range of the sample, but forcing the regression through the origin makes the slope appear to be significantly positive.

© 2011 Pearson Addison-Wesley. All rights reserved.

Alternative Functional Forms

- An equation is **linear** in the **variables** if plotting the function in terms of X and Y generates a **straight line**
- For example, Equation 7.1:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (7.1)$$

is linear in the variables but Equation 7.2:

$$Y = \beta_0 + \beta_1 X^2 + \varepsilon \qquad (7.2)$$

is **not** linear in the variables

- Similarly, an equation is **linear** in the **coefficients** only if the **coefficients** appear in their simplest form—they:
  - are **not raised** to any powers (other than one)
  - are **not multiplied** or **divided** by other coefficients
  - do **not** themselves **include** some sort of **function** (like **logs** or **exponents**)

© 2011 Pearson Addison-Wesley. All rights reserved.

Alternative Functional Forms (cont.)

- For example, Equations 7.1 and 7.2 **are linear** in the coefficients, while Equation 7:3:

$$Y = \beta_0 + X^{\beta_1} \qquad (7.3)$$

is **not linear** in the coefficients

- In fact, of **all possible equations** for a single explanatory variable, **only** functions of the general form:

$$f(Y) = \beta_0 + \beta_1 f(X) \qquad (7.4)$$

are linear in the coefficients β_0 and β_1

© 2011 Pearson Addison-Wesley. All rights reserved.

Linear Form

- This is based on the assumption that the slope of the relationship between the independent variable and the dependent variable is constant:

$$\frac{\Delta Y}{\Delta X_k} = \beta_k \qquad k = 1, 2, \ldots, K$$

- For the linear case, the **elasticity** of Y with respect to X (the percentage change in the dependent variable caused by a 1-percent increase in the independent variable, holding the other variables in the equation constant) is:

$$\text{Elasticity}_{Y,X_k} = \frac{\Delta Y / Y}{\Delta X_k / X_k} = \frac{\Delta Y}{\Delta X_k} \cdot \frac{X_k}{Y} = \beta_k \frac{X_k}{Y}$$

© 2011 Pearson Addison-Wesley. All rights reserved.

What Is a Log?

- If **e** (a **constant** equal to 2.71828) to the "*b*th power" produces x, then b is the **log** of x:

  b is the log of x to the base e if: $e^b = x$

- Thus, a log (or logarithm) is the exponent to which a given base must be taken in order to produce a specific number
- While logs come in more than one variety, we'll use only natural logs (logs to the base e) in this text
- The symbol for a natural log is "ln," so ln(x) = b means that $(2.71828)^b = x$ or, more simply,

  $\ln(x) = b$ means that $e^b = x$

- For example, since $e^2 = (2.71828)^2 = 7.389$, we can state that:

  $\ln(7.389) = 2$

  Thus, the natural log of 7.389 is 2! Again, why? Two is the power of e that produces 7.389

© 2011 Pearson Addison-Wesley. All rights reserved.

What Is a Log? (cont.)

- Let’s look at some other natural log calculations:

  ln(100) = 4.605

  ln(1000) = 6.908

  ln(10000) = 9.210

  ln(1000000) = 13.816

  n(100000) = 11.513

- Note that as a number goes from 100 to 1,000,000, its natural log goes from 4.605 to only 13.816! As a result, logs can be used in econometrics if a researcher wants to **reduce** the **absolute size** of the numbers associated with the same actual meaning
- One useful property of natural logs in econometrics is that they make it easier to figure out impacts in **percentage terms** (we’ll see this when we get to the **double-log** specification)

© 2011 Pearson Addison-Wesley. All rights reserved.

Double-Log Form

- Here, the natural log of Y is the dependent variable and the natural log of X is the independent variable:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \epsilon \qquad (7.5)$$

- In a double-log equation, an individual regression coefficient can be interpreted as an **elasticity** because:

$$\beta_k = \frac{\Delta(\ln Y)}{\Delta(\ln X_k)} = \frac{\Delta Y / Y}{\Delta X_k / X_k} = \text{Elasticity}_{Y,X_k} \qquad (7.6)$$

- Note that the **elasticities** of the model are **constant** and the **slopes** are **not**
- This is in **contrast** to the **linear model**, in which the **slopes** are **constant** but the **elasticities** are **not**

© 2011 Pearson Addison-Wesley. All rights reserved.

Figure 7.2 Double-Log Functions

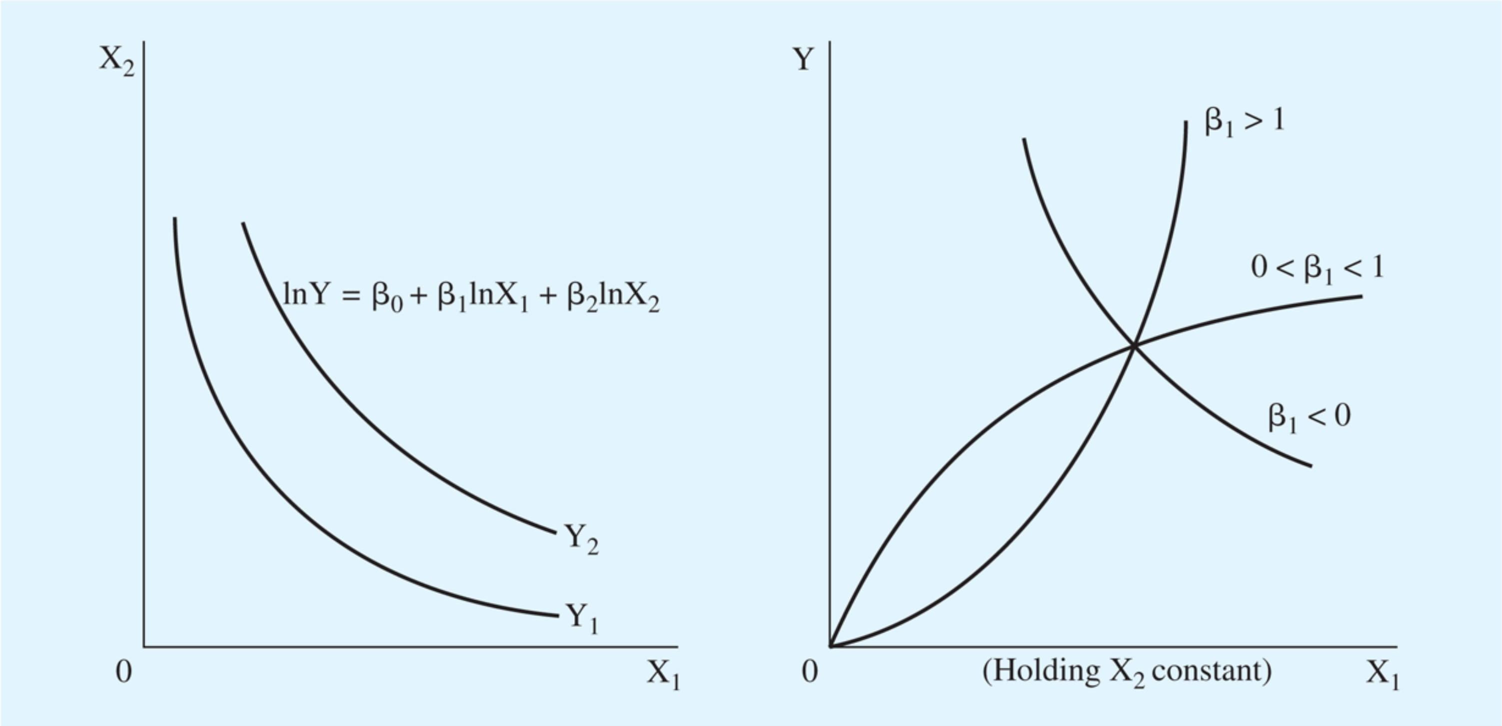


Figure 7.2 Double-Log Functions

Depending on the values of the regression coefficients, the double-log functional form can take on a number of shapes. The left panel shows the use of a double-log function to depict a shape useful in describing the economic concept of an isoquant or an indifference curve. The right panel shows various shapes that can be achieved with a double-log function if X_2 is held constant or is not included in the equation.

© 2011 Pearson Addison-Wesley. All rights reserved.

Semilog Form

- The **semilog** functional form is a variant of the double-log equation in which some but not all of the variables (dependent and independent) are expressed in terms of their natural logs.
- It can be on the **right-hand** side, as in:

$$Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad (7.7)$$

- Or it can be on the **left-hand** side, as in:

$$\ln Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (7.9)$$

- Figure 7.3 illustrates these two different cases

© 2011 Pearson Addison-Wesley. All rights reserved.

Figure 7.3 Semilog Functions

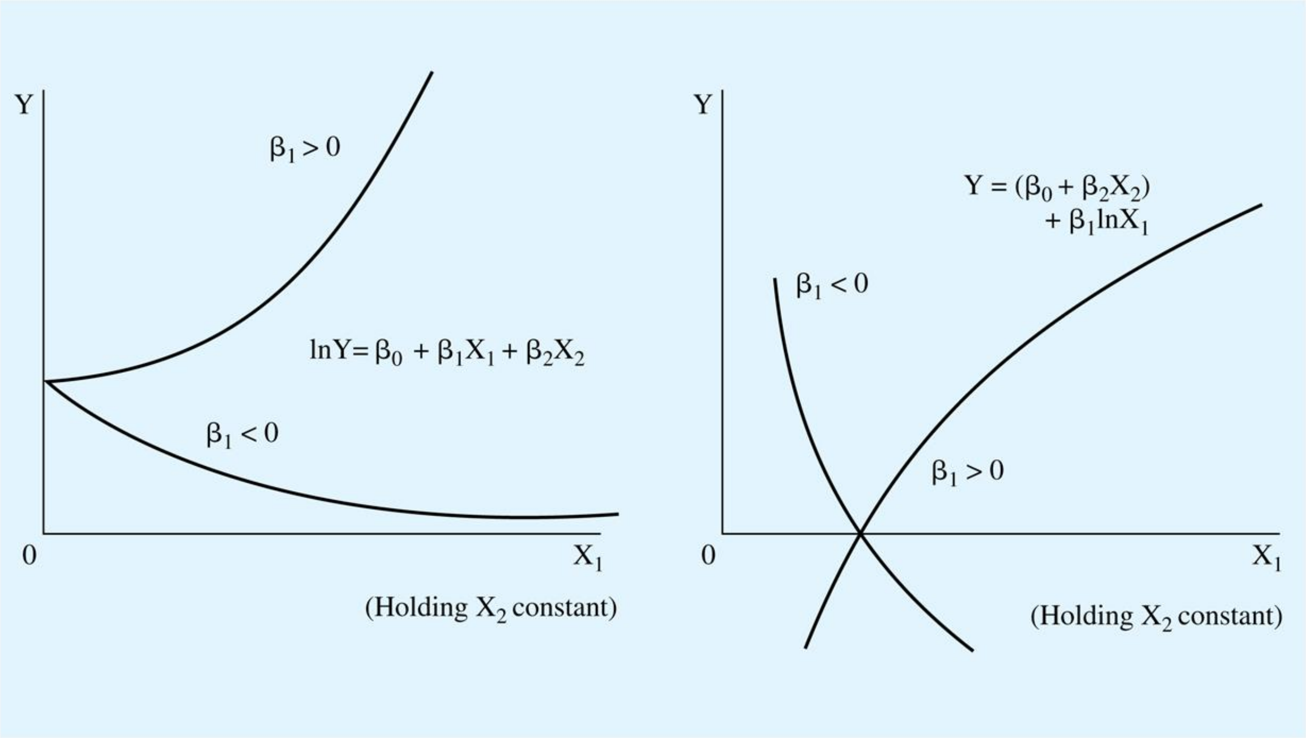


Figure 7.3 Semilog Functions

The semilog functional form on the right (lnX) can be used to depict a situation in which the impact of X_1 on Y is expected to increase at a decreasing rate as X_1 gets bigger as long as β_1 is greater than zero (holding X_2 constant). The semilog functional form on the left (lnY) can be used to depict a situation in which an increase in X_1 causes Y to increase at an increasing rate.

© 2011 Pearson Addison-Wesley. All rights reserved.

Polynomial Form

- **Polynomial functional** forms express Y as a function of independent variables, some of which are raised to powers other than 1
- For example, in a **second-degree** polynomial (also called a quadratic) equation, at least one independent variable is **squared**:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 (X_{1i})^2 + \beta_3 X_{2i} + \varepsilon_i \qquad (7.10)$$

- The **slope** of Y with respect to X_1 in Equation 7.10 is:

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + 2\beta_2 X_1 \qquad (7.11)$$

- Note that the **slope** depends on the **level** of X_1

© 2011 Pearson Addison-Wesley. All rights reserved.

Figure 7.4 Polynomial Functions

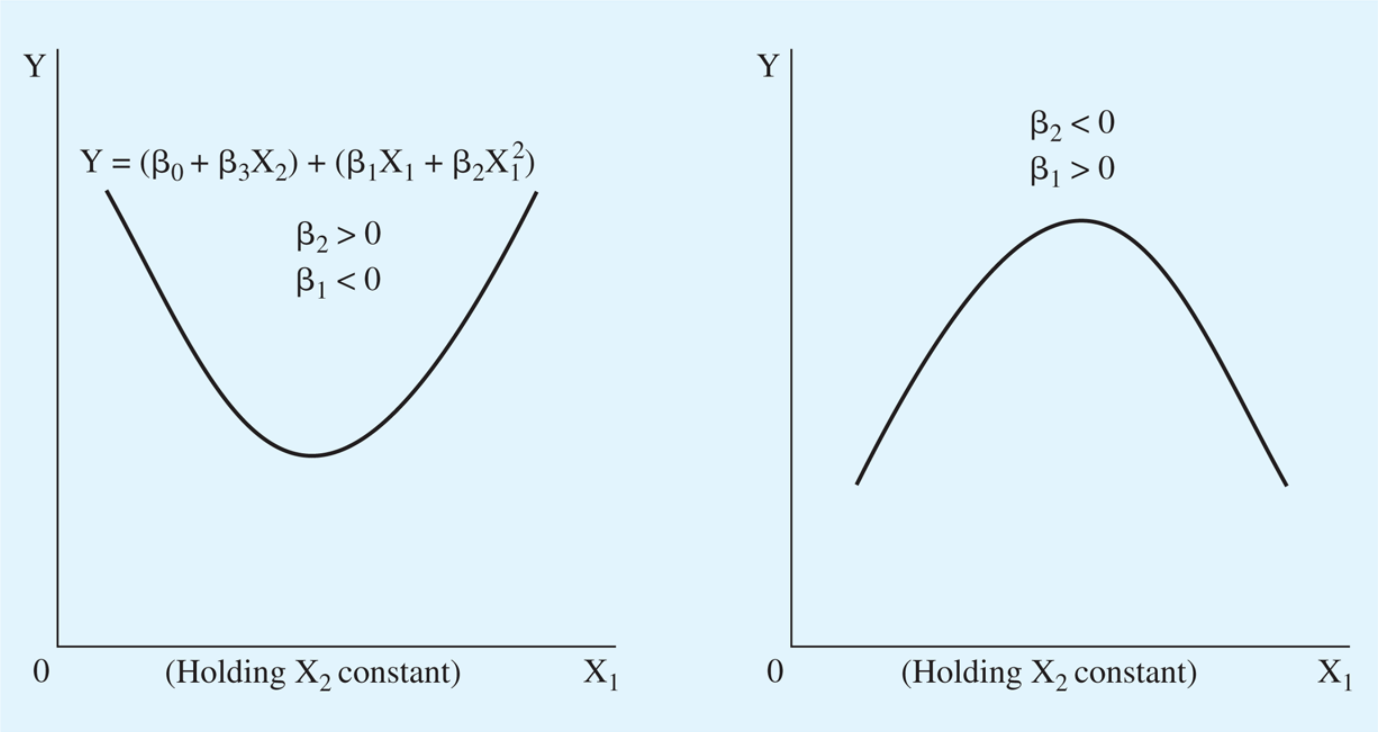


Figure 7.4 Polynomial Functions

Quadratic functional forms (polynomials with squared terms) take on U or inverted U shapes, depending on the values of the coefficients (holding X_2 constant). The left panel shows the shape of a quadratic function that could be used to show a typical cost curve; the right panel allows the description of an impact that rises and then falls (like the impact of age on earnings).

© 2011 Pearson Addison-Wesley. All rights reserved.

Inverse Form

- The **inverse** functional form expresses Y as a function of the **reciprocal** (or **inverse**) of one or more of the independent variables (in this case, X_1):

$$Y_i = \beta_0 + \beta_1(1/X_{1i}) + \beta_2 X_{2i} + \varepsilon_i \qquad (7.13)$$

- So X_1 **cannot equal zero**
- This functional form is relevant when the impact of a particular independent variable is expected to approach zero as that independent variable approaches infinity
- The slope with respect to X_1 is:

$$\frac{\Delta Y}{\Delta X_1} = \frac{-\beta_1}{X_1^2} \qquad (7.14)$$

- The slopes for X_1 fall into **two categories**, depending on the sign of β_1 (illustrated in Figure 7.5)

© 2011 Pearson Addison-Wesley. All rights reserved.

Figure 7.5 Inverse Functions

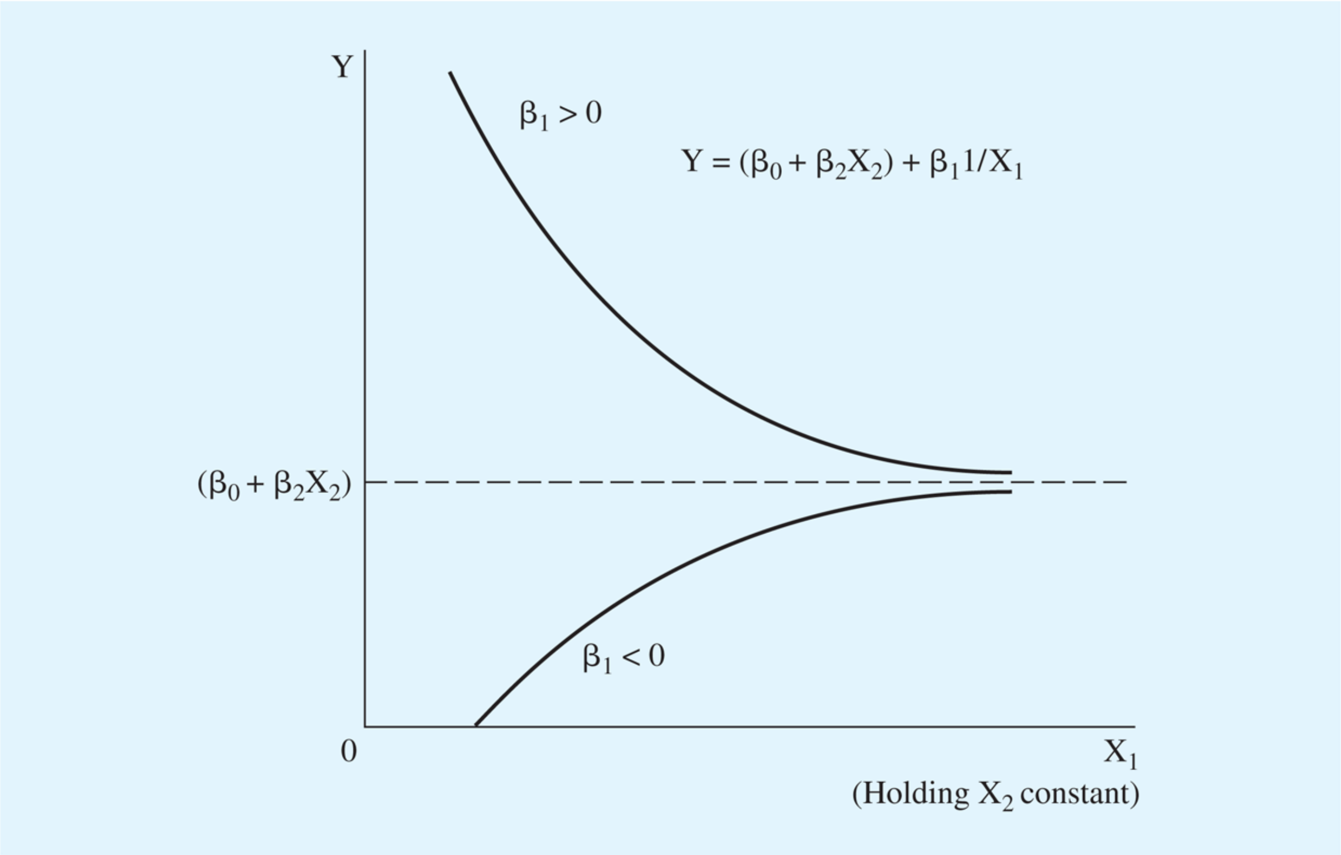


Figure 7.5 Inverse Functions

Inverse (or reciprocal) functional forms allow the impact of an X_1 on Y to approach zero as X_1 increases in size. The inverse function approaches the same value (the asymptote) from the top or bottom depending on the sign of β_1.

© 2011 Pearson Addison-Wesley. All rights reserved.

Table 7.1 Summary of Alternative Functional Forms

Table 7.1 Summary of Alternative Functional Forms

Functional Form	Equation (one X only)	The Meaning of β_1
Linear	$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	The slope of Y with respect to X
Double-log	$\ln Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	The elasticity of Y with respect to X
Semilog (lnX)	$Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	The change in Y (in units) related to a 1 percent increase in X
Semilog (lnY)	$\ln Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	The percent change in Y related to a one-unit increase in X
Polynomial	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$	Roughly, the slope of Y with respect to X for small X
Inverse	$Y_i = \beta_0 + \beta_1 \left(\frac{1}{X_i}\right) + \epsilon_i$	Roughly, the inverse of the slope of Y with respect to X for small X

© 2011 Pearson Addison-Wesley. All rights reserved.