

Παλινδρόμηση

1. Γενικά

Εξετάζοντας την ανάλυση διασποράς (ANOVA) και την παλινδρόμηση, περνάμε από τις μεθόδους της **μονομεταβλητής** (univariate) στην **πολυμεταβλητή** (multivariate) στατιστική ανάλυση.

Με τη μέθοδο της **παλινδρόμησης** (regression) εκτιμούμε ένα **μοντέλο** (μαθηματικό πρότυπο) που αναπαριστά τη σχέση ανάμεσα σε μια **εξαρτημένη** μεταβλητή Y αφενός και μια ή περισσότερες **ανεξάρτητες** μεταβλητές X αφετέρου.

Αν και δεν θα το αναφέρουμε ξανά σε αυτό το φυλλάδιο, εξετάζουμε μόνο μοντέλα **γραμμικής παλινδρόμησης** (linear regression) δηλαδή μαθηματικά πρότυπα στα οποία η σχέση ανάμεσα στο Y και τα X είναι γραμμική.

Όπως και στην ANOVA, η **εξαρτημένη** μεταβλητή πρέπει να είναι **ποσοτική** (quantitative). Όμως, ενώ στην ANOVA οι **ανεξάρτητες** μεταβλητές ήταν ποιοτικές (qualitative), στην παλινδρόμηση πρέπει να είναι, εν γένει, **ποσοτικές**.

Έτσι λοιπόν, η μαθηματική σχέση ανάμεσα σε μια εξαρτημένη μεταβλητή Y και πολλές ανεξάρτητες μεταβλητές X αναπαρίσταται σε γενική μορφή ως εξής:

$$Y = f(X_1, X_2, \dots, X_n)$$

ή

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

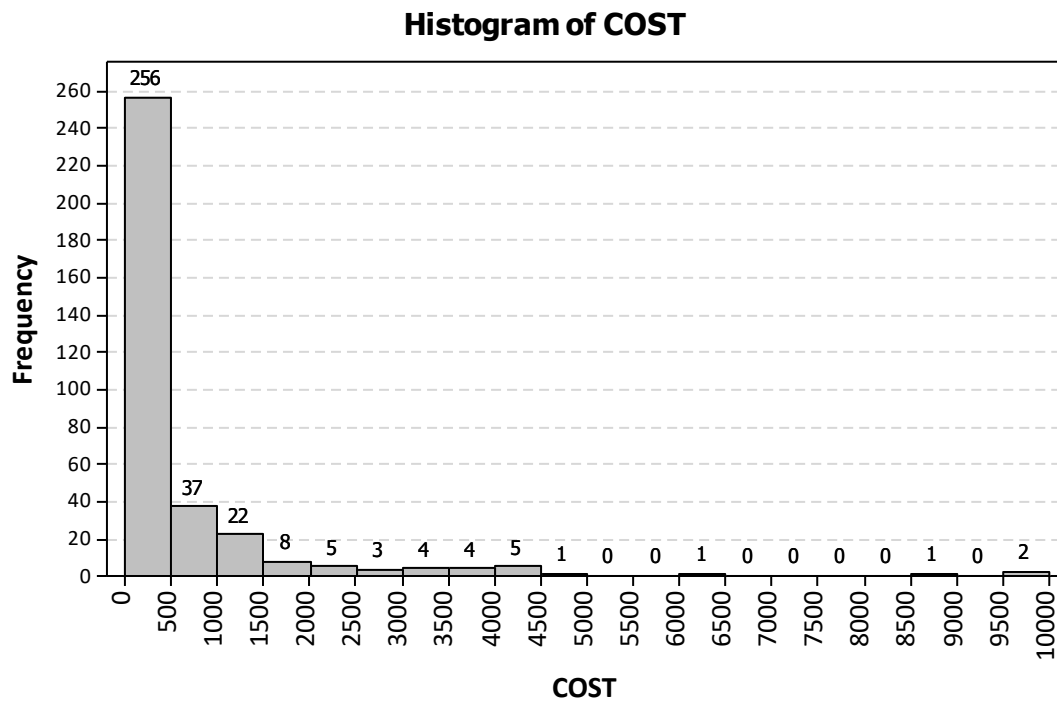
Στα Αγγλικά, η εξαρτημένη μεταβλητή Y εκτός από **dependent** variable λέγεται και **response** (γιατί «ανταποκρίνεται» στις ανεξάρτητες μεταβλητές). Ομοίως, οι ανεξάρτητες μεταβλητές X εκτός από **independent** variables λέγονται και **predictors** (επειδή μπορούν να «προβλέψουν» τις τιμές της εξαρτημένης μεταβλητής).

Όταν έχουμε μια μόνο ανεξάρτητη μεταβλητή X , τότε λέμε ότι κάνουμε **απλή παλινδρόμηση** (simple regression). Όταν εξετάζουμε δυο ή περισσότερες ανεξάρτητες μεταβλητές X , τότε έχουμε **πολλαπλή παλινδρόμηση** (multiple regression).

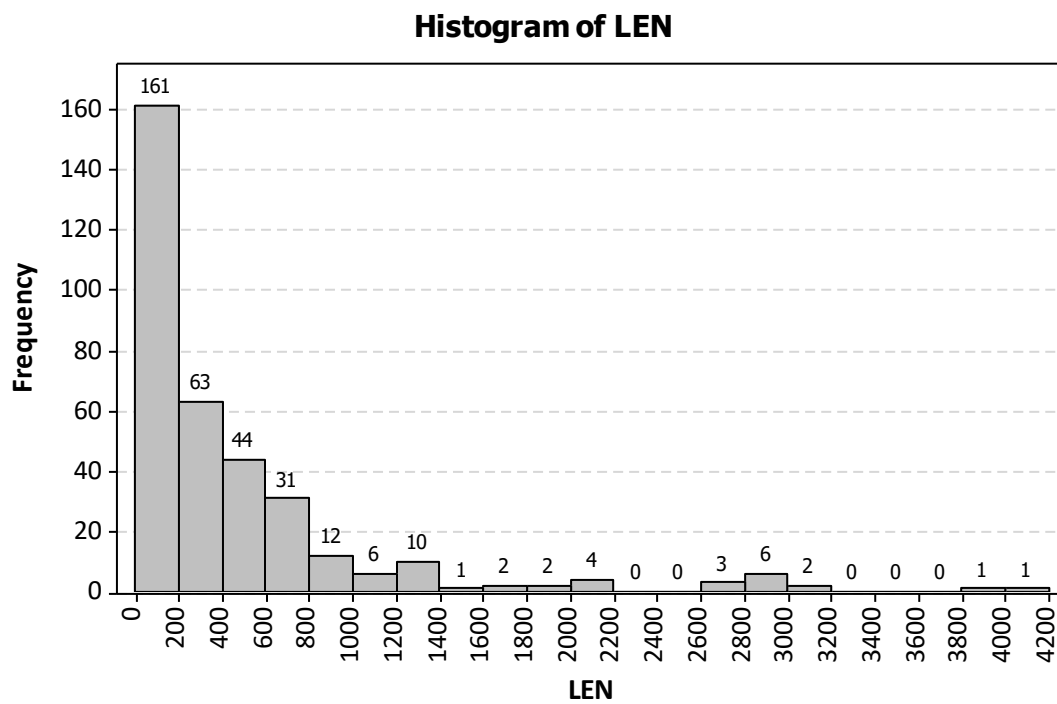
Στις επόμενες ενότητες αυτού του φυλλαδίου, εξετάζουμε την εφαρμογή της μεθόδου της πολλαπλής παλινδρόμησης μέσα από ένα πραγματικό **παράδειγμα**.

2. Γραφική και περιγραφική ανάλυση μεταβλητών

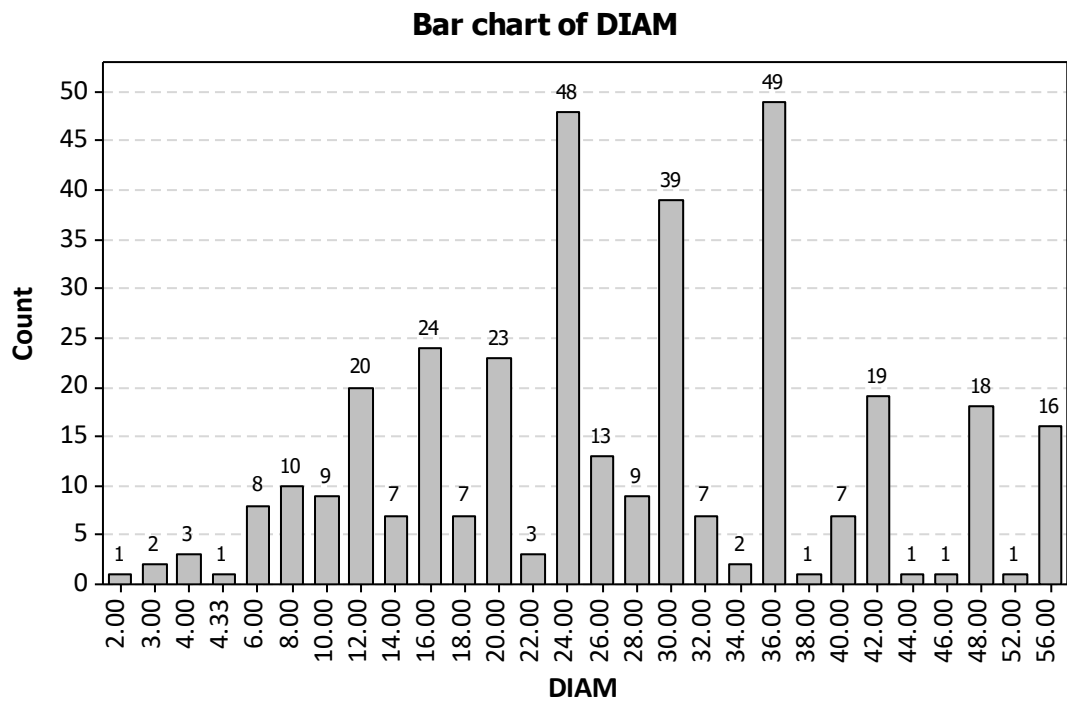
@@



@@



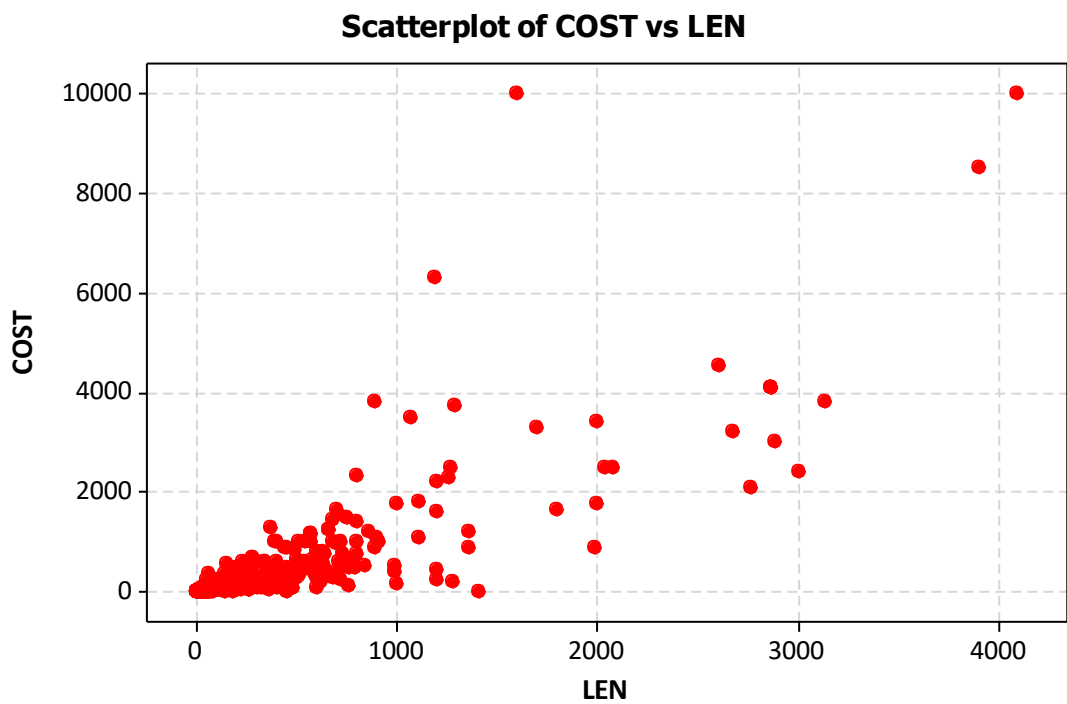
@@



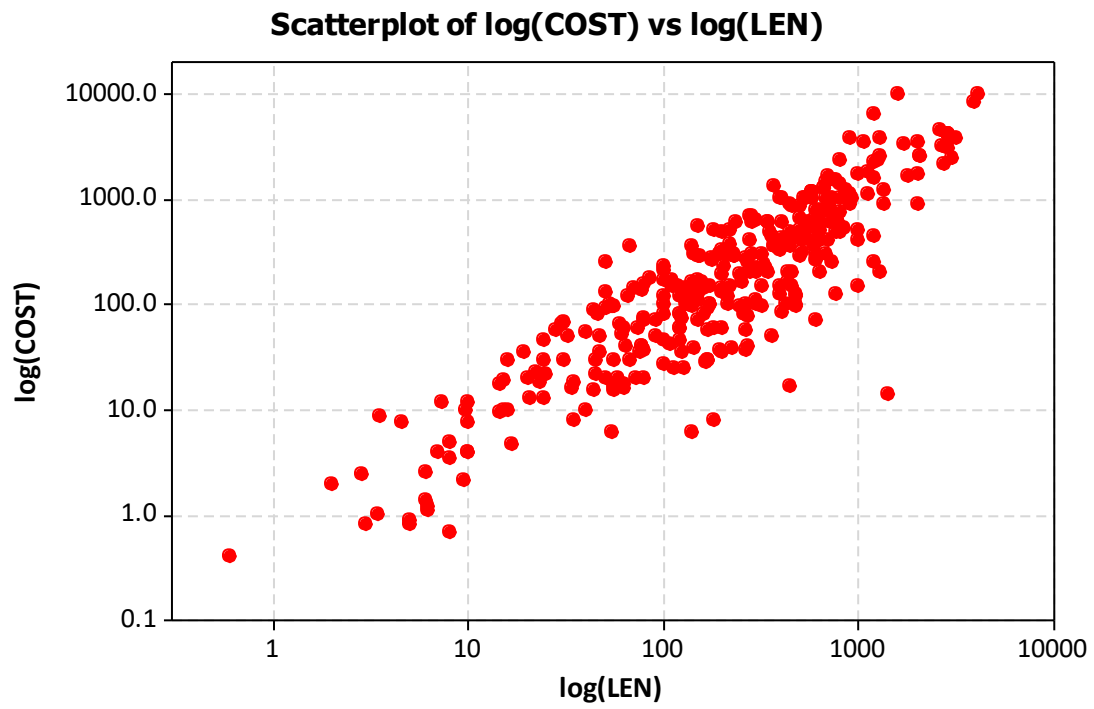
@@

3. Συσχέτιση εξαρτημένης με ανεξάρτητες μεταβλητές

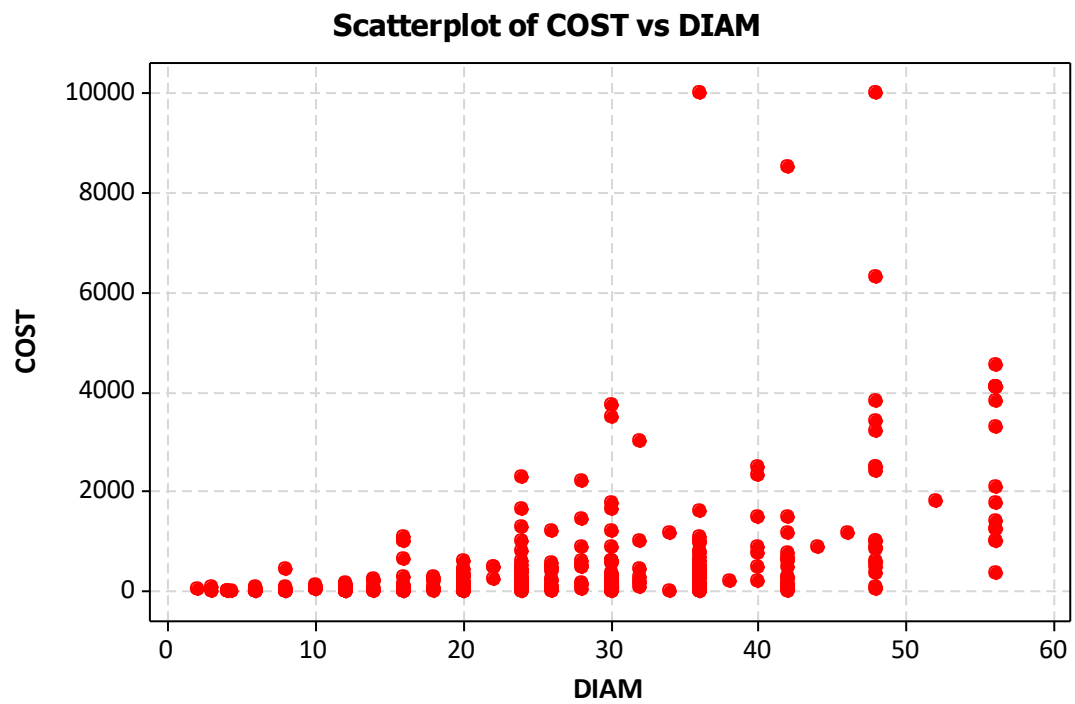
@@



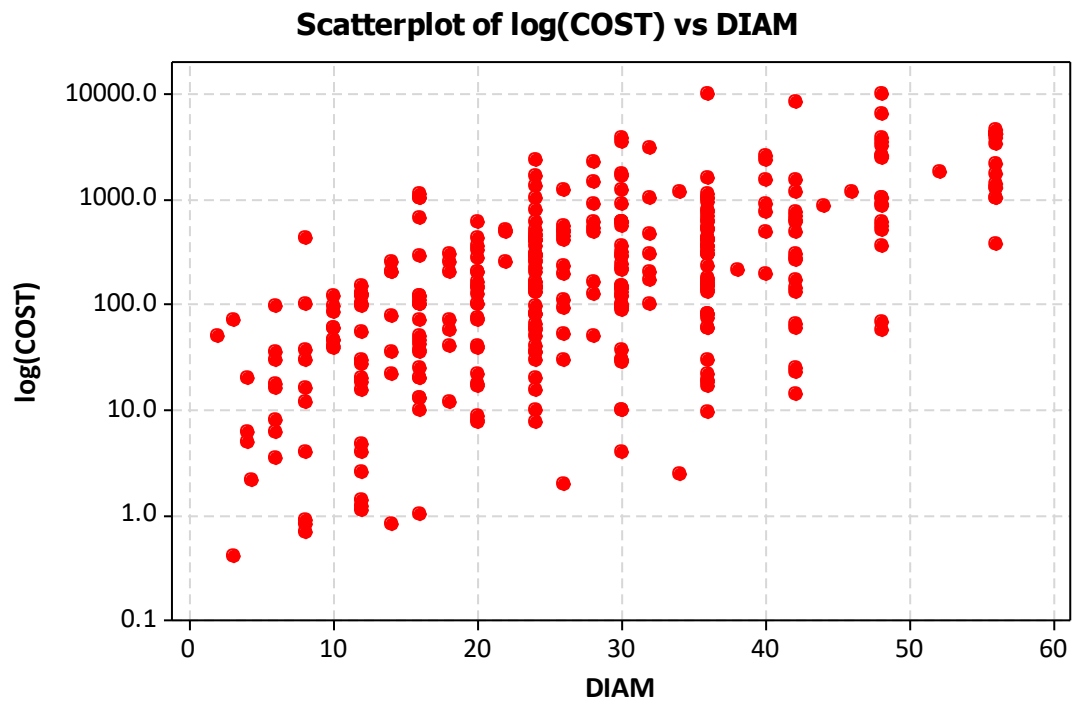
@@



@@



@@



@@

3.1. Μετασχηματισμοί μεταβλητών

(Studenmund, 2001) @@

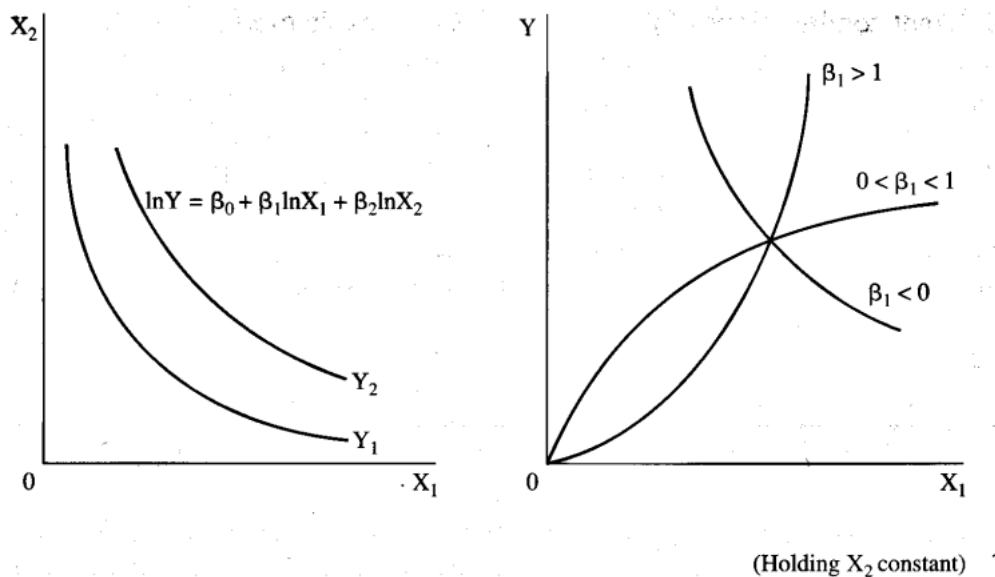


Figure 7.2 Double-Log Functions

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 a production function (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.

@@

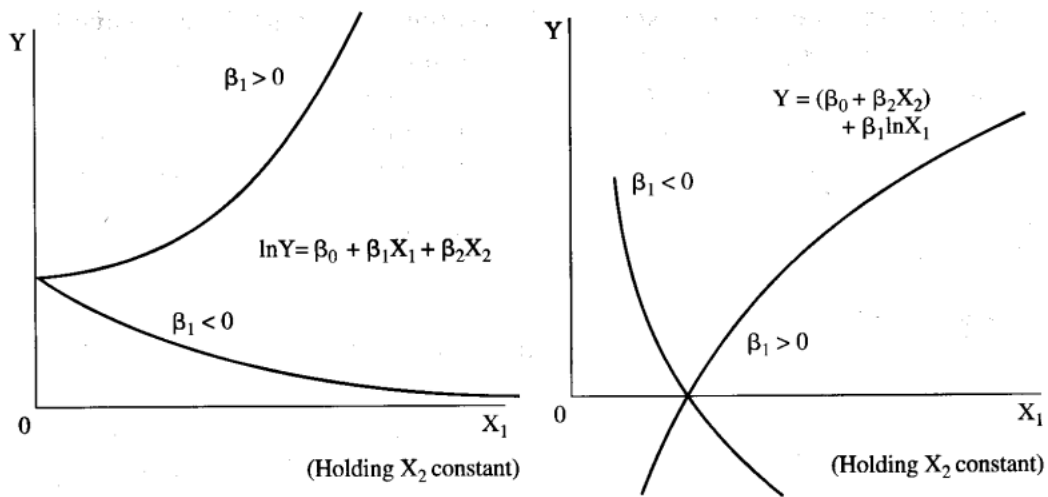


Figure 7.3 Semilog Functions

The semilog functional form on the right ($\ln X$) 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 ($\ln Y$) can be used to depict a situation in which an increase in X_1 causes Y to increase at an increasing rate.

@@

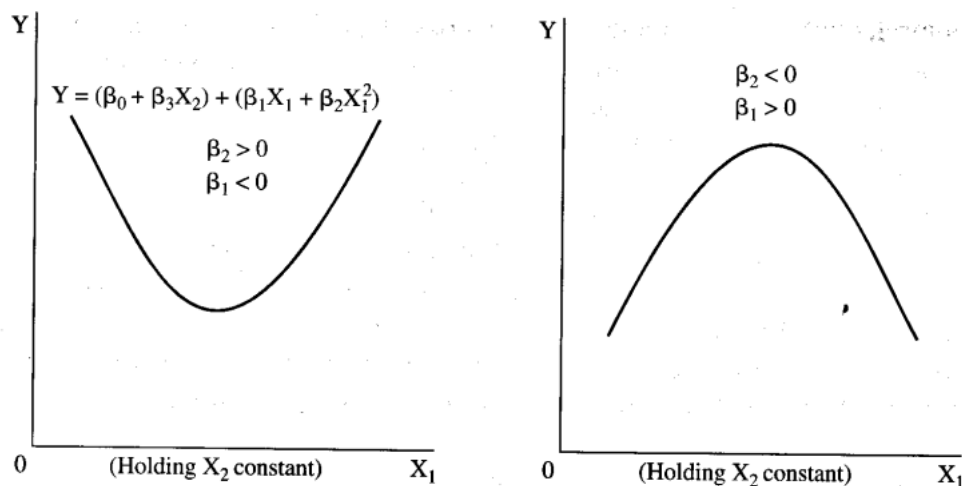


Figure 7.4 Polynomial Functions

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).

@@

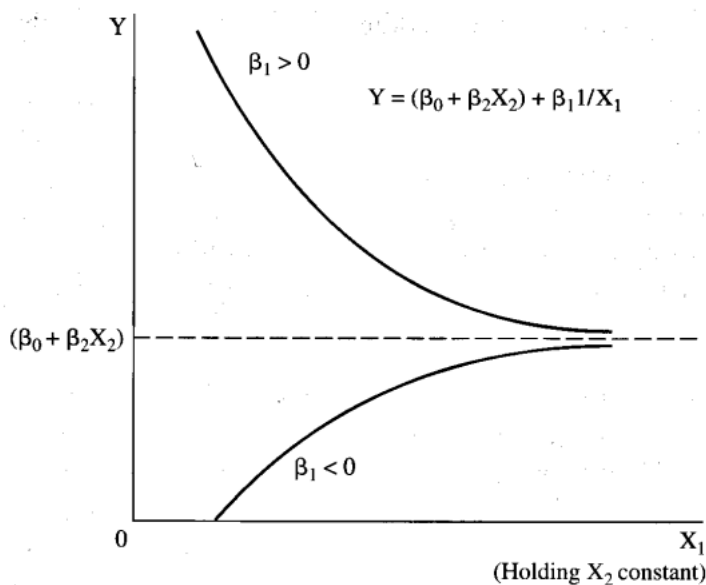


Figure 7.5 Inverse Functions

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 .

@@

TABLE 7.1 SUMMARY OF ALTERNATIVE FUNCTIONAL FORMS

Functional Form	Equation (one X only)	Slope $= \left(\frac{\Delta Y}{\Delta X} \right)$	Elasticity $= \left(\frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y} \right)$
Linear	$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	β_1	$\beta_1 \left(\frac{X_i}{Y_i} \right)$
Double-log	$\ln Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	$\beta_1 \left(\frac{Y_i}{X_i} \right)$	β_1
Semilog (lnX)	$Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	$\beta_1 \left(\frac{1}{X_i} \right)$	$\beta_1 \left(\frac{1}{Y_i} \right)$
Semilog (lnY)	$\ln Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	$\beta_1 Y_i$	$\beta_1 X_i$
Polynomial	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$	$\beta_1 + 2\beta_2 X_i$	$\beta_1 \left(\frac{X_i}{Y_i} \right) + 2\beta_2 \left(\frac{X_i^2}{Y_i} \right)$
Inverse	$Y_i = \beta_0 + \beta_1 \left(\frac{1}{X_i} \right) + \epsilon_i$	$-\beta_1 \left(\frac{1}{X_i^2} \right)$	$-\beta_1 \left(\frac{1}{X_i Y_i} \right)$

Note: Slopes and elasticities that include X_i or Y_i are not constant; they vary from point to point, depending on the value of X_i or Y_i . If general slopes or elasticities are desired, $\bar{X}$ and $\bar{Y}$ can be substituted into the equations.

4. Εκτίμηση μοντέλου πολλαπλής παλινδρόμησης

@@

Βιβλιογραφικές αναφορές

Studenmund, A.H. (2001): *Using Econometric: A Practical Guide*, 4th edition, Addison Wesley Longman.