ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Ανάλυση Δεδομένων στη Λογιστική και Χρηματοοικονομική

Ενότητα $4^{η}$

Περαιτέρω Συμπεράσματα από Το Απλό Γραμμικό Υπόδειγμα Παλινδρόμησης

(Further inference on the simple regression model. Prediction, goodness-of-fit, modelling issues)

Περιγραφή Ενότητας

- Προβλέψεις με τη γραμμική παλινδρόμηση,
- Επεξηγηματική ικανότητα (Goodness-of-fit),
- Θέματα εξειδίκευσης ενός οικονομετρικού μοντέλου,
- Μη καλά ορισμένα υποδείγματα και ανάλυση καταλοίπων,
- Linear-Log, Log-linear and Log-log Models.

Προβλέψεις (Forecasts) και Επεξηγηματική Ικανότητα

Υπάρχουν δύο κυρίως λόγοι για να εκτιμήσουμε ένα οικονομετρικό μοντέλο γραμμικής παλινδρόμησης όπως το ακόλουθο:

$$y_i = \beta_1 + \beta_2 x_i + e_i$$

1. Να προβλέψουμε το *y* δοθέντος του *x*.

2. Να εξηγήσουμε τις αλλαγές στην εξαρτημένη μεταβλητή y_i με τη βοήθεια των αλλαγών στην ανεξάρτητη μεταβλητή x_i.

Προβλέψεις (Forecasts)

Η ικανότητα ενός οικονομετρικού μοντέλου να παράγει (σχετικά) αξιόπιστες προβλέψεις είναι σημαντική:

- Για τους αναλυτές οικονομολόγους οι οποίοι ενδιαφέρονται για προβλέψεις των πωλήσεων των επιχειρήσεων
- Για τους κυβερνητικούς οργανισμούς (και όχι μόνο) οι οποίοι ασκούνε πολιτική και θέτουν (προτείνουν) νομοθετικά πλαίσια. Για παράδειγμα τέτοιοι οργανισμοί ενδιαφέρονται για τη μεταβολή στο ΑΕΠ, τον πληθωρισμό, τις επενδύσεις, τις αποταμιεύσεις, τα έξοδα για την κοινωνική ασφάλιση και τα φορολογικά έσοδα.
- Για τις τοπικές επιχειρήσεις οι οποίες ενδιαφέρονται να προβλέψουν τη μεταβολή στον τοπικό πληθυσμό και το εισόδημα του ώστε να διαμορφώσουν αντίστοιχα την ικανότητα παραγωγής τους και τη στρατηγική τους.

Με άλλα λόγια, οι αξιόπιστες προβλέψεις ενός οικονομετρικού μοντέλου οδηγούν σε ορθότερη λήψη αποφάσεων και επιχειρηματικό (χρηματοοικονομικό) προγραμματισμό.

Προβλέψεις (Forecasts)

- Παρατηρήστε ότι όσο σημαντικές είναι οι σωστές προβλέψεις τόσο εξίσου δύσκολες είναι.

- Για αυτό η προβλεπτική ικανότητα ενός μοντέλου πρέπει να αξιολογείται με αρκετά και διαφορετικά μέτρα προβλεπτικής ικανότητας.

- Σημειώστε ότι
  - (i) η πρόβλεψη στις κοινωνικές-οικονομικές επιστήμες είναι εξαιρετικά δύσκολη καθώς τα φαινόμενα είναι δυναμικά,
  - (ii) ένα αξιόπιστο μοντέλο πρόβλεψης είναι στατιστικά (δηλαδή κατά μέσο όρο) αξιόπιστο, δηλαδή μπορεί να πέφτει έξω αρκετά συχνά αλλά κατά μέσο όρο να είναι σωστό.

Προβλέψεις (Forecasts)

- Ένα μοντέλο γραμμικής παλινδρόμησης μπορεί να παράγει προβλέψεις για τη μεταβλητή Y, δοθείσης μίας τιμής για τη μεταβλητή X.
- Για να υπάρχει κάποιος βαθμός αξιοπιστίας στο μοντέλο (και άρα και στις προβλέψεις που αυτό κάνει) θα πρέπει να ισχύουν οι υποθέσεις του γραμμικού μοντέλου παλινδρόμησης που είδαμε.

$$y_0 = \beta_1 + \beta_2 x_0 + e_{0,}$$ όπου e_0 είναι ένας τυχαίος όρος.

- Εφόσον $y_0 = E(y_0) + e_0 = \beta_1 + \beta_2 x_0 + e_0$
- Η πρόβλεψη για το $\boldsymbol{y_0}$ θα είναι ισοδύναμη με το να εκτιμήσουμε το $\boldsymbol{E(y_0)} = \beta_1 + \beta_2 x_{0,}$ κάτω από τις υποθέσεις ότι:

$$E(e_0)=0, Var \quad (e_0)= \sigma^2, Cov(e_0, e_i)= 0$$

- Παρά το γεγονός ότι το $E(y_0) = \beta_1 + \beta_2 x_0$ δεν είναι τυχαίο, το αποτέλεσμα για το y_0 είναι τυχαίο (περιέχει τον τυχαίο όρο e_0). Συνεπώς, όπως θα δούμε, υπάρχει διαφορά μεταξύ του εκτιμημένου εύρους για το $E(y_0) = \beta_1 + \beta_2 x_0$ και του εύρους της πρόβλεψης για το $y_{0.}$

Προβλέψεις (Forecasts)

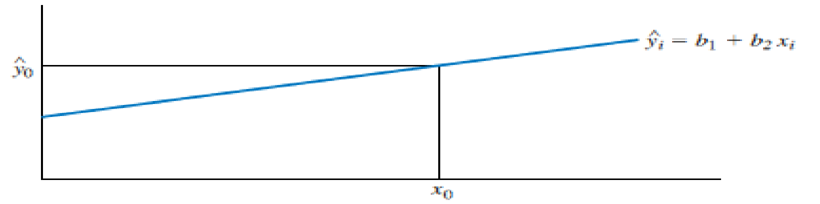
- Για να αξιολογήσουμε πόσο καλά λειτουργεί η προβλεπτική ικανότητα του μοντέλου μπορούμε να ορίσουμε το σφάλμα πρόβλεψης (forecast errors) το οποίο είναι ανάλογο των εκτιμημένων καταλοίπων

$$f = y_0 - \hat{y}_0 = \left(\beta_1 + \beta_2 x_0 + e_0\right) - \left(b_1 + b_2 x_0\right)$$

  – Θα θέλαμε το σφάλμα πρόβλεψης να είναι όσο μικρότερο γίνεται, δηλαδή η αναμενόμενη τιμή του σφάλματος πρόβλεψης να είναι μηδέν:

$$\begin{aligned} E(f) &= \beta_1 + \beta_2 x_0 + E(e_0) - \left[E(b_1) + E(b_2)x_0\right] \\ &= \beta_1 + \beta_2 x_0 + 0 - \left[\beta_1 + \beta_2 x_0\right] \\ &= 0 \end{aligned}$$

Το οποίο σημαίνει ότι το σφάλμα πρόβλεψης είναι κατά μέσο όρο μηδέν και άρα ότι το $\hat{y}_0$ μπορεί να προβλέπει αμερόληπτα τις πραγματικές τιμές y_0.



Προβλέψεις (Forecasts)

- Η διακύμανση των προβλέψεων είναι:

$$\operatorname{var}(f) = \sigma^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$$

- Η διακύμανση είναι μικρότερη (επιθυμητή ιδιότητα) όταν:
  - Η συνολική αβεβαιότητα για το μοντέλο είναι μικρότερη (δηλαδή η διακύμανση των τυχαίων σφαλμάτων, σ^2).
  - Το μέγεθος του δείγματος N είναι μεγαλύτερο.
  - Οι μεταβολές στην επεξηγηματική μεταβλητή (x_i) είναι μεγαλύτερες.

Επεξηγηματική Ικανότητα

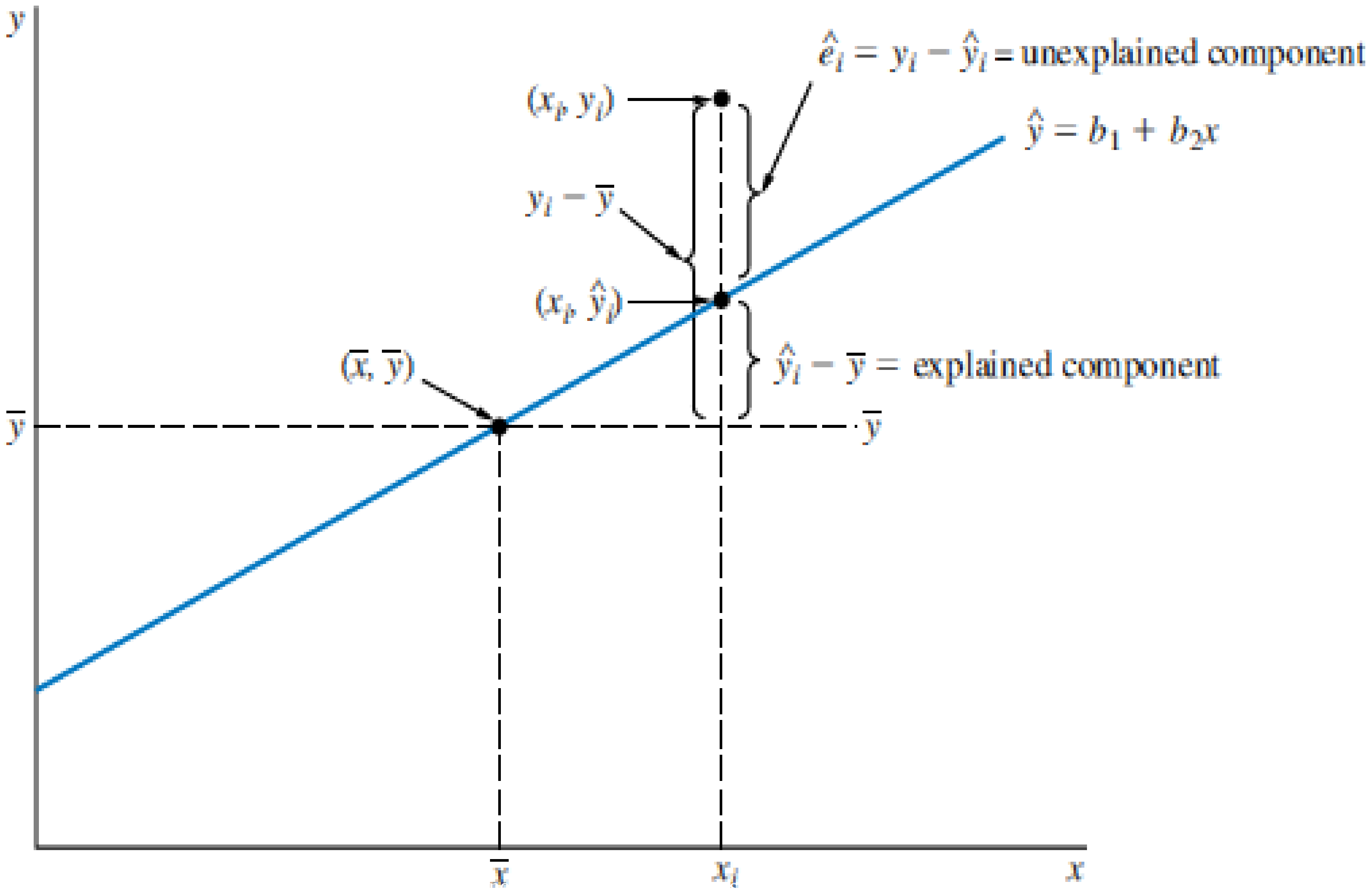
- Ένα εξίσου σημαντικό πρόβλημα με τη διενέργεια αξιόπιστων προβλέψεων είναι η επεξηγηματική ικανότητα ενός μοντέλου. Η επιθυμητή ιδιότητα είναι η ανεξάρτητη μεταβλητή x_i να εξηγεί όσο περισσότερο γίνεται από τις μεταβολές της εξαρτημένης μεταβλητής y_i.
- Η μεταβλητή y_i μπορεί να διαχωριστεί σε ένα μέρος που είναι εξηγήσιμο και ένα ανεξήγητο ως ακολούθως:

$$y_i = E(y_i) + e_i$$

  - Όπου το $E(y_i)$ είναι το εξηγήσιμο (ή συστηματικό) μέρος
  - Και όπου e_i i είναι το τυχαίο, μη-συστηματικό και ανεξήγητο μέρος
  - Την παραπάνω εξίσωση μπορούμε να τη γράψουμε και ως: $y_i = \hat{y}_i + \hat{e}_i$
  - Και αφαιρώντας το δειγματικό μέσο $(\bar{y})$και από τις δύο πλευρές:

$$y_i - \bar{y} = \left(\hat{y}_i - \bar{y}\right) + \hat{e}_i$$

Explained and unexplained components of y_i



Επεξηγηματική Ικανότητα

- Υψώνοντας στο τετράγωνο (squaring) και αθροίζοντας (summing) και τις δύο πλευρές της προηγούμενης εξίσωσης
  (επίσης εκμεταλλευόμενοι ότι $\sum(\hat{y}_i - \bar{y})\hat{e}_i = 0$):

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum \hat{e}_i^2$$

- Η παραπάνω εξίσωση κατηγοριοποιεί τις μεταβολές (variation) του συνολικού δείγματος (total sample variation) σε δύο μέρη το εξηγήσιμο και το μη εξηγήσιμο. Αυτά τα ονομάζουμε αθροίσματα τετραγώνων (sums of squares) ως εξής:

$$\sum(y_i - \bar{y})^2 = \text{total sum of squares} = \text{SST}$$

$$\sum(\hat{y}_i - \bar{y})^2 = \text{sum of squares due to regression} = \text{SSR}$$

$$\sum \hat{e}_i^2 = \text{sum of squares due to error} = \text{SSE}$$

Ο συντελεστής προσδιορισμού R²

- Από τα προηγούμενα είναι φανερό ότι: $\text{SST} = \text{SSR} + \text{SSE}$
- Ο συντελεστής προσδιορισμού (coefficient of determination, R^2) μετρά τι ποσοστό των μεταβολών στη μεταβλητή y μπορεί να εξηγηθεί από τις μεταβολές στη μεταβλητή χ (η τιμή του κυμαίνεται: $0 < R^2 < 1$):

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

– Όσο πιο κοντά είναι στη μονάδα, τόσο πιο κοντά είναι οι παρατηρήσεις της μεταβλητής y στην εκτιμημένη γραμμική εξίσωση παλινδρόμησης. Δηλαδή έχουμε καλύτερο fit του μοντέλου στα δεδομένα (goodness-of-fit).

– Εάν τα δεδομένα για τις y και x είναι μη-συσχετισμένα και δεν έχουν γραμμική σχέση μεταξύ τους, τότε $R^2 = 0$.

R^2 και ρ_{xy}, r_{xy}

- Ο συντελεστής συσχέτισης ρ_{xy} μεταξύ των y και x ορίζεται ως (αντίστοιχα ο r_{xy} είναι ο συντελεστής συσχέτισης για το δείγμα που εξετάζουμε):

$$\rho_{xy} = \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)}\sqrt{\operatorname{var}(y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

where:

$$s_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) / (N-1)$$

$$s_x = \sqrt{\sum (x_i - \bar{x})^2 / (N-1)}$$

$$s_y = \sqrt{\sum (y_i - \bar{y})^2 / (N-1)}$$

- Η σχέση μεταξύ R^2 and r_{xy} είναι: $\boldsymbol{r^2_{xy} = R^2}$

R^2 και ρ_{xy}, r_{xy}: Παράδειγμα

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{304505.176}{495132.160} = 0.385$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{478.75}{(6.848)(112.675)} = 0.62$$

- Έστω ότι $R^2 = 0.385$, τότε συμπεραίνουμε ότι 38.5% των μεταβολών (variation) της επεξηγηματικής μας μεταβλητής (εδώ είναι τα έξοδα για διατροφή ενός νοικοκυριού) εξηγούνται από τις μεταβολές της επεξηγηματικής μεταβλητής (εδώ είναι το εισόδημα του νοικοκυριού).

- Παρατηρήστε και ότι: $r_{xy}^2 = 0.62^2 = 0.385 = R^2$

Το μοντέλο παλινδρόμησης μεταξύ εξόδων διατροφή ενός νοικοκυριού (Y) και εισοδήματος ενός νοικοκυριού (X)

FOOD_EXP = weekly food expenditure by a household of size 3, in dollars

INCOME = weekly household income, in \$100 units

$$\underset{(se)}{FOOD\_EXP} = \underset{(43.41)}{83.42} + \underset{(2.09)^{***}}{10.21} INCOME, \qquad R^2 = 0.385$$

where

* indicates significant at the 10% level

** indicates significant at the 5% level

*** indicates significant at the 1% level

Θέματα εξειδίκευσης ενός οικονομετρικού μοντέλου

Econometric Specification

Επιδράσεις της κλίμακας (Scaling Effects)

- Υπάρχουν μία σειρά από αποφάσεις μοντελοποίησης που μπορεί να προκύψουν σε ένα οικονομετρικό μοντέλο. Για παράδειγμα είναι πολύ σημαντικοί πιθανοί μετασχηματισμοί των μεταβλητών πριν εκτιμηθεί μία γραμμική παλινδρόμηση.
- Ένα από τα πιο γνωστά ζητήματα είναι οι μονάδες στις οποίες αναφέρονται οι μεταβλητές και το εάν θα τις διαιρέσουμε ή όχι με κάποια άλλη μεταβλητή ή με κάποιο αριθμό (effects of scaling the variables).

- Στο παράδειγμα μας με τα έξοδα διατροφής και το εισόδημα ενός νοικοκυριού:
  - Τα εβδομαδιαία έξοδα αναφέρονται σε δολάρια, αλλά το εισόδημα αναφέρεται σε εκατοντάδες δολάρια ($100 units).
  - Άρα το εισόδημα ίσο με $2,000 θα αναφέρεται ως x = 20 (διαιρέσαμε το 2.000 με το 100).

Επιδράσεις της κλίμακας (Scaling Effects): Παράδειγμα

Η εκτίμηση μας με το εισόδημα σε δολλάρια θα ήταν:

$$FOOD\_EXP = 83.42 + 0.1021 INCOME(\$) \qquad R^2 = 0.385$$
$$(\text{se}) \qquad (43.41)(0.0209)^{***}$$

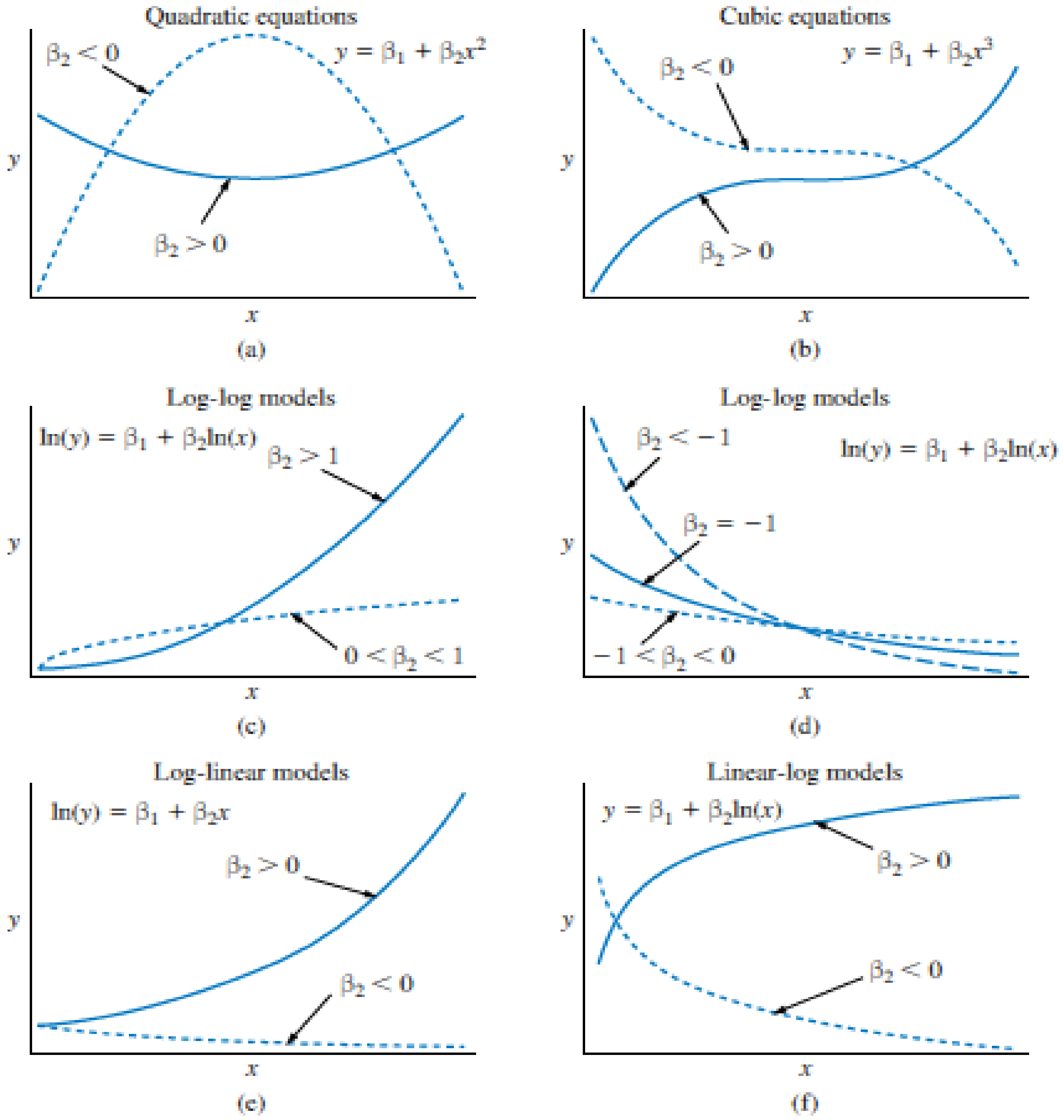
- Παρατηρείστε τις αλλαγές σε σχέση με το αρχικό μοντέλο:
  1. Ο εκτιμημένος συντελεστής της μεταβλητής εισόδημα (income) είναι τώρα 0.1021 (από 10.21 αρχικά).
  2. Τα τυπικά σφάλματα (standard errors, se) γίνονται μικρότερα κατά τον παράγοντα του 100.
- Εφόσον και ο συντελεστής (0.1021) και τα τυπικά σφάλματα μειώθηκαν (διαιρέθηκαν) κατά 100, αυτό αφήνει ανέπαφα τα t-statistics και την τελική ερμηνεία σε ότι αφορά τη στατιστική σημαντικότητα του μοντέλου.

Επιδράσεις της κλίμακας (Scaling Effects)

1. Αν αλλάξουμε την κλίμακα της μεταβλητής χ, τότε όπως είδαμε αλλάζει ο συντελεστής της και το τυπικό της σφάλμα κατά c, τον παράγοντα της κλίμακας (στο παράδειγμα μας 100). Όμως το t-statistic και όλα τα άλλα αποτελέσματα παραμένουν ανεπηρέαστα.
2. Αν αλλάξουμε την κλίμακα της μεταβλητής y, τότε αλλάζουν όλοι οι συντελεστές και τα τυπικά σφάλματα (όπως και τα κατάλοιπα) αλλά δεν αλλάζουν τα t-statistics και το R^2.
3. Αν αλλάξουμε την κλίματα των μεταβλητών χ και y κατά τον ίδιο παράγοντα τότε δεν θα αλλάξει ο συντελεστής της επεξηγηματικής μεταβλητής (b_2) αλλά θα αλλάξει ο εκτιμημένος σταθερός όρος και τα κατάλοιπα. Τα t-statistics και το R^2 παραμένουν.
4. Σε κάθε περίπτωση η οικονομική θεωρία πρέπει να είναι η αφετηρία για ένα οικονομετρικό μοντέλο. Για παράδειγμα θα περιμέναμε μία θετική σχέση μεταξύ εξόδων διατροφής και εισοδήματος καθώς τα είδη διατροφής είναι κανονικά αγαθά. Αλλά η θεωρία δεν λέει ότι θα πρέπει η σχέση τους να είναι γραμμική ….!

Γραμμική σχέση ή όχι ;

- Εάν μετασχηματίσουμε τις X και Y μπορούμε να αλλάξουμε τη σχέση που τις συνδέει σε μη γραμμική (π.χ. Quadratic (X^2), Cubic (X^3), φυσικοί λογάριθμοι, etc.) αλλά να συνεχίσουμε να χρησιμοποιούμε το κλασσικό γραμμικό μοντέλο παλινδρόμησης (χρησιμοποιώντας τις μετασχηματισμένες μεταβλητές X και Y)



Επιδράσεις της κλίμακας (Scaling Effects)

- Τρεις κλασσικοί μετασχηματισμοί είναι οι ακόλουθοι:

1. Στο μοντέλο **linear-log** η μεταβλητή *x* μετασχηματίζεται με τον φυσικό λογάριθμο (ln)

$$y = b_1 + b_2 \ln(x)$$

2. Στο μοντέλο **log-linear** η μεταβλητή *y* μετασχηματίζεται με το φυσικό λογάριθμο (ln)

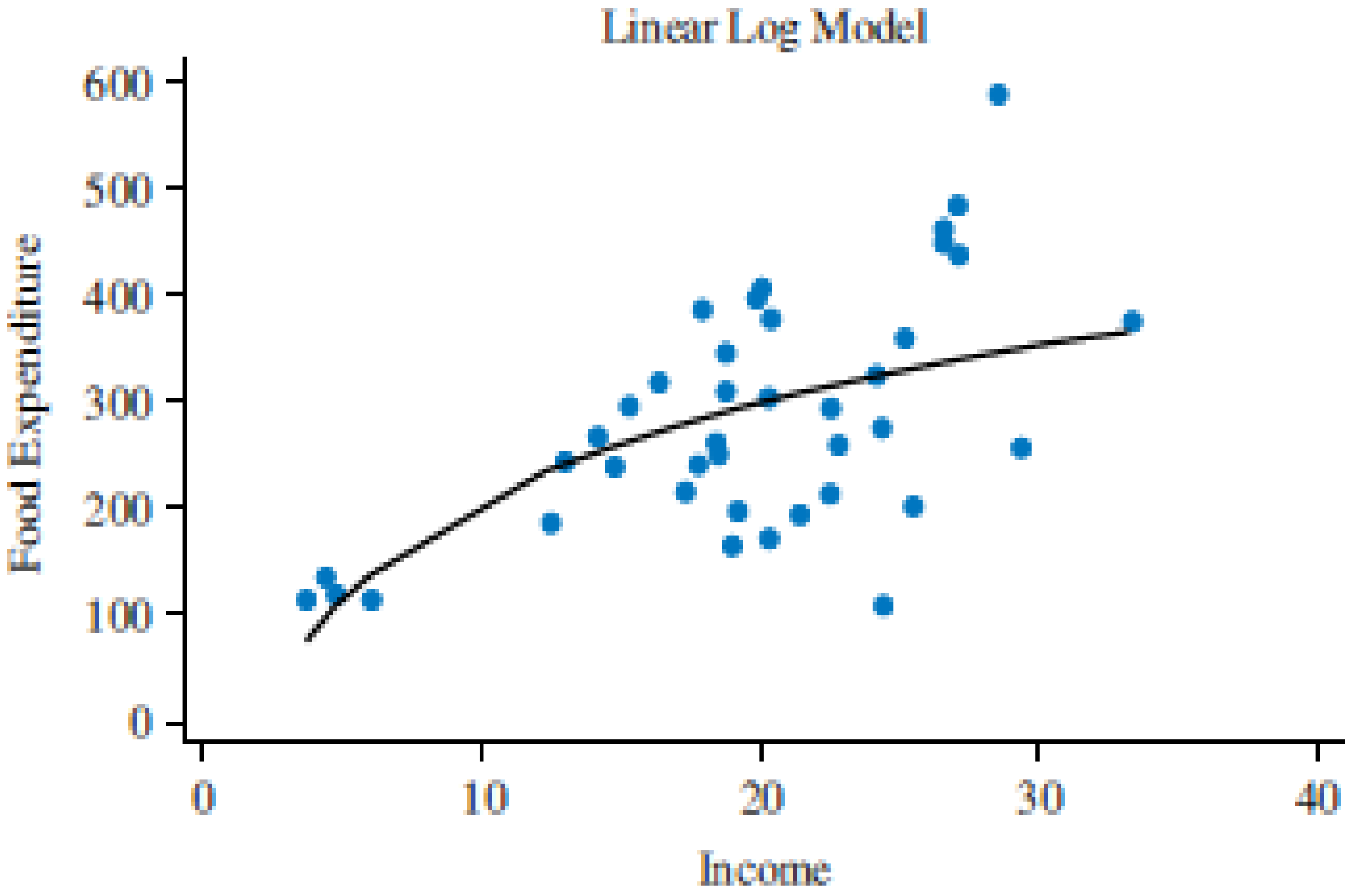
$$\ln(y) = b_1 + b_2 x$$

3. Στο μοντέλο **log-log** και οι δύο μεταβλητές (x και y) μετασχηματίζονται με το φυσικό λογάριθμο (ln). Σε αυτή τη περίπτωση η παράμετρος β_2 είναι η ελαστικότητα της y ως προς τη *x*.

$$\ln(y) = b_1 + b_2 \ln(x)$$

$$\widehat{FOOD\_EXP} = -97.19 + 132.17\ln(INCOME) \qquad R^2 = 0.357$$

$$\text{(se)} \qquad (84.24)\ (28.80)^{***}$$



Linear-Log Interpretation

$$FOOD\_EXP = -97.19 + 132.17 \ln(INCOME) \qquad R^2 = 0.357$$

$$\text{(se)} \qquad (84.24)\ (28.80)^{***}$$

Με την παραπάνω εκτίμηση βλέπουμε ότι ένα νοικοκυριό με εβδομαδιαίο εισόδημα $1,000 θα ξοδεύει κατά μέσο όρο $13.22 περισσότερα σε έξοδα διατροφής για κάθε αύξηση $100 στο εισόδημα του. Ενώ ένα νοικοκυριό με εβδομαδιαίο εισόδημα $2,000 θα ξοδεύει $6.61 περισσότερα σε έξοδα διατροφής για κάθε αύξηση $100 στο εισόδημα του.

– Έτσι βλέπουμε ότι η οριακή επίδραση του εισοδήματος στα έξοδα διατροφής είναι μικρότερη για υψηλότερα επίπεδα εισοδήματος. Αυτή είναι μια αλλαγή από την κλασσική γραμμική παλινδρόμηση, η οποία προέβλεπε ότι για κάθε αύξηση στο εισόδημα κατά $100 τα έξοδα διατροφής θα αυξανόταν κατά $10.21 για οποιοδήποτε ύψος εισοδήματος.
– Εναλλακτικά μπορούμε να πούμε ότι για κάθε 1% αύξηση στο εισόδημα τα έξοδα για διατροφή θα αυξηθούν κατά $1.32 ανά εβδομάδα.

Πως να διαλέξω το σωστό μετασχηματισμό ;

1. Διάλεξε ένα μετασχηματισμό συνεπή με την οικονομική θεωρία σε ότι αφορά τη σχέση των δύο μεταβλητών που εξετάζεις.
2. Διάλεξε ένα μετασχηματισμό ο οποίος να είναι αρκετά ευέλικτος ώστε να περιγράψει τα δεδομένα (fit the data).
3. Διάλεξε ένα μετασχηματισμό ο οποίος να ικανοποιεί τις υποθέσεις του γραμμικού μοντέλου παλινδρόμησής (να οδηγεί σε BLUE coefficients).

Name	Function	Slope = dy/dx	Elasticity
Linear	$y = \beta_1 + \beta_2 x$	β_2	$\beta_2 \frac{x}{y}$
Quadratic	$y = \beta_1 + \beta_2 x^2$	$2\beta_2 x$	$(2\beta_2 x)\frac{x}{y}$
Cubic	$y = \beta_1 + \beta_2 x^3$	$3\beta_2 x^2$	$(3\beta_2 x^2)\frac{x}{y}$
Log-Log	$\ln(y) = \beta_1 + \beta_2 \ln(x)$	$\beta_2 \frac{y}{x}$	β_2
Log-Linear	$\ln(y) = \beta_1 + \beta_2 x$	$\beta_2 y$	$\beta_2 x$
	or, a 1 unit change in x leads to (approximately) a 100 β_2% change in y		
Linear-Log	$y = \beta_1 + \beta_2 \ln(x)$	$\beta_2 \frac{1}{x}$	$\beta_2 \frac{1}{y}$
	or, a 1% change in x leads to (approximately) a $\beta_2/100$ unit change in y		

Μη καλά ορισμένο υπόδειγμα και ανάλυση καταλοίπων

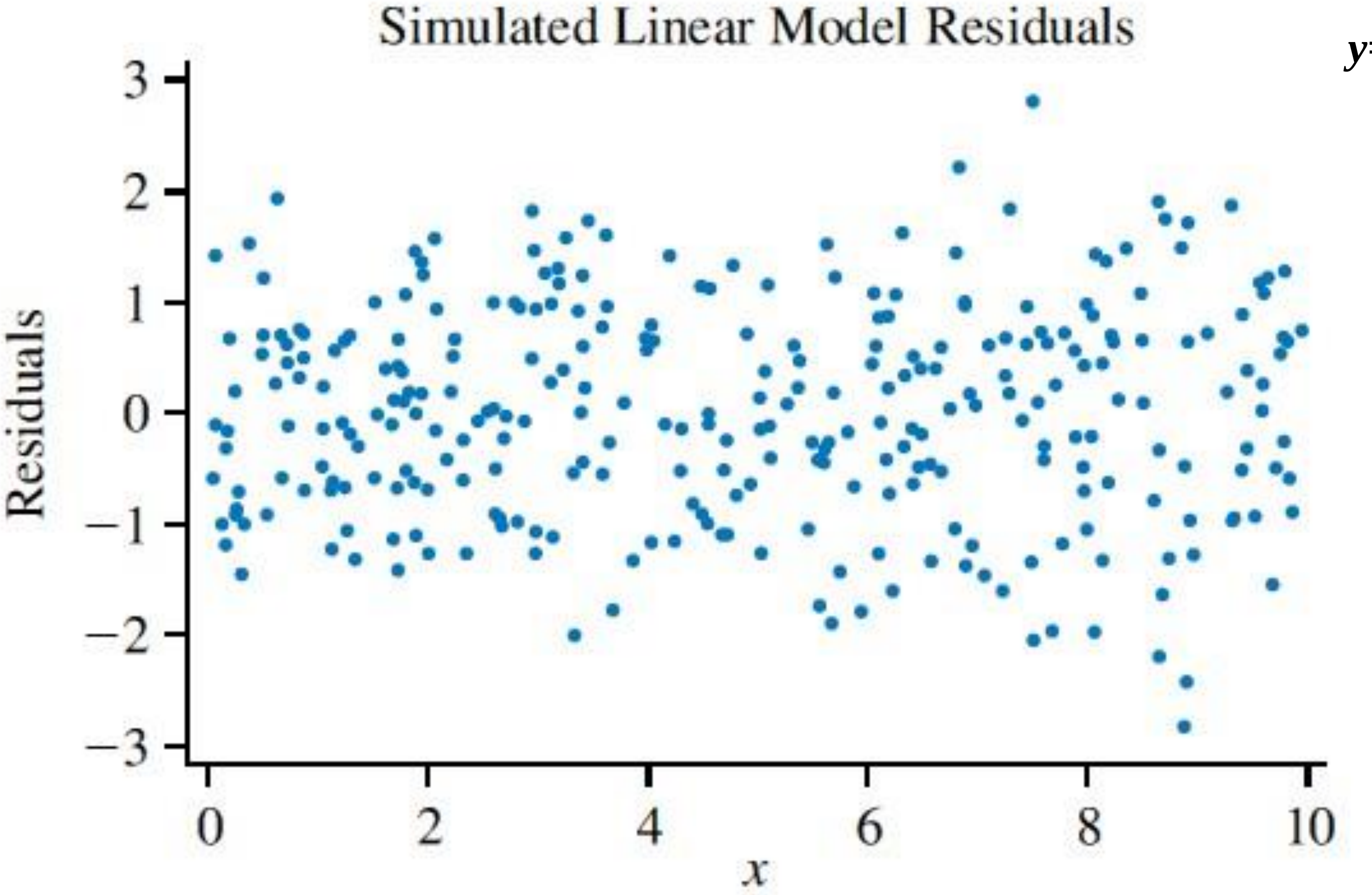
Mis-specified model and analysis of residuals

Ανάλυση Καταλοίπων

- Όταν εξειδικεύουμε ένα οικονομετρικό μοντέλο μπορεί να επιλέξουμε έναν μη επαρκή ή λανθασμένο μετασχηματισμό των μεταβλητών που εξετάζουμε (και άρα και της σχέσης που πιστεύουμε ότι τις συνδέει).

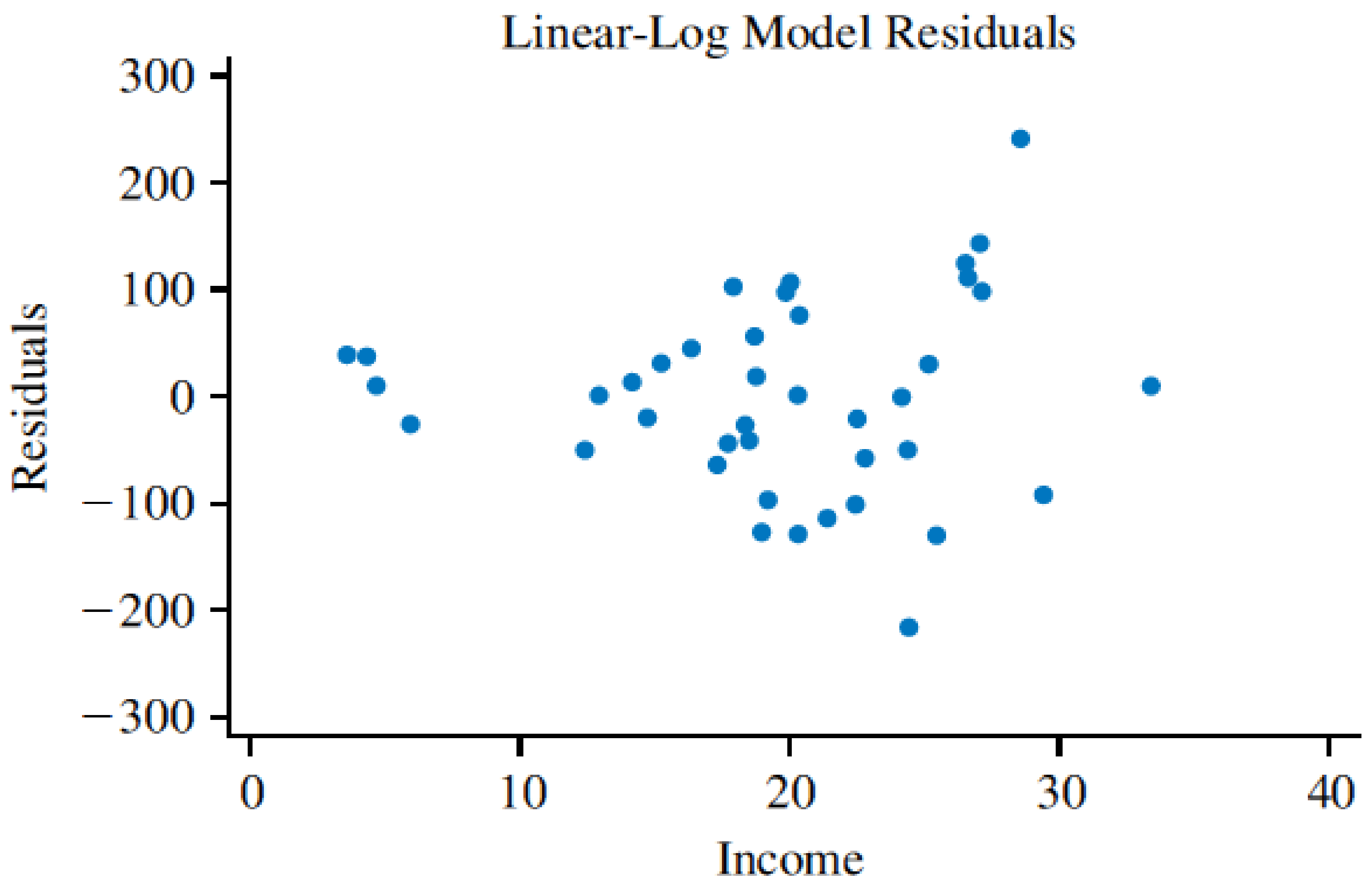
- Για αυτό θα πρέπει να εξετάσουμε τα αποτελέσματα μίας παλινδρόμησης για έννοιες όπως η ετεροσκεδαστικότητα και η αυτοσυσχέτιση (οι οποίες θα μας απασχολήσουν αργότερα) αλλά κυρίως η ανάλυση καταλοίπων

Γράφημα Τυχαιοποιημένων Καταλοίπων
Randomly scattered residuals

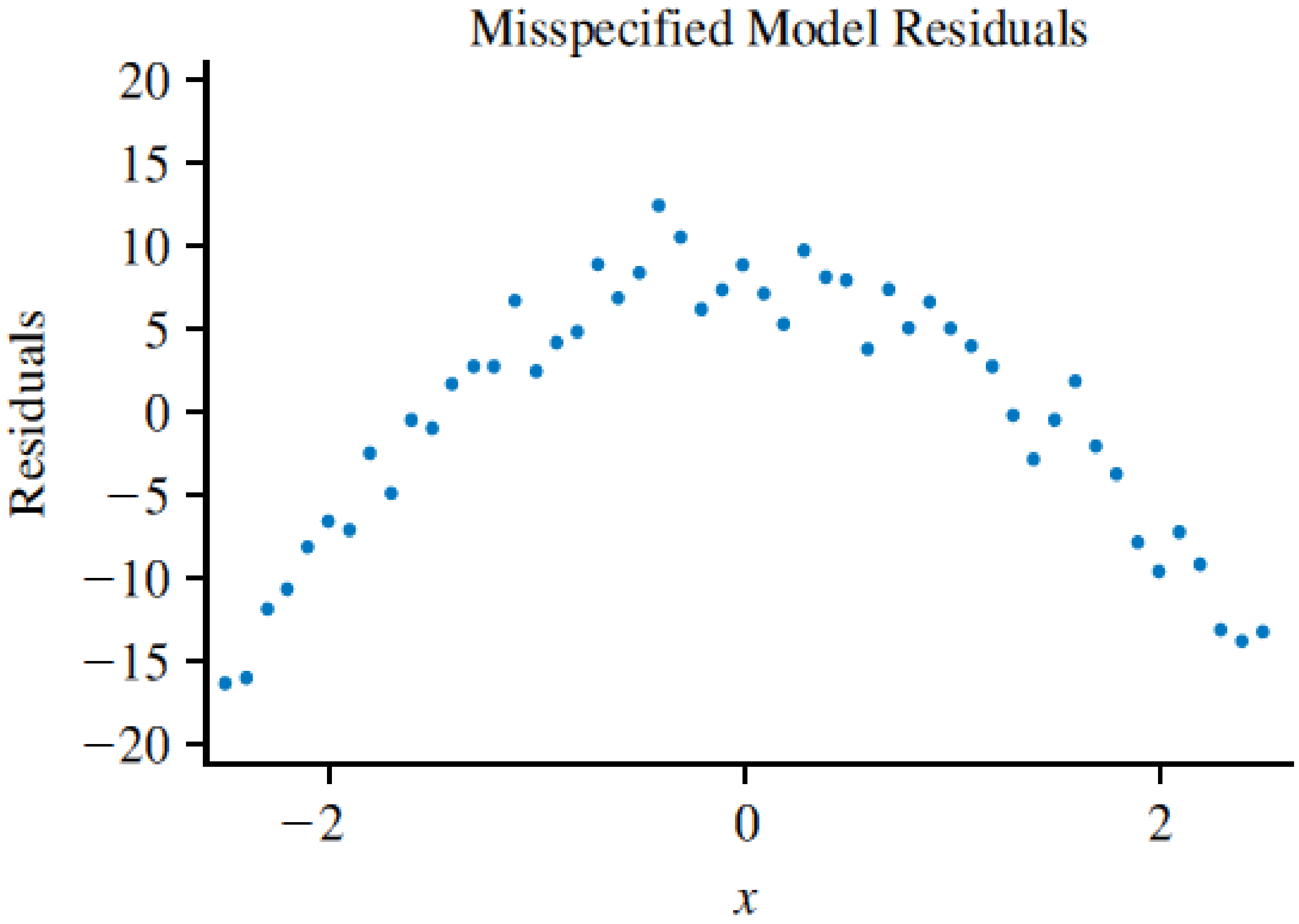


y*=1+*x*+*e

Γράφημα Καταλοίπων από το linear log μοντέλο για τα έξοδα διατροφής
Residuals from linear-log food expenditure model



Για δείτε για παράδειγμα στο παρακάτω γράφημα καταλοίπων φαίνεται καθαρά ότι υπάρχει μία συστηματική σχέση (δευτέρου βαθμού - quadratic) που παραμένει στα εκτιμημένα κατάλοιπα της παλινδρόμησης

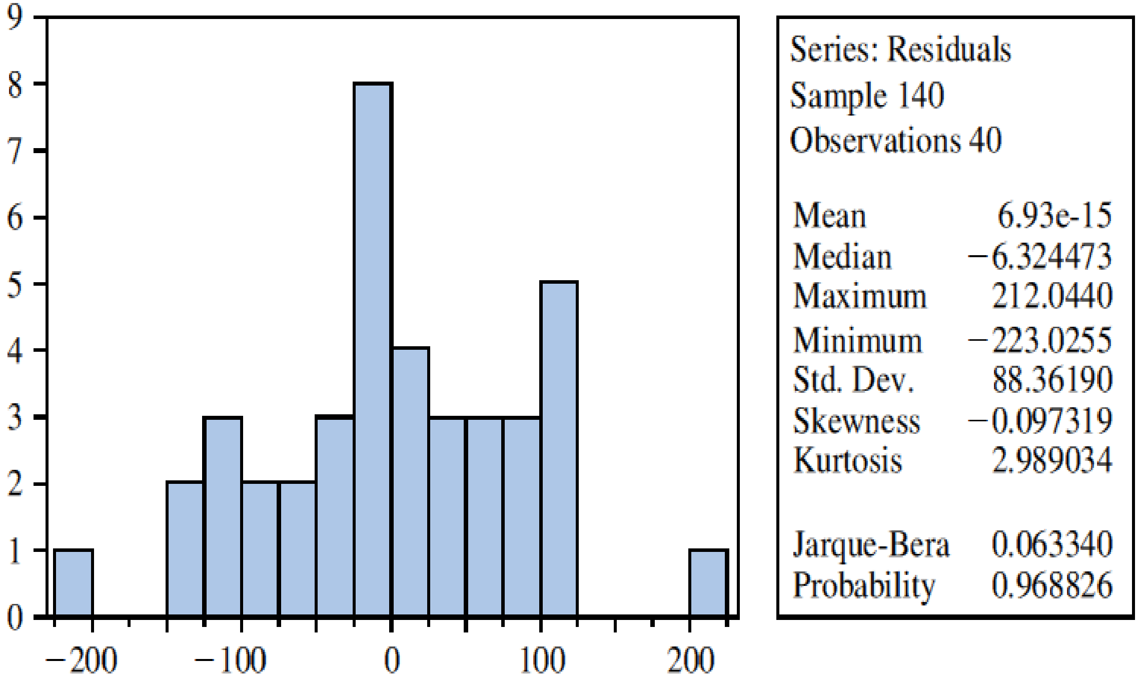


Έλεγχος Κανονικής Κατανομής στα Κατάλοιπα

- Ο έλεγχος υποθέσεων και τα διαστήματα εμπιστοσύνης προϋποθέτουν την υπόθεση ότι τα σφάλματα (και άρα η εξαρτημένη μεταβλητή Y) είναι κανονικά κατανεμημένη. Είναι όμως ;

- Μπορούμε να ελέγξουμε τα κατάλοιπα χρησιμοποιώντας:
  - Το ιστόγραμμα τους
  - Ένα στατιστικό τεστ (για παράδειγμα το Jarque-Bera)

EViews output: residuals histogram and summary statistics for food expenditure



Η μηδενική υπόθεση του JB test είναι ότι τα δεδομένα προέρχονται από την κανονική κατανομή (με ασυμμετρία 0 και κύρτωση ίση με 3. Όπως είδαμε όταν **p-value < (α=5%)**, τότε απορρίπτουμε την H0. Εδώ p-value = 0.9688 >> α = 0.05, άρα δεν απορρίπτουμε την H0 και άρα τα κατάλοιπα κατανέμονται κανονικά.

The Jarque–Bera statistic is given by:

$$JB = \frac{N}{6}\left(S^2 + \frac{(K-3)^2}{4} \right) \sim \chi^2(2)$$

where

N = sample size

S = skewness

K = kurtosis

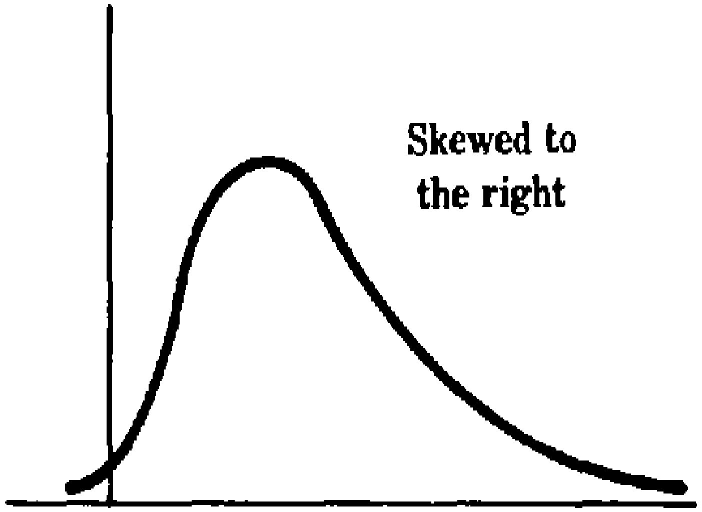


Fig. 3-3

Positive Skewness

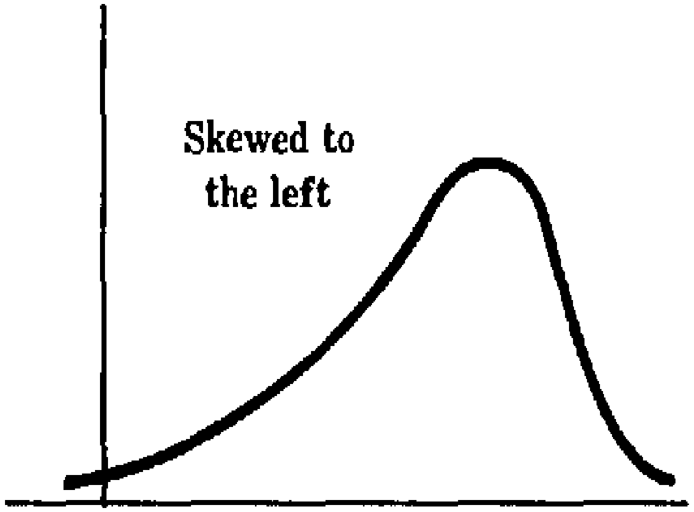


Fig. 3-4

Negative Skewness

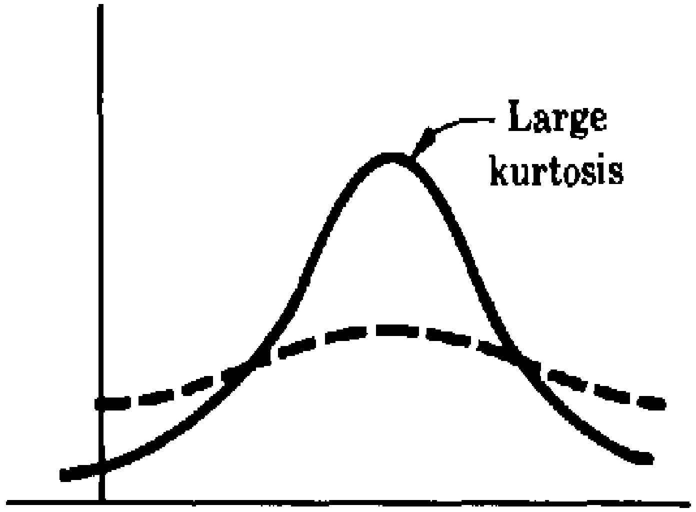


Fig. 3-5

Excess Kurtosis

Log-Linear

Μοντέλα Log-Linear

- Τα υποδείγματα με την εξαρτημένη μεταβλητή (Y) μετασχηματισμένη με τον φυσικό λογάριθμο είναι πολύ κοινά, ιδιαίτερα όπου εμπλέκονται χρηματικές τιμές (monetary values), για παράδειγμα: μισθοί, εισόδημα, τιμές, πωλήσεις, έξοδα, κτλ.
- Αυτές οι μεταβλητές έχουν ως χαρακτηριστικό τους ότι παίρνουν θετικές τιμές και έχουν κατανομές με θετική ασυμμετρία και μία μακριά δεξιά ουρά (ακραίες υψηλές τιμές).
- Για παράδειγμα η σχέση μεταξύ μισθού και εκπαίδευσης είναι ένα θεμελιώδες ερώτημα των οικονομικών της εργασίας.
- Υποθέστε ότι η προστιθέμενη αξία στο μισθό κάποιου ενός προστιθέμενου έτους εκπαίδευσης είναι σταθερή ως r.
- Τότε ισχύει ότι $WAGE_1=(1+r)\ WAGE_0$, για δύο έξτρα χρόνια εκπαίδευσης ισχύει ότι: $WAGE_2=(1+r)^2\ WAGE_0$, κ.ο.κ.
- Τότε ένα μοντέλο θα μπορούσε να είναι:
- $\ln WAGE = \ln WAGE_0 + \ln(1+r) \times EDUC$
- $\ln WAGE = \beta_1 + \beta_2 \times \mathrm{EDUC}$

Μοντέλα Log-Linear: Παράδειγμα

- Εκτιμώντας αυτό το μοντέλο:

$$\ln(WAGE) = 1.6094 + 0.0904 \times EDUC$$
$$(\text{se}) \qquad (0.0864) \quad (0.0061)$$

- Που σημαίνει ότι ένας έξτρα χρόνος εκπαίδευσης αυξάνει το μισθό κατά 9%. Το διάστημα εμπιστοσύνης (95%) για την προστιθέμενη αξία στο μισθό ενός πρόσθετου έτους εκπαίδευσης παίρνει τις τιμές 7.8% έως 10.2%.

- Επειδή σε ένα log-linear μοντέλο παλινδρόμησης το R^2 αναφέρεται στο ln(y) και όχι στο y (επίσης οι προβλέψεις αναφέρονται στο ln(y) και όχι στο y) το λογισμικό κάνει μία σειρά από μετασχηματισμούς στον υπολογισμό αυτών των μεταβλητών.

Log-Log

Μοντέλα Log-Log: Παράδειγμα

- Τα μοντέλα log-log, $\ln(y) = \beta_1 + \beta_2 \ln(x)$, χρησιμοποιούνται ευρέως για να περιγράψουν καμπύλες ζήτησης και συναρτήσεις παραγωγής, μεταξύ άλλων.
- Για να ορίζονται οι λογάριθμοι θα πρέπει οι X και Y να είναι αποκλειστικά θετικές.
- Οι εκτιμημένοι συντελεστές είναι ελαστικότητες.
- Ένα κλασσικό παράδειγμα εδώ είναι η κατανάλωση και τιμή κοτόπουλου, όπου:

Q: per capita consumption of chicken

P: real price of chicken

Μοντέλα Log-Log: Παράδειγμα

- Το εκτιμημένο μοντέλο σε αυτή την περίπτωση είναι:

$$\ln(Q) = 3.717 - 1.121 \times \ln(P) \qquad R_g^2 = 0.8817$$
$$(\text{se}) \quad (0.022) \quad (0.049)$$

- Με αυτόν τον τρόπο έχουμε εκτιμήσει την ελαστικότητα ζήτησης για κοτόπουλο ως 1.121.
- Δηλαδή, για κάθε 1% αύξηση στην πραγματική τιμή του κοτόπουλου, η κατανάλωση του μειώνεται κατά 1.121%.