ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
AGRICULTURAL UNIVERSITY OF ATHENS

ΔΠΜΣ «ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ ΔΙΟΙΚΗΣΗΣ»

Μάθημα: Ποσοτικές Μέθοδοι

Διδάσκουσα: Επίκουρη Καθηγήτρια Φωτεινή Κυριαζή

4η Εκπαιδευτική Συνάντηση:

Εισαγωγή στην Οικονομετρία, Ανάλυση Παλινδρόμησης , Διμεταβλητή Παλινδρόμηση, Ανάλυση Συσχέτισης

Αθήνα, 2024

Ι.1 Τι Είναι η Οικονομετρία;

- Η κυριολεκτική ερμηνεία της λέξης, *οικονομετρία* είναι «οικονομική μέτρηση»
- Παρά το γεγονός ότι η μέτρηση αποτελεί ένα σημαντικό μέρος της οικονομετρίας, το πλαίσιο της οικονομετρίας είναι πολύ ευρύτερο.
- Η οικονομετρία είναι ένα αμάλγαμα οικονομικής θεωρίας, οικονομικών μαθηματικών, στατιστικής.
- Ωστόσο, το γνωστικό αυτό αντικείμενο αξίζει να μελετηθεί μεμονωμένα από τα παραπάνω.

Ι.2 Γιατί ένα Ξεχωριστό Γνωστικό Αντικείμενο;

- Η οικονομική θεωρία διατυπώνει δηλώσεις ή υποθέσεις που είναι ως επί το πλείστον ποιοτικής φύσης.
- Το κύριο ενδιαφέρον των οικονομικών μαθηματικών είναι η έκφραση της οικονομικής θεωρίας σε μαθηματική μορφή (εξισώσεις) χωρίς να λαμβάνεται υπόψη η μετρησιμότητα ή η εμπειρική επαλήθευση της θεωρίας.
- Η οικονομική στατιστική ασχολείται κατά κύριο λόγο με τη συλλογή, την επεξεργασία και την παρουσίαση οικονομικών στοιχείων με τη μορφή διαγραμμάτων και πινάκων.
- Παρόλο που η μαθηματική στατιστική παρέχει πολλά εργαλεία, ο οικονομέτρης χρειάζεται συχνά ειδικές μεθόδους λόγω της ιδιαίτερης φύσης της πλειοψηφίας των οικονομικών στοιχείων.

Ι.2 Γιατί ένα Ξεχωριστό Γνωστικό Αντικείμενο;

- Στην οικονομετρία ο ερευνητής έρχεται συχνά αντιμέτωπος με στοιχεία που είναι το αποτέλεσμα **παρατήρησης** και όχι **πειράματος**.
- Δύο σημαντικές συνέπειες για την εμπειρική υποδειγματοποίηση στην οικονομετρία.
  1. Ο ερευνητής απαιτείται να κατέχει πολύ διαφορετικές δεξιότητες από αυτές που χρειάζονται για την ανάλυση πειραματικών στοιχείων
  2. Ο διαχωρισμός μεταξύ αυτού που συνέλλεξε τα στοιχεία και αυτού που θα τα αναλύσει απαιτεί ο ερευνητής να εξοικειωθεί σημαντικά με τη φύση και τη δομή των υπό εξέταση στοιχείων.

I.3 Μεθοδολογία της Οικονομετρίας

- Η **παραδοσιακή** ή **κλασική** οικονομετρική μεθοδολογία ακολουθεί τα ακόλουθα στάδια:
  1. Δήλωση της θεωρίας ή της υπόθεσης.
  2. Προσδιορισμός του μαθηματικού υποδείγματος της θεωρίας.
  3. Προσδιορισμός του στατιστικού, ή οικονομετρικού υποδείγματος.
  4. Συλλογή των στοιχείων.
  5. Εκτίμηση των παραμέτρων του οικονομετρικού υποδείγματος.
  6. Έλεγχοι υποθέσεων.
  7. Πρόβλεψη.
  8. Χρήση του υποδείγματος για σκοπούς ελέγχου ή πολιτικής.

Ι.3 Μεθοδολογία της Οικονομετρίας

- Ως παράδειγμα των προηγούμενων σταδίων, ας εξετάσουμε τη γνωστή Κεϋνσιανή θεωρία της κατανάλωσης.

1. **Δήλωση της Θεωρίας ή της Υπόθεσης**

   Ο Keynes διατύπωσε την άποψη ότι η **οριακή ροπή προς κατανάλωση** (Marginal Propensity to Consume - MPC), δηλαδή ο ρυθμός μεταβολής της κατανάλωσης για μία μεταβολή του εισοδήματος κατά μία μονάδα (π.χ., ένα ευρώ), είναι μεγαλύτερη από το μηδέν αλλά μικρότερη από το 1.

I.3 Μεθοδολογία της Οικονομετρίας

2. Προσδιορισμός του Μαθηματικού Υποδείγματος της Κατανάλωσης

Παρά το γεγονός ότι ο Keynes αξίωνε μία θετική σχέση μεταξύ της κατανάλωσης και του εισοδήματος, δεν προσδιόρισε την ακριβή μορφή της συναρτησιακής σχέσης μεταξύ των δύο.

Έχουμε την ακόλουθη μορφή για την Κεϋνσιανή συνάρτηση κατανάλωσης:

$$Y = \beta_1 + \beta_2 X \quad 0 < \beta_2 < 1 \text{ (I.3.1)}$$

όπου

Y = καταναλωτικές δαπάνες

X = εισόδημα

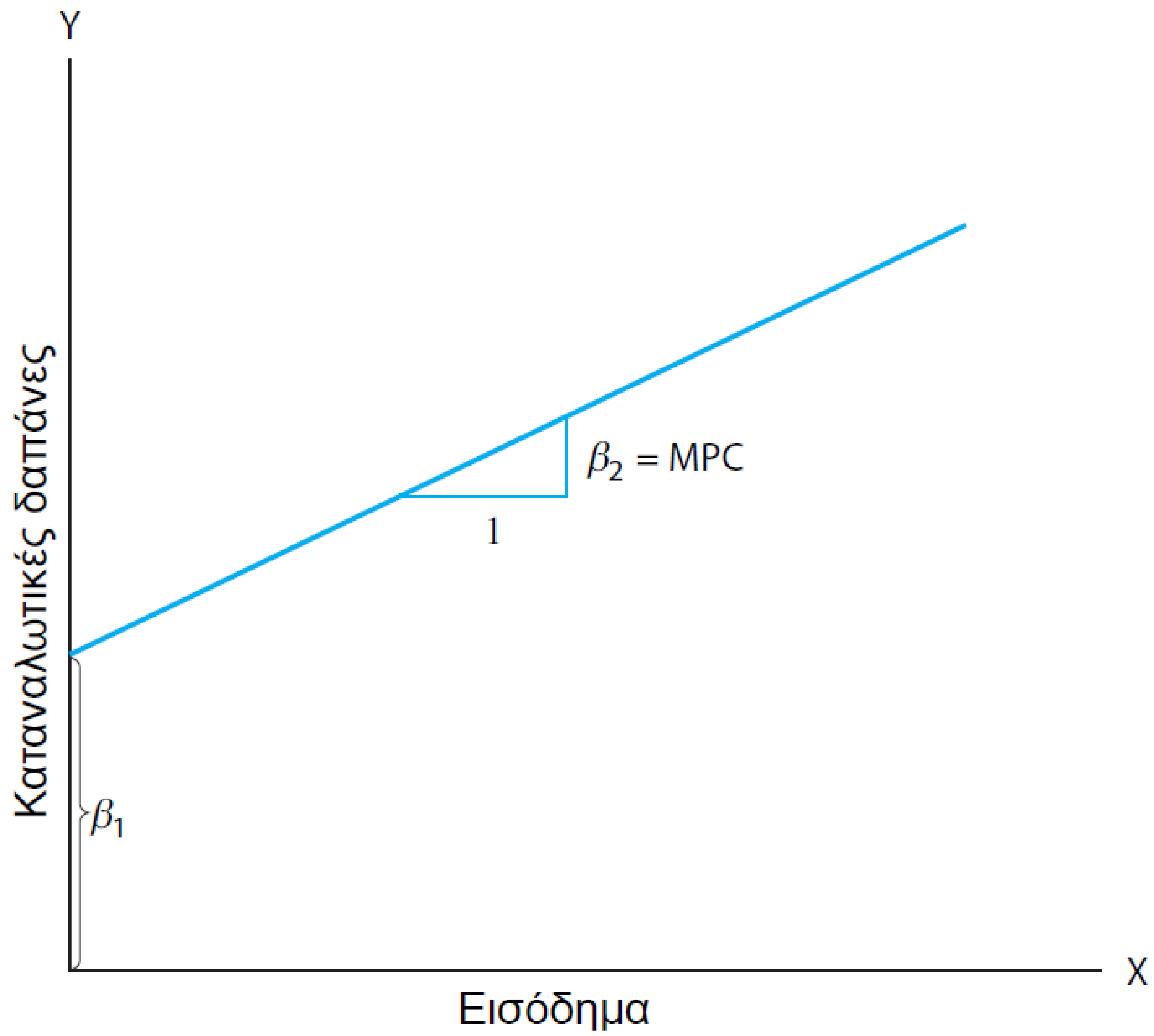
β_1 και β_2 = οι **παράμετροι** του υποδείγματος. Συντελεστές **σταθεράς** (intercept) και **κλίσης** (slope).

I.3 Μεθοδολογία της Οικονομετρίας

2. Προσδιορισμός του Μαθηματικού Υποδείγματος της Κατανάλωσης

- Η εξίσωση αυτή ονομάζεται **συνάρτηση κατανάλωσης** (consumption function).
- Ένα υπόδειγμα είναι απλά ένα σύνολο μαθηματικών εξισώσεων.
- Αν το υπόδειγμα έχει μόνο μία εξίσωση, ονομάζεται **υπόδειγμα μίας εξίσωσης** (single-equation model).
- Αν το υπόδειγμα έχει περισσότερες από μία εξίσωση, είναι γνωστό ως **υπόδειγμα πολλαπλών εξισώσεων** (multiple equation model).
- Η μεταβλητή που εμφανίζεται στην αριστερή πλευρά του συμβόλου της ισότητας ονομάζεται **εξαρτημένη μεταβλητή** (dependent variable)
- Η μεταβλητή που βρίσκεται στη δεξιά πλευρά ονομάζεται **ανεξάρτητη** (independent variable), ή **ερμηνευτική** (explanatory) μεταβλητή.

ΔΙΑΓΡΑΜΜΑ Ι.1 Η Κεϋνσιανή συνάρτηση κατανάλωσης



I.3 Μεθοδολογία της Οικονομετρίας

3. Προσδιορισμός του Οικονομετρικού Υποδείγματος της Κατανάλωσης

- Για να λάβει υπόψη του τις μη-ακριβείς σχέσεις μεταξύ των οικονομικών μεταβλητών, ο οικονομέτρης θα τροποποιήσει την *προσδιοριστική* συνάρτηση κατανάλωσης της Εξ. (I.3.1), ως εξής:

$$Y = \beta_1 + \beta_2 X + u \quad \text{(I.3.2)}$$

- όπου
- u = **διαταρακτικός όρος** (disturbance term), ή **όρος σφάλματος** (error term), είναι μία **τυχαία (στοχαστική) μεταβλητή** (random stochastic variable) που έχει καλά καθορισμένες πιθανολογικές ιδιότητες.
- Ο διαταρακτικός όρος u μπορεί να αντιπροσωπεύει όλους εκείνους τους παράγοντες που επηρεάζουν την κατανάλωση, οι οποίοι όμως δε λαμβάνονται ρητά υπόψη.

I.3 Μεθοδολογία της Οικονομετρίας

3. Προσδιορισμός του Οικονομετρικού Υποδείγματος της Κατανάλωσης

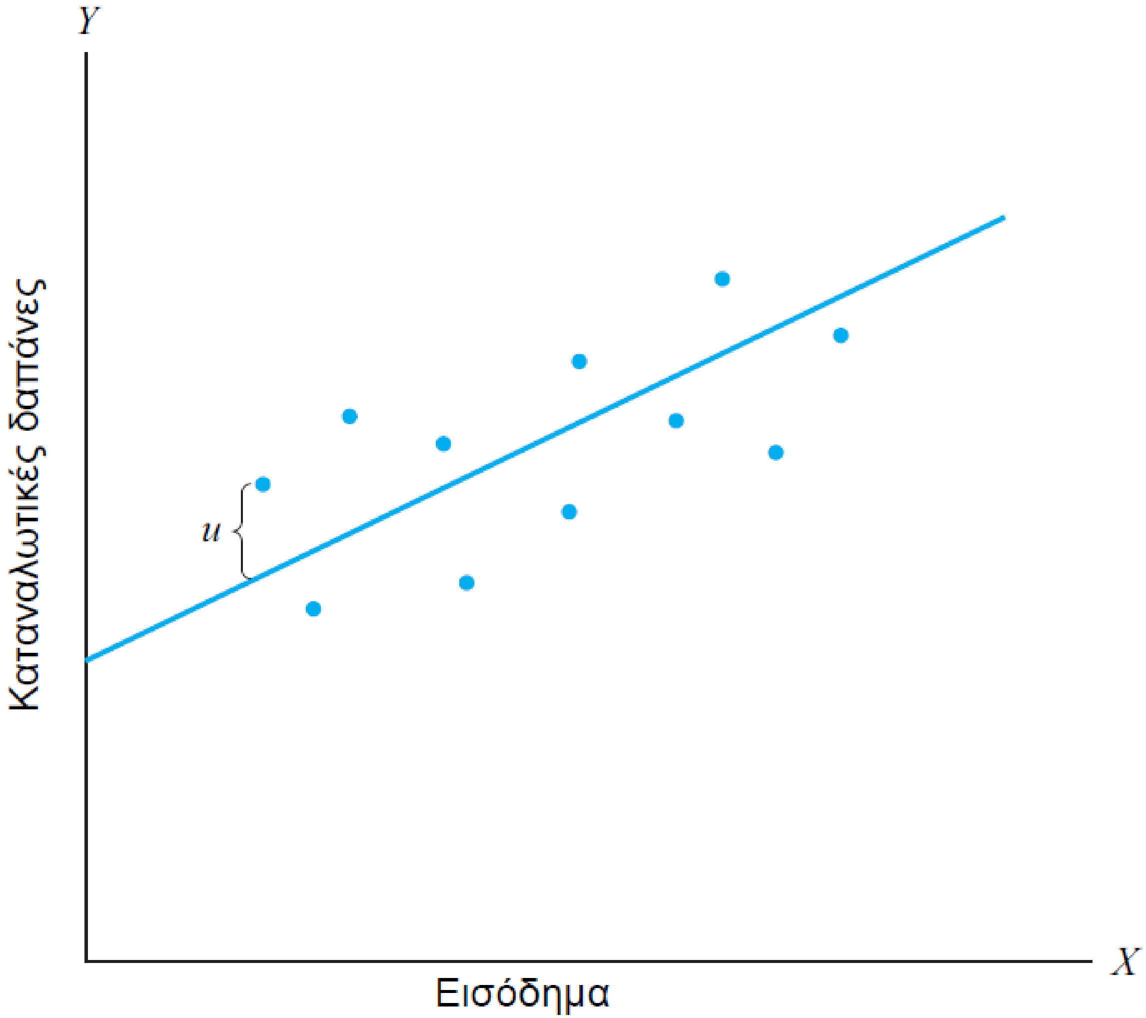
- Η Εξίσωση (I.3.2) είναι ένα παράδειγμα ενός **οικονομετρικού υποδείγματος** (econometric model).
- Είναι ένα παράδειγμα ενός **γραμμικού υποδείγματος παλινδρόμησης** (linear regression model).
- Η οικονομετρική συνάρτηση κατανάλωσης υποθέτει ότι η εξαρτημένη μεταβλητή *Y* (κατανάλωση) σχετίζεται γραμμικά με την ερμηνευτική μεταβλητή *X* (εισόδημα), όμως η σχέση μεταξύ των δύο δεν είναι ακριβής˙ **υπόκειται σε μεμονωμένη μεταβολή**.

I.3 Μεθοδολογία της Οικονομετρίας

4. Συλλογή των Στοιχείων

- Για να μπορέσουμε να εκτιμήσουμε το οικονομετρικό υπόδειγμα που παρουσιάζεται στην Εξ. (I.3.2), δηλαδή, για να αποκτήσουμε τις αριθμητικές τιμές των β_1 και β_2, χρειαζόμαστε στοιχεία (δεδομένα).
- Η μεταβλητή *Y* είναι η *συνολική* (για την οικονομία στο σύνολό της) προσωπική καταναλωτική δαπάνη (ΠΚΔ).
- Η μεταβλητή *X* είναι το ακαθάριστο εγχώριο προϊόν (ΑΕΠ), ένα μέτρο του συνολικού εισοδήματος.
- Και οι δύο μετρώνται σε δισεκατομμύρια δολάρια σε τιμές του 2000.

ΔΙΑΓΡΑΜΜΑ Ι.2 Οικονομετρικό υπόδειγμα της Κεϋνσιανής συνάρτησης κατανάλωσης



ΠΙΝΑΚΑΣ Ι.1

Y (Προσωπική Καταναλωτική Δαπάνη (Personal Consumption Expenditure – PCE) και *X* (Ακαθάριστο Εγχώριο Προϊόν (Gross Domestic Product – GDP, 1960-2005),

Σημείωση: Δισεκατομμύρια Δολάρια (σε τιμές του 2000).

Έτος	PCE(*Y*)	GDP(*X*)
1960	1597,4	2501,8
1961	1630,3	2560,0
1962	1711,1	2715,2
1963	1781,6	2834,0
1964	1888,4	2998,6
1965	2007,7	3191,1
1966	2121,8	3399,1
1967	2185,0	3484,6
1968	2310,5	3652,7
1969	2396,4	3765,4
1970	2451,9	3771,9
1971	2545,5	3898,6
1972	2701,3	4105,0
1973	2833,8	4341,5
1974	2812,3	4319,6
1975	2876,9	4311,2
1976	3035,5	4540,9
1977	3164,1	4750,5
1978	3303,1	5015,0
1979	3383,4	5173,4
1980	3374,1	5161,7
1981	3422,2	5291,7
1982	3470,3	5189,3
1983	3668,6	5423,8
1984	3863,3	5813,6
1985	4064,0	6053,7
1986	4228,9	6263,6
1987	4369,8	6475,1
1988	4546,9	6742,7
1989	4675,0	6981,4
1990	4770,3	7112,5
1991	4778,4	7100,5
1992	4934,8	7336,6
1993	5099,8	7532,7
1994	5290,7	7835,5
1995	5433,5	8031,7
1996	5619,4	8328,9
1997	5831,8	8703,5
1998	6125,8	9066,9
1999	6438,6	9470,3
2000	6739,4	9817,0
2001	6910,4	9890,7
2002	7099,3	10048,8
2003	7295,3	10301,0
2004	7577,1	10703,5
2005	7841,2	11048,6

I.3 Μεθοδολογία της Οικονομετρίας

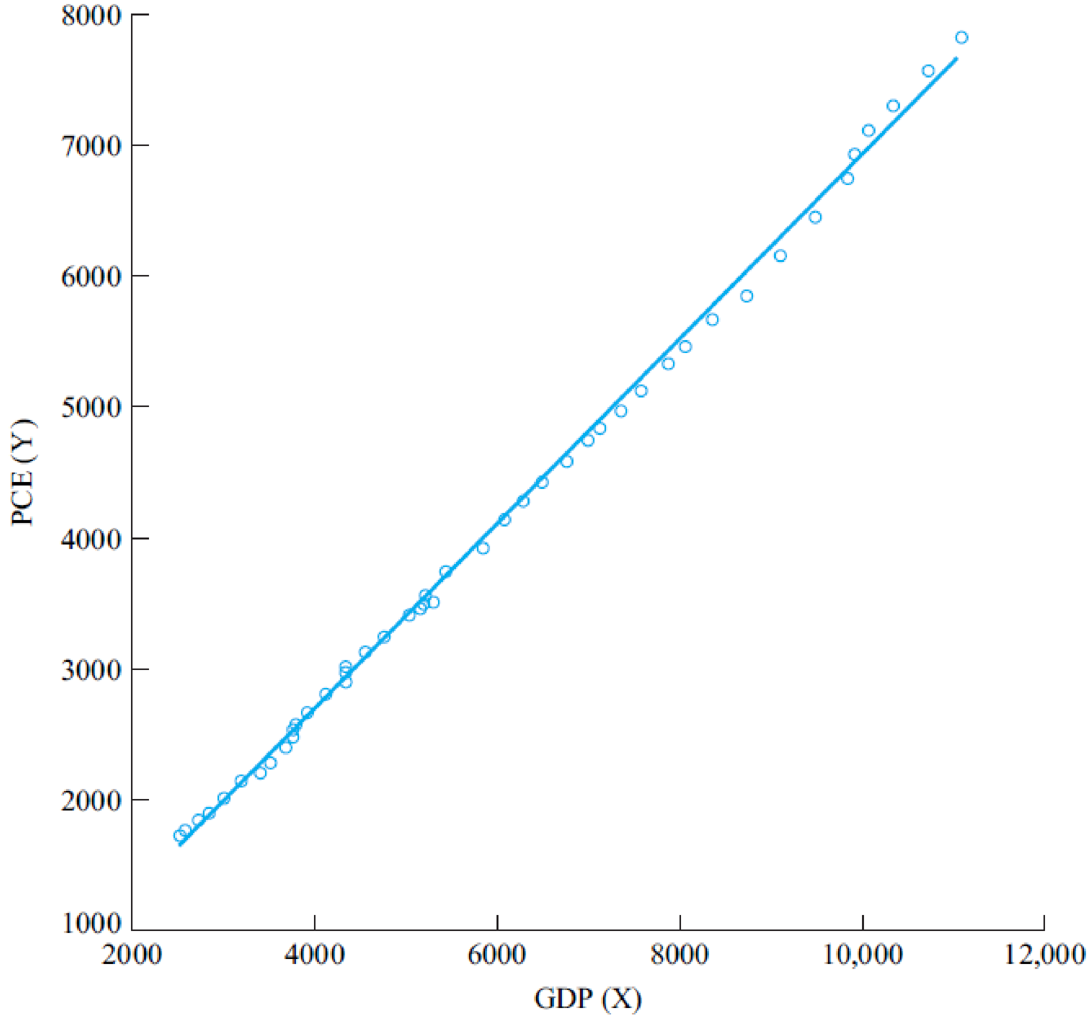
5. Εκτίμηση του Οικονομετρικού Υποδείγματος

- Το επόμενο βήμα είναι να εκτιμήσουμε τις παραμέτρους της συνάρτησης κατανάλωσης.
- Η στατιστική τεχνική της **ανάλυσης παλινδρόμησης** (regression analysis) είναι το βασικό εργαλείο που χρησιμοποιείται προκειμένου να λάβουμε τις εκτιμήσεις.
- η εκτιμημένη συνάρτηση κατανάλωσης είναι:

$$\hat{Y}_t = -299{,}5913 + 0{,}7218X_t \quad \text{(I.3.3)}$$

- Το καπέλο στη *Y* δείχνει ότι πρόκειται για μία εκτίμηση.
- Η εκτιμημένη συνάρτηση κατανάλωσης (η γραμμή παλινδρόμησης) παρουσιάζεται στο Διάγραμμα I.3.

ΔΙΑΓΡΑΜΜΑ Ι.3 Προσωπική καταναλωτική δαπάνη (*Y*) σε σχέση με το ΑΕΠ (*X*), 1960-2005, σε δισεκατομμύρια δολάρια (σε τιμές του 2000).



Ι.3 Μεθοδολογία της Οικονομετρίας

6. Έλεγχοι Υποθέσεων

- Η επιβεβαίωση ή διάψευση των οικονομικών θεωριών, βάσει στοιχείων του δείγματος στηρίζεται σε ένα κλάδο της στατιστικής θεωρίας που είναι γνωστός ως **συμπερασματική ή επαγωγική στατιστική** (statistical inference) (**έλεγχος υποθέσεων –** hypothesis testing).

7. Πρόβλεψη

- Εάν το υπόδειγμα που έχει επιλεχθεί δεν αντικρούει την υπόθεση ή τη θεωρία που είναι υπό εξέταση, μπορούμε να το χρησιμοποιήσουμε για να προβλέψουμε τη μελλοντική τιμή(ες) της εξαρτημένης μεταβλητής ή **προγνωστικής μεταβλητής** (forecast variable) Y βάσει της γνωστής ή προσδοκώμενης μελλοντικής τιμής (ών) της ερμηνευτικής, ή **μεταβλητής πρόβλεψης** (predictor variable) X.

I.3 Μεθοδολογία της Οικονομετρίας

7. Πρόβλεψη

- Ας υποθέσουμε ότι θέλουμε να προβλέψουμε τη μέση καταναλωτική δαπάνη για το 2006.
- Το ΑΕΠ για το 2006 ήταν 11319,4 δισεκατομμύρια δολάρια. Αντικαθιστώντας την τιμή αυτή του ΑΕΠ στο δεξί μέλος της Εξ. (I.3.3), έχουμε:

$$\hat{Y}_{2006} = -299{,}5913 + 0{,}7218(11319{,}4) = 7870{,}7516 \quad (I.3.4)$$

- Η πραγματική τιμή της καταναλωτικής δαπάνης που καταγράφηκε το 2006 ήταν 8044 δισεκατομμύρια δολάρια.
- Επομένως το εκτιμημένο υπόδειγμα της Εξ. (I.3.3) **υποεκτίμησε** (underpredicted) την πραγματική καταναλωτική δαπάνη κατά περίπου 174 δισεκατομμύρια δολάρια.
- Θα μπορούσαμε να πούμε ότι το **σφάλμα πρόβλεψης** (forecast error) είναι περίπου 174 δισεκατομμύρια δολάρια, που είναι περίπου 1,5 τοις εκατό του πραγματικού ΑΕΠ για το 2006.

I.3 Μεθοδολογία της Οικονομετρίας

8. Χρήση του Υποδείγματος για Σκοπούς Ελέγχου ή Πολιτικής

- Ας υποθέσουμε ότι έχουμε εκτιμήσει τη συνάρτηση κατανάλωσης που παρουσιάζεται στην Εξ. (I.3.3).
- Έστω πως η κυβέρνηση πιστεύει ότι οι καταναλωτικές δαπάνες των 8750 (δισεκατομμυρίων δολαρίων σε τιμές του 2000) θα διατηρήσουν το ποσοστό ανεργίας στο τρέχον επίπεδο της τάξης του 4,2 τοις εκατό (αρχές 2006).
- Ποιο είναι το επίπεδο του εισοδήματος που εγγυάται το ποσό-στόχο των καταναλωτικών δαπανών;

$$8750 = -299{,}5913 + 0{,}7218(GDP_{2006}) \quad \text{(I.3.6)}$$

- Οπότε X = 12537.
- Δηλαδή, δεδομένης μίας MPC της τάξης του 0,72, ένα επίπεδο εισοδήματος ύψους 12537 (δισεκατομμυρίων) δολαρίων θα οδηγήσει σε δαπάνες ύψους περίπου 8750 δισεκατομμυρίων δολαρίων.

I.3 Μεθοδολογία της Οικονομετρίας

8. Χρήση του Υποδείγματος για Σκοπούς Ελέγχου ή Πολιτικής

- Ένα εκτιμημένο υπόδειγμα μπορεί να χρησιμοποιηθεί για σκοπούς ελέγχου, ή για σκοπούς πολιτικής.
- Με το κατάλληλο μείγμα δημοσιονομικής και νομισματικής πολιτικής, η κυβέρνηση μπορεί να χειραγωγήσει τη **μεταβλητή ελέγχου** (control variable) *X* για την παραγωγή του επιθυμητού επιπέδου της **μεταβλητής στόχου** (target variable) *Y*.

Ι.4 Κλάδοι της Οικονομετρίας

- Η οικονομετρία μπορεί να χωριστεί σε δύο μεγάλες κατηγορίες:
  1. **θεωρητική οικονομετρία** (theoretical econometrics)
  2. **εφαρμοσμένη οικονομετρία** (applied econometrics).
- Σε κάθε κατηγορία, κάποιος μπορεί να προσεγγίσει το θέμα μέσω της κλασικής προσέγγισης ή της προσέγγισης του Bayes.
- Η θεωρητική οικονομετρία ασχολείται με την ανάπτυξη των κατάλληλων μεθόδων για τη μέτρηση των οικονομικών σχέσεων που καθορίζονται από τα οικονομετρικά υποδείγματα.
- Μία από τις μεθόδους που χρησιμοποιείται εκτενώς είναι η μέθοδος των **ελαχίστων τετραγώνων** (least squares).
- Στην εφαρμοσμένη οικονομετρία χρησιμοποιούμε τα εργαλεία της θεωρητικής οικονομετρίας για να μελετήσουμε κάποιο ειδικό πεδίο(α) της οικονομίας και των επιχειρήσεων, όπως η *συνάρτηση παραγωγής*, η *συνάρτηση επενδύσεων*, οι *συναρτήσεις ζήτησης και προσφοράς*, η *θεωρία χαρτοφυλακίου*, κ.λπ.

Η Φύση της Ανάλυσης Παλινδρόμησης

1.1 Ιστορική Προέλευση του Όρου *Παλινδρόμηση*

- Ο όρος παλινδρόμηση εισήχθη από τον Francis Galton.
- Διαπίστωσε ότι, αν και υπήρχε μία τάση οι ψηλοί γονείς να έχουν ψηλά παιδιά και οι κοντοί γονείς να έχουν κοντά παιδιά, το μέσο ύψος των παιδιών που γεννιόνταν από γονείς ενός ορισμένου ύψους έτεινε να μετακινείται ή να «παλινδρομεί» προς το μέσο ύψος του συνολικού πληθυσμού.
- Με άλλα λόγια, το ύψος των παιδιών των ασυνήθιστα ψηλών ή ασυνήθιστα κοντών γονέων τείνει να κινείται προς το μέσο ύψος του πληθυσμού.
- Ο *νόμος της καθολικής παλινδρόμησης* του Galton επιβεβαιώθηκε από τον φίλο του Karl Pearson, ο οποίος συνέλεξε περισσότερες από χίλιες εγγραφές αναφορικά με τα ύψη μελών διαφόρων οικογενειών.
- Βρήκε ότι το μέσο ύψος των γιων μίας ομάδας ψηλών πατέρων ήταν χαμηλότερο από το ύψος των πατέρων και το μέσο ύψος των γιων μίας ομάδας κοντών πατέρων ήταν υψηλότερο από το ύψος των πατέρων τους, με αποτέλεσμα οι ψηλοί και οι κοντοί γιοι να «παλινδρομούν» προς το μέσο ύψος όλων των ανδρών.

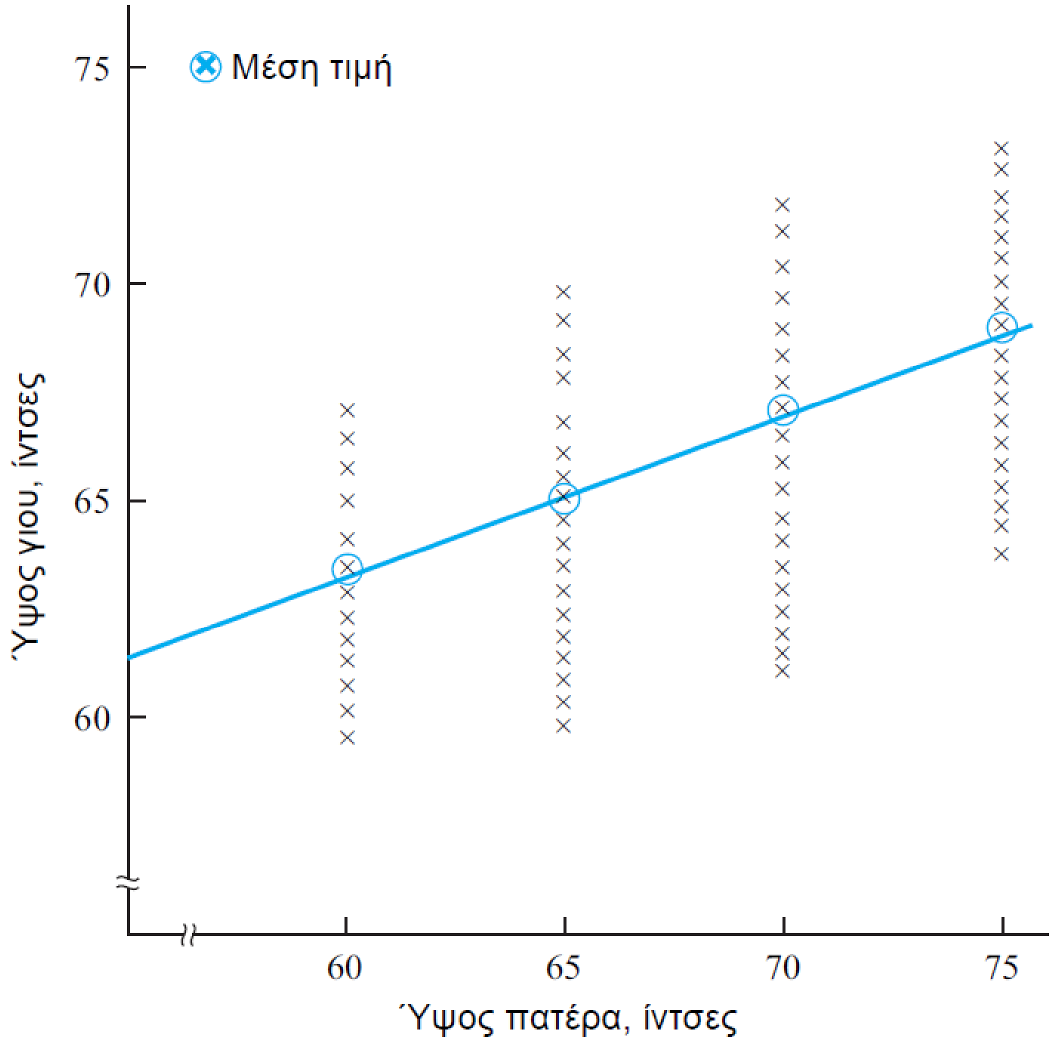
1.2 Η Σύγχρονη Ερμηνεία της Παλινδρόμησης

- Η ανάλυση παλινδρόμησης ασχολείται με τη μελέτη της εξάρτησης μίας μεταβλητής, της *εξαρτημένης μεταβλητής*, από μία ή από περισσότερες άλλες μεταβλητές, τις *ερμηνευτικές μεταβλητές*.
- Σκοπός είναι η εκτίμηση ή/και η πρόβλεψη της τιμής του (πληθυσμιακού) μέσου όρου σε όρους των γνωστών ή σταθερών τιμών (σε επαναλαμβανόμενη δειγματοληψία) των τελευταίων.

1.2 Η Σύγχρονη Ερμηνεία της Παλινδρόμησης

- **Παραδείγματα**
- Το **διάγραμμα διασποράς** δείχνει την κατανομή των υψών των γιων σε έναν υποθετικό πληθυσμό που αντιστοιχούν σε συγκεκριμένες ή *σταθερές* τιμές ύψους του πατέρα.
- Σε κάθε δεδομένο ύψος του πατέρα αντιστοιχεί μία *σειρά* ή κατανομή υψών των γιων.
- Παρά τη μεταβλητότητα του ύψους των γιων για μία δεδομένη τιμή του ύψους του πατέρα, ο μέσος όρος του ύψους των γιων αυξάνει γενικά καθώς αυξάνει το ύψος του πατέρα.
- Οι σταυροί που περιέχονται σε κύκλο στο Διάγραμμα δείχνουν το *μέσο* ύψος των γιων που αντιστοιχεί σε ένα συγκεκριμένο ύψος του πατέρα.
- Η ένωση αυτών μας δίνει τη **γραμμή παλινδρόμησης** (regression line).
- Η γραμμή παλινδρόμησης δείχνει πώς το *μέσο* ύψος των γιων αυξάνεται καθώς αυξάνεται το ύψος του πατέρα.

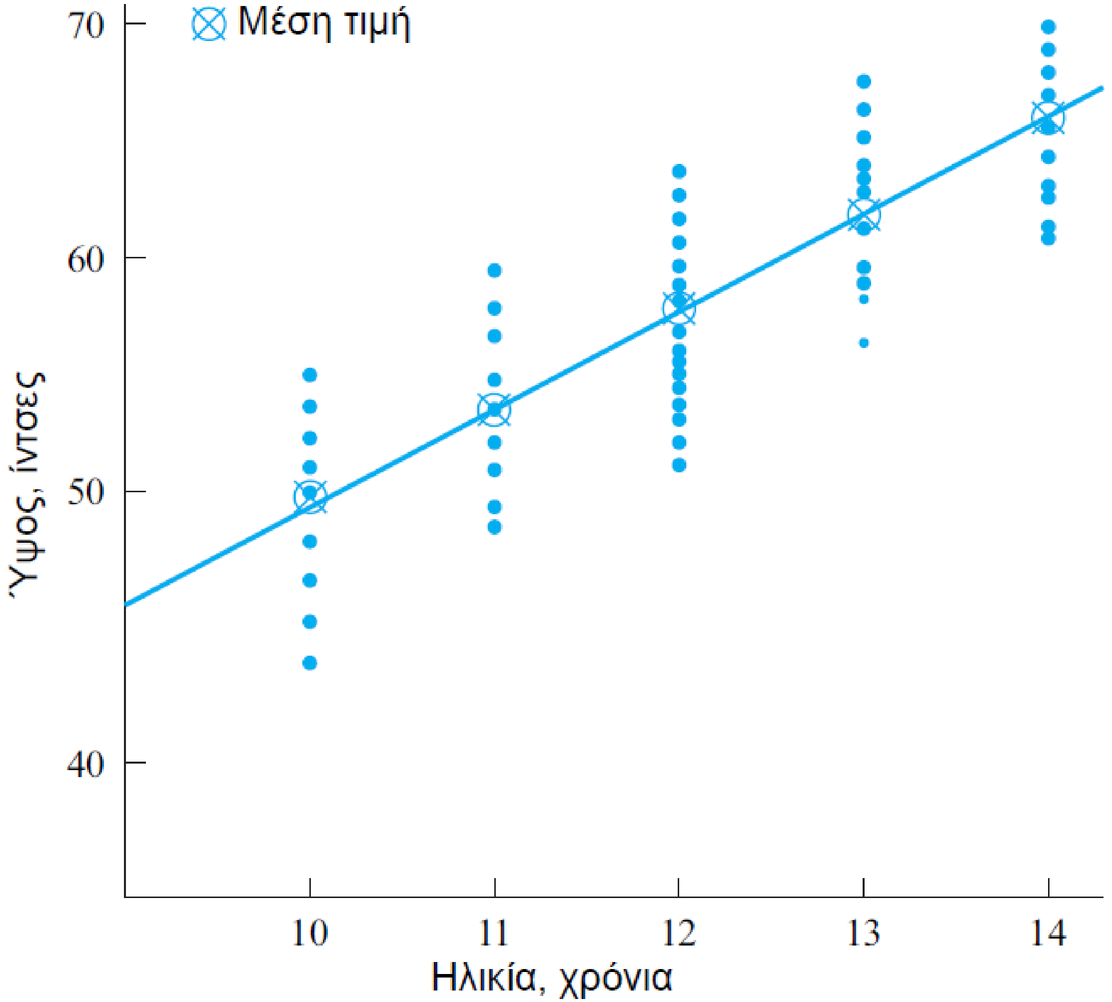
ΔΙΑΓΡΑΜΜΑ 1.1 Υποθετική κατανομή υψών γιών που αντιστοιχούν σε συγκεκριμένα ύψη πατέρων



1.2 Η Σύγχρονη Ερμηνεία της Παλινδρόμησης

- **Παραδείγματα**
- Το διάγραμμα διασποράς 1.1 παρουσιάζει την κατανομή του ύψους των αγοριών, που μετράται σε *σταθερές* ηλικίες, ενός υποθετικού πληθυσμού.
- Σε κάθε ηλικία, έχουμε μία σειρά, ή κατανομή, υψών.
- Το ύψος, *κατά μέσο όρο*, αυξάνει με την ηλικία (φυσικά, μέχρι μία ορισμένη ηλικία), κάτι που γίνεται προφανές, αν σχεδιάσουμε μία γραμμή (τη γραμμή παλινδρόμησης).
- Η γραμμή αυτή διατρέχει τα σημεία που περιέχονται στους κύκλους και τα οποία αντιπροσωπεύουν το μέσο ύψος στις συγκεκριμένες ηλικίες.

ΔΙΑΓΡΑΜΜΑ 1.2 Υποθετική κατανομή υψών γιών που αντιστοιχούν σε συγκεκριμένες ηλικίες



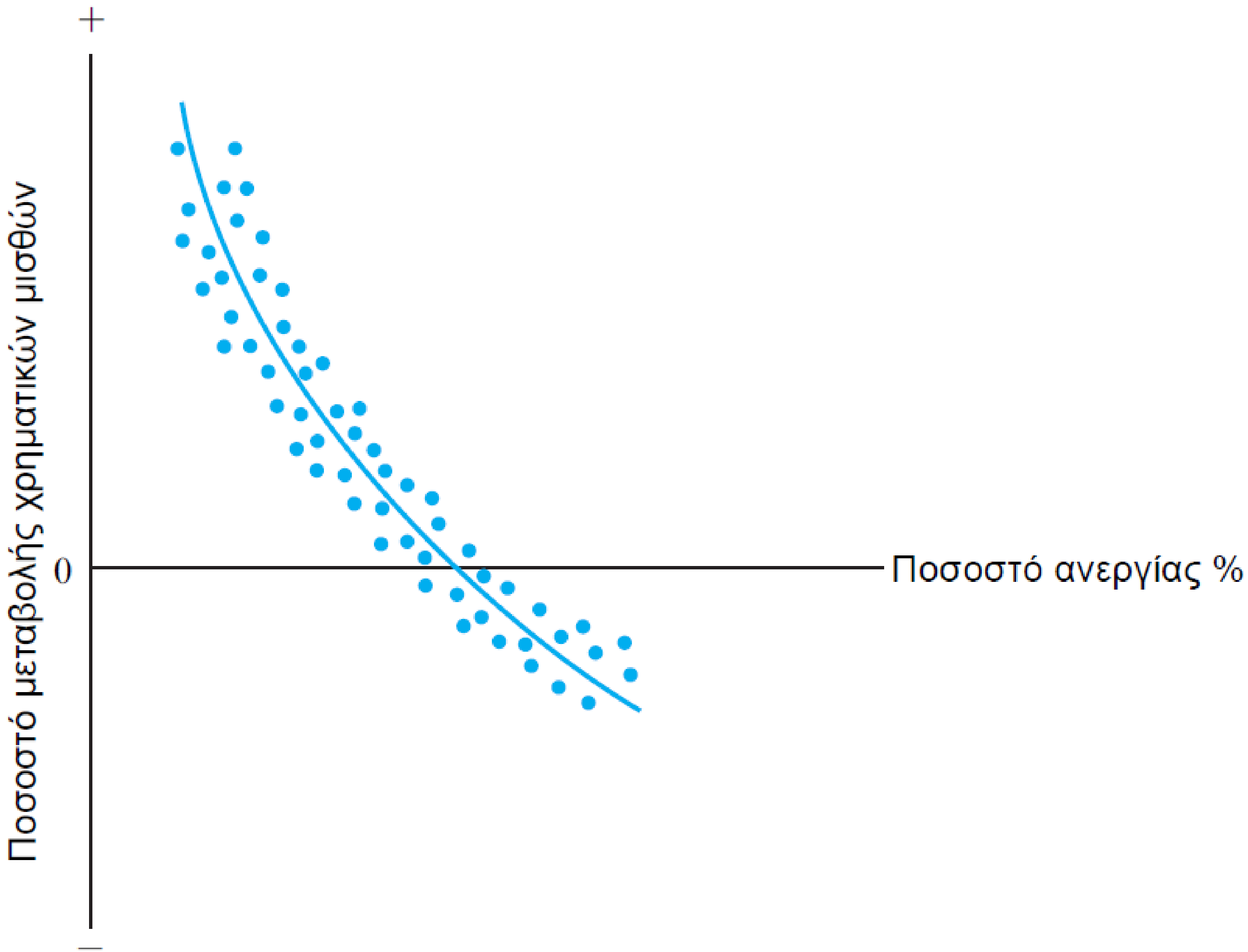
1.2 Η Σύγχρονη Ερμηνεία της Παλινδρόμησης

- **Παραδείγματα**
- Οικονομικά παραδείγματα.
- Ένας οικονομολόγος μπορεί να ενδιαφέρεται για τη μελέτη της εξάρτησης των προσωπικών καταναλωτικών δαπανών από το μετά φόρων ή διαθέσιμο πραγματικό προσωπικό εισόδημα.
- Εκτίμηση της οριακής ροπής προς κατανάλωση (MPC) για μία, π.χ., μεταβολή του πραγματικού εισοδήματος κατά ένα δολάριο.
- **Παραδείγματα**
- **Οικονομικά παραδείγματα.**
- Ένας μονοπωλητής που μπορεί να καθορίσει την τιμή ή την εκροή (αλλά όχι και τα δύο) μπορεί να θέλει να μάθει την ανταπόκριση της ζήτησης για ένα προϊόν στις μεταβολές των τιμών.
- Ένα τέτοιο πείραμα εκτιμά την **ελαστικότητα ζήτησης ως προς την τιμή** (price elasticity).

1.2 Η Σύγχρονη Ερμηνεία της Παλινδρόμησης

- **Παραδείγματα**
- **Οικονομικά παραδείγματα.**
- Ρυθμός μεταβολής των μισθών σε σχέση με το ποσοστό ανεργίας.
- Τα ιστορικά στοιχεία παρουσιάζονται στο διάγραμμα διασποράς (Διάγραμμα 1.3).
- Η καμπύλη στο Διάγραμμα 1.3 είναι ένα παράδειγμα της *καμπύλης Phillips* η οποία συσχετίζει τις μεταβολές των μισθών με το ποσοστό ανεργίας.

ΔΙΑΓΡΑΜΜΑ 1.3 Υποθετική καμπύλη Phillips



1.2 Η Σύγχρονη Ερμηνεία της Παλινδρόμησης

- **Παραδείγματα**
- **Οικονομικά παραδείγματα.**
- Γνωρίζουμε από τη νομισματική θεωρία ότι, με όλους τους άλλους παράγοντες σταθερούς, όσο υψηλότερο είναι το ποσοστό πληθωρισμού π, τόσο χαμηλότερο είναι το ποσοστό k του εισοδήματος που οι άνθρωποι θα θέλουν να διατηρήσουν με τη μορφή χρήματος.
- Η κλίση αυτής της γραμμής αντιπροσωπεύει τη μεταβολή στο k δοθείσης μίας μεταβολής του ποσοστού πληθωρισμού.

1.2 Η Σύγχρονη Ερμηνεία της Παλινδρόμησης

- **Παραδείγματα**
- **Οικονομικά παραδείγματα.**
- Ο διευθυντής μάρκετινγκ μίας εταιρείας μπορεί να θέλει να γνωρίζει πως σχετίζεται η ζήτηση για το προϊόν της εταιρείας με, π.χ., τις διαφημιστικές δαπάνες.
- **Ελαστικότητα της ζήτησης** (elasticity of demand) σε σχέση με τις διαφημιστικές δαπάνες, δηλαδή, την ποσοστιαία μεταβολή της ζήτησης σε μία μεταβολή του διαφημιστικού προϋπολογισμού κατά 1 τοις εκατό.
- Ένας γεωπόνος μπορεί να ενδιαφέρεται για τη μελέτη της εξάρτησης της απόδοσης μίας καλλιέργειας πχ του σιταριού από τη θερμοκρασία, τις βροχοπτώσεις κλπ.

1.4 Παλινδρόμηση έναντι Αιτιότητας

- Στο παράδειγμα της απόδοσης των καλλιεργειών, δεν υπάρχει **κανένας *στατιστικός λόγος*** να υποθέσουμε ότι οι βροχοπτώσεις δεν εξαρτώνται από την απόδοση των καλλιεργειών.
- Αντιμετωπίζουμε την απόδοση των καλλιεργειών ως ένα παράγοντα που εξαρτάται από τη βροχόπτωση (μεταξύ άλλων)
- Αυτό οφείλεται σε μη-στατιστικές εκτιμήσεις: δε μπορούμε να ελέγξουμε τις βροχοπτώσεις μεταβάλλοντας την απόδοση των καλλιεργειών.
- **Μία στατιστική σχέση καθεαυτή δε συνεπάγεται λογικά αιτιότητα**.
- Για να καταλογίσει κάποιος αιτιότητα, πρέπει να επικαλεστεί *a priori* θεωρητικές απόψεις.

1.5 Ανάλυση Συσχέτισης

- Η **ανάλυση συσχέτισης** (correlation analysis) είναι στενά συνδεδεμένη, αλλά εννοιολογικά πολύ διαφορετική, από την ανάλυση παλινδρόμησης, όπου ο πρωταρχικός στόχος είναι η μέτρηση της *δύναμης* ή του *βαθμού γραμμικής συσχέτισης* μεταξύ δύο μεταβλητών.
- Ο **συντελεστής συσχέτισης** (correlation coefficient), μετρά τη δύναμη αυτής της (γραμμικής) σχέσης.
- **Παράδειγμα:** συσχέτιση (το συντελεστή συσχέτισης) μεταξύ του καπνίσματος και του καρκίνου του πνεύμονα, μεταξύ των βαθμών των μαθητών στη στατιστική και στα μαθηματικά, μεταξύ των βαθμών στο λύκειο και των βαθμών στο πανεπιστήμιο και στις ώρες που διαβάζουν.
- Στην ανάλυση παλινδρόμησης προσπαθούμε να εκτιμήσουμε ή να προβλέψουμε τη μέση τιμή μίας μεταβλητής βάσει των σταθερών τιμών άλλων μεταβλητών.

1.6 Ορολογία και Συμβολισμοί

- Στη βιβλιογραφία οι όροι *εξαρτημένη μεταβλητή* και *ερμηνευτική μεταβλητή* περιγράφονται ποικιλοτρόπως.

Εξαρτημένη μεταβλητή	**Ερμηνευτική μεταβλητή**
↕	↕
Ερμηνευμένη μεταβλητή	Ανεξάρτητη μεταβλητή
↕	↕
Προβλεφθείσα μεταβλητή	Μεταβλητή πρόβλεψης
↕	↕
Παλινδρομούμενη μεταβλητή	Παλινδρομητής
↕	↕
Απόκρισης	Ερέθισμα
↕	↕
Ενδογενής	Εξωγενής
↕	↕
Αποτέλεσμα	Συμμεταβλητή
↕	↕
Ελεγχόμενη μεταβλητή	Μεταβλητή ελέγχου

1.6 Ορολογία και Συμβολισμοί

- Αν μελετούμε την εξάρτηση μίας μεταβλητής από μία μόνο ερμηνευτική μεταβλητή, όπως αυτή της καταναλωτικής δαπάνης από το πραγματικό εισόδημα, μία τέτοια μελέτη είναι γνωστή ως απλή, ή **διμεταβλητή ανάλυση παλινδρόμησης** (two-variable regression analysis).
- Αν μελετούμε την εξάρτηση μίας μεταβλητής από περισσότερες από μία ερμηνευτικές μεταβλητές, όπως στο παράδειγμα της απόδοσης των καλλιεργειών και της βροχόπτωσης, της θερμοκρασίας, της ηλιοφάνειας και των λιπασμάτων, τότε είναι γνωστή ως **ανάλυση πολλαπλής παλινδρόμησης** (multiple regression analysis).

1.6 Ορολογία και Συμβολισμοί

- Ο όρος **τυχαίο** (random) είναι συνώνυμος του όρου **στοχαστικό** (stochastic).
- Μία τυχαία ή στοχαστική μεταβλητή είναι μία μεταβλητή που μπορεί να πάρει οποιοδήποτε σύνολο τιμών, θετικών ή αρνητικών, με μία δεδομένη πιθανότητα.

1.6 Ορολογία και Συμβολισμοί

- Το γράμμα ***Y*** συνήθως δηλώνει την εξαρτημένη μεταβλητή
- Το γράμμα ***X*** (X_1, X_2,..., X_k) συνήθως δηλώνει τις ερμηνευτικές μεταβλητές, X_{k} είναι η $k^{η}$ ερμηνευτική μεταβλητή.
- Ο δείκτης ***i*** ή ***t*** συνήθως δηλώνει την $i^{η}$ ή $t^{η}$ παρατήρηση ή τιμή.
- $\boldsymbol{X_{ki}}$ (ή $\boldsymbol{X_{kt}}$) συνήθως δηλώνει την $i^{η}$ (ή $t^{η}$) παρατήρηση της μεταβλητής X_k.
- ***N*** (ή ***T***) συνήθως δηλώνει το συνολικό αριθμό των παρατηρήσεων ή των τιμών στον πληθυσμό
- ***n*** (ή ***t***) συνήθως δηλώνει το συνολικό αριθμό των παρατηρήσεων στο δείγμα.
- Ο δείκτης ***i*** χρησιμοποιείται για **διαστρωματικά στοιχεία** (cross-sectional data) (δηλαδή, στοιχεία που έχουν συλλεχθεί σε μία χρονική στιγμή)
- Ο δείκτης ***t*** χρησιμοποιείται για **χρονοσειρές στοιχείων** (time series data) (δηλαδή, στοιχεία που έχουν συλλεχθεί κατά τη διάρκεια μίας χρονικής περιόδου).

1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- **Τύποι Στοιχείων**
- *Τρεις τύποι στοιχείων:*

1. **Χρονοσειρές στοιχείων**
2. **Διαστρωματικά στοιχεία**
3. **Ομαδοποιημένα στοιχεία** (δηλαδή, ο συνδυασμός των χρονοσειρών και των διαστρωματικών στοιχείων)

1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- **Τύποι Στοιχείων**

1. **Χρονοσειρές στοιχείων**

- Συλλέγονται:
- **καθημερινά** (π.χ., οι τιμές των μετοχών, δελτία καιρού)
- **εβδομαδιαία** (π.χ., στοιχεία προσφοράς χρήματος)
- **μηνιαία** [π.χ., το ποσοστό ανεργίας, ο Δείκτης Τιμών Καταναλωτή (ΔΤΚ)]
- **τριμηνιαία** (για παράδειγμα, το ΑΕΠ)
- **ετήσια** (π.χ., κρατικοί προϋπολογισμοί)
- **ανά πενταετία**, κάθε 5 χρόνια (για παράδειγμα, η απογραφή των κατασκευών)
- **ανά δεκαετία** (π.χ., η απογραφή του πληθυσμού).

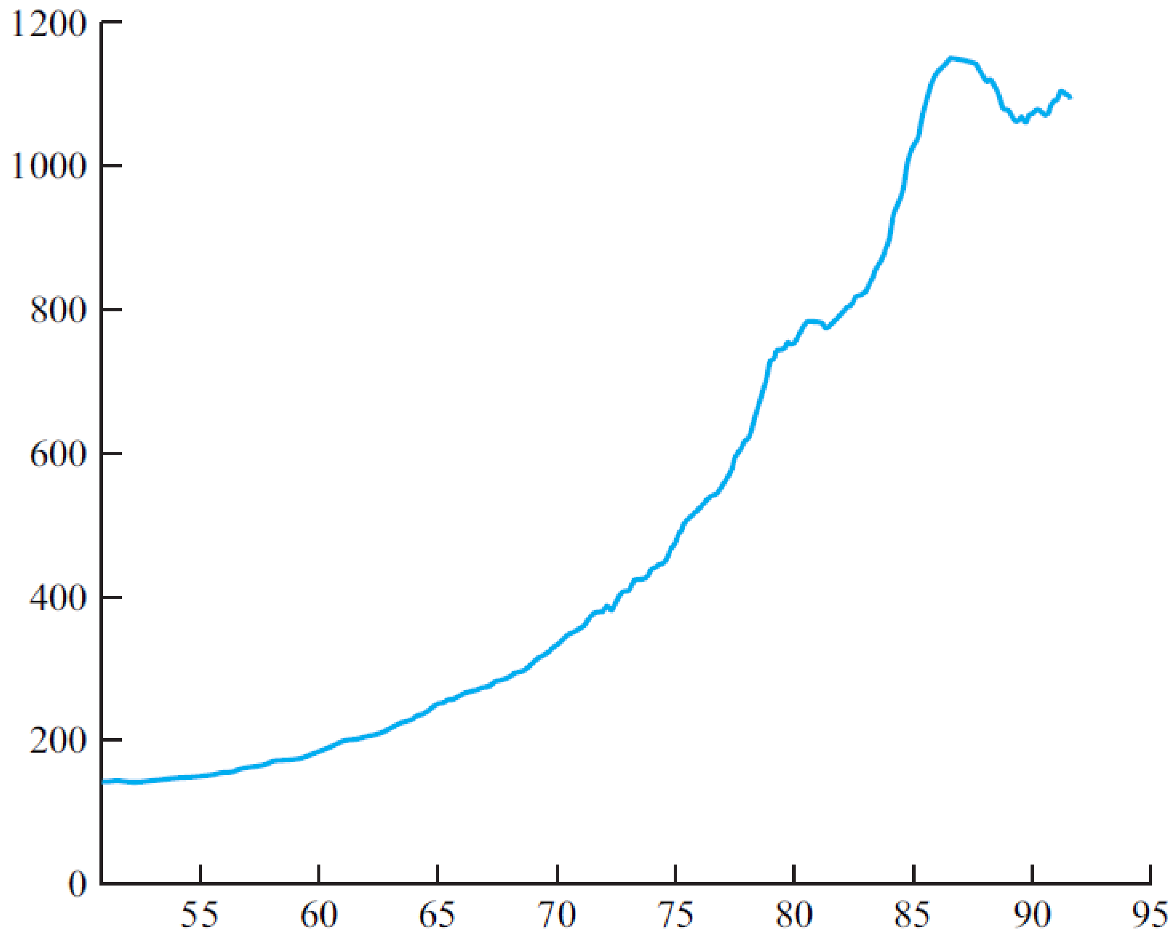
1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- **Τύποι Στοιχείων**

1. **Χρονοσειρές στοιχείων**

- Παρουσιάζουν ιδιαίτερα προβλήματα για τους οικονομέτρες:
- Οι περισσότερες εμπειρικές εργασίες που βασίζονται σε χρονοσειρές στοιχείων προϋποθέτουν ότι οι υποκείμενες χρονοσειρές είναι **στάσιμες** (stationary).
- *Μία χρονοσειρά είναι στάσιμη, αν η μέση τιμή της και η διακύμανσή της δε διαφέρουν συστηματικά διαχρονικά*

ΔΙΑΓΡΑΜΜΑ 1.5 Προσφορά χρήματος M1: Ηνωμένες Πολιτείες, 1951:01-1999:09.



1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- **Τύποι Στοιχείων**

2. Διαστρωματικά στοιχεία

- Τα διαστρωματικά στοιχεία είναι στοιχεία για μία ή περισσότερες μεταβλητές που αναφέρονται *στην ίδια χρονική στιγμή*, όπως είναι η απογραφή του πληθυσμού κάθε 10 χρόνια, οι έρευνες των καταναλωτικών δαπανών, και, φυσικά, οι δημοσκοπήσεις.

1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- **Τύποι Στοιχείων**

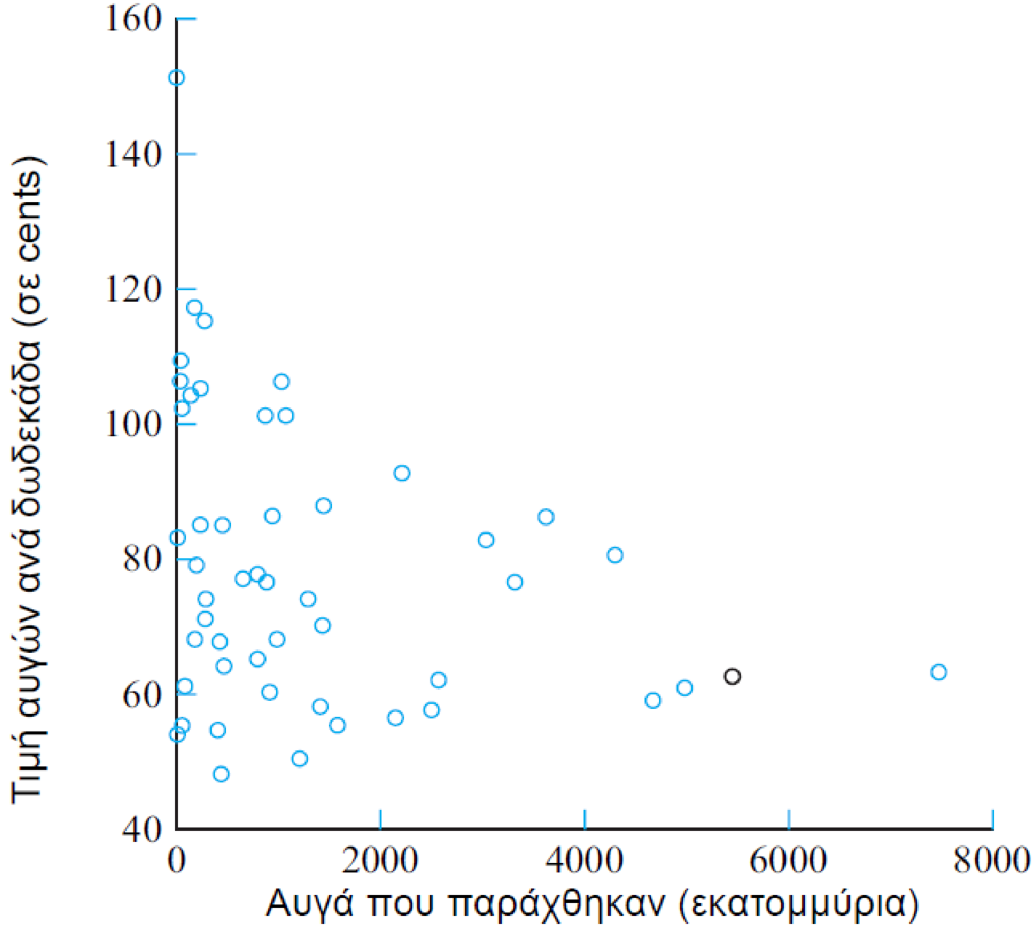
2. Διαστρωματικά στοιχεία

- Τα διαστρωματικά στοιχεία παρουσιάζουν επίσης προβλήματα
- Σημαντικότερο πρόβλημα είναι αυτό της *ετερογένειας (heterogeneity)*.
- Όταν συμπεριλαμβάνουμε ανομοιογενείς μονάδες σε μία στατιστική ανάλυση, πρέπει να λαμβάνουμε υπόψη μας το αποτέλεσμα του **μεγέθους** ή της **κλίμακας**, ώστε να μη συγχέουμε μήλα με πορτοκάλια.

ΠΙΝΑΚΑΣ 1.1 Παραγωγή αυγών Η.Π.Α.

Πολιτεία	Y_1	Y_2	X_1	X_2	Πολιτεία	Y_1	Y_2	X_1	X_2
AL	2.206	2.186	92,7	91,4	MT	172	164	68,0	66,0
AK	0,7	0,7	151,0	149,0	NE	1.202	1.400	50,3	48,9
AZ	73	74	61,0	56,0	NV	2,2	1,8	53,9	52,7
AR	3.620	3.737	86,3	91,8	NH	43	49	109,0	104,0
CA	7.472	7.444	63,4	58,4	NJ	442	491	85,0	83,0
CO	788	873	77,8	73,0	NM	283	302	74,0	70,0
CT	1.029	948	106,0	104,0	NY	975	987	68,1	64,0
DE	168	164	117,0	113,0	NC	3.033	3.045	82,8	78,7
FL	2.586	2.537	62,0	57,2	ND	51	45	55,2	48,0
GA	4.302	4.301	80,6	80,8	OH	4.667	4.637	59,1	54,7
HI	227,5	224,5	85,0	85,5	OK	869	830	101,0	100,0
ID	187	203	79,1	72,9	OR	652	686	77,0	74,6
IL	793	809	65,0	70,5	PA	4.976	5.130	61,0	52,0
IN	5.445	5.290	62,7	60,1	RI	53	50	102,0	99,0
IA	2.151	2.247	56,5	53,0	SC	1.422	1.420	70,1	65,9
KS	404	389	54,5	47,8	SD	435	602	48,0	45,8
KY	412	483	67,7	73,5	TN	277	279	71,0	80,7
LA	273	254	115,0	115,0	TX	3.317	3.356	76,7	72,6
ME	1.069	1.070	101,0	97,0	UT	456	486	64,0	59,0
MD	885	898	76,6	75,4	VT	31	30	106,0	102,0
MA	235	237	105,0	102,0	VA	943	988	86,3	81,2
MI	1.406	1.396	58,0	53,8	WA	1.287	1.313	74,1	71,5
MN	2.499	2.697	57,7	54,0	WV	136	174	104,0	109,0
MS	1.434	1.468	87,8	86,7	WI	910	873	60,1	54,0
MO	1.580	1.622	55,4	51,5	WY	1,7	1,7	83,0	83,0

ΔΙΑΓΡΑΜΜΑ 1.6 Σχέση μεταξύ των παραχθέντων αυγών και των τιμών, 1990



1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- **Τύποι Στοιχείων**

3. **Ομαδοποιημένα στοιχεία** (δηλαδή, ο συνδυασμός των χρονοσειρών και των διαστρωματικών στοιχείων)

- Τα στοιχεία αφορούν τόσο χρονοσειρές όσο και διαστρωματικά στοιχεία.
- ***Χρονοσειρές διαστρωματικών στοιχείων – Panel, Longitudinal, Micropanel Data***
- Ένας ειδικός τύπος ομαδοποιημένων στοιχείων σύμφωνα με τον οποίο η ***ίδια*** **διαστρωματική μονάδα** (ας πούμε, μία οικογένεια ή μία επιχείρηση) είναι αντικείμενο μελέτης **για μία χρονική περίοδο**.

1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- ## Οι Πηγές των Στοιχείων
- Τα στοιχεία που συλλέγονται από διάφορους φορείς μπορεί να είναι **πειραματικά** ή **μη πειραματικά**.
- Στα **πειραματικά στοιχεία**, τα οποία συλλέγονται συχνά στις φυσικές επιστήμες, ο ερευνητής μπορεί να επιθυμεί να συλλέξει στοιχεία, διατηρώντας ορισμένους παράγοντες σταθερούς, προκειμένου να αξιολογηθεί ο αντίκτυπος ορισμένων παραγόντων σε ένα συγκεκριμένο φαινόμενο.
- Στις κοινωνικές επιστήμες, τα στοιχεία που συναντά συνήθως κάποιος είναι **μη πειραματικής φύσης**, δηλαδή, δεν υπόκεινται στον έλεγχο του ερευνητή.

1.7 Η Φύση και οι Πηγές των Στοιχείων για την Οικονομική Ανάλυση

- **Η Ακρίβεια των Στοιχείων**
- *Η ποιότητα των στοιχείων συχνά δεν είναι καλή*

1. Η πλειοψηφία των στοιχείων που αφορούν τις κοινωνικές επιστήμες είναι μη πειραματικής φύσης.
2. Προκύπτουν σφάλματα μετρήσεων από προσεγγίσεις και στρογγυλοποιήσεις.
3. Συχνά έχουμε **μεροληψία λόγω επιλογής δείγματος** (sample selectivity bias).
4. Οι μέθοδοι δειγματοληψίας μπορεί να διαφέρουν τόσο πολύ που είναι συχνά δύσκολο να συγκριθούν τα αποτελέσματα που προκύπτουν από τα διάφορα δείγματα.
5. Τα οικονομικά στοιχεία είναι διαθέσιμα σε πολύ συγκεντρωτικό επίπεδο. Αυτά ίσως δε μπορούν να μας δώσουν επαρκείς πληροφορίες για τα άτομα ή τις μικροοικονομικές μονάδες που μπορεί να είναι το βασικό αντικείμενο της μελέτης.
6. Ενδεχομένως να είναι απαραίτητο ορισμένα στοιχεία να δημοσιεύονται μόνο σε πολύ συγκεντρωτική μορφή για λόγους εμπιστευτικότητας.

2. Διμεταβλητό Υπόδειγμα Παλινδρόμησης: Το πρόβλημα της εκτίμησης

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- Υπενθυμίζεται η διμεταβλητή PRF (population regression function):

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

- Ωστόσο, όπως επισημάναμε προηγουμένως, η PRF δεν είναι άμεσα παρατηρήσιμη. Μπορούμε να την εκτιμήσουμε από την SRF (sample regression function):

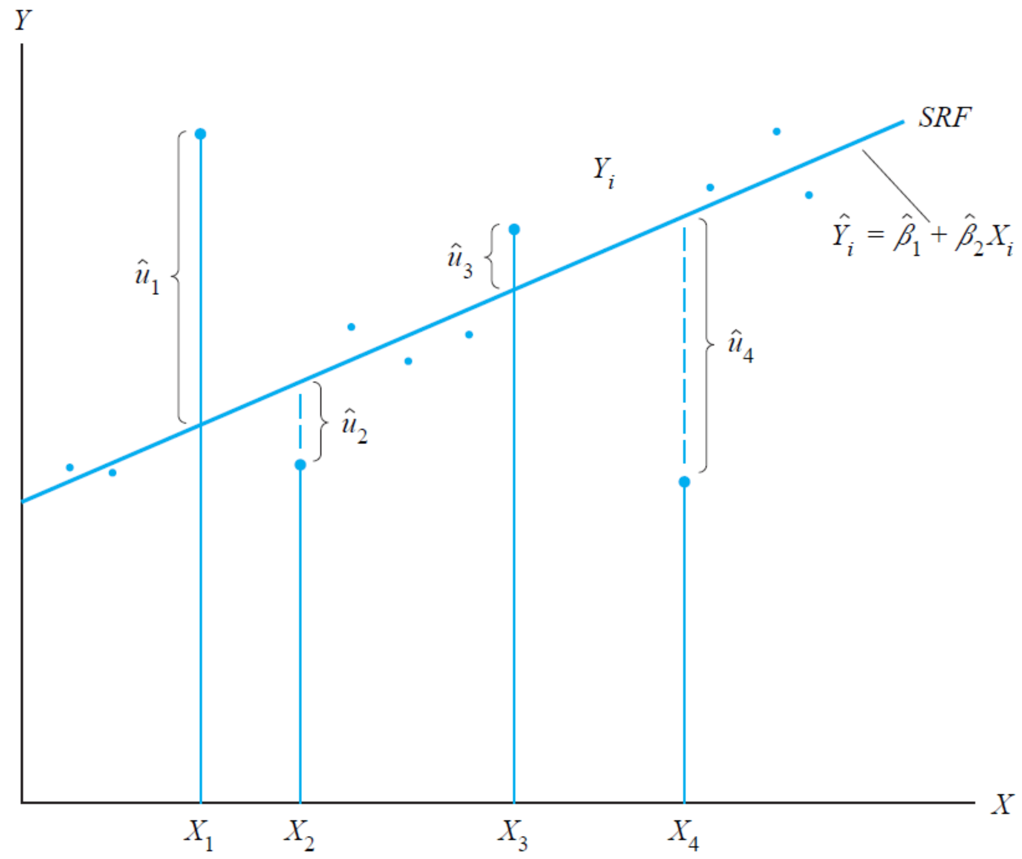
$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned}$$

- όπου $\hat{Y}_i$ είναι η εκτιμηθείσα (υπό συνθήκη μέση) τιμή της Y_i.
- Πώς καθορίζεται όμως η ίδια η SRF; Για να το κατανοήσουμε αυτό, ας προχωρήσουμε ως εξής. Κατ' αρχάς, παρουσιάζουμε την Εξίσωση $Y_i = \hat{Y}_i + \hat{u}_i$ ως

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \end{aligned}$$

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΔΙΑΓΡΑΜΜΑ 2.1** Κριτήριο των ελαχίστων τετραγώνων



2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- Το αλγεβρικό άθροισμα αυτών των καταλοίπων είναι μηδέν παρόλο που τα $\hat{u}_1$ και $\hat{u}_4$ είναι διάσπαρτα ευρύτερα γύρω από την SRF από ότι τα $\hat{u}_2$ και $\hat{u}_3$.
- Μπορούμε να αποφύγουμε αυτό το πρόβλημα αν υιοθετήσουμε το *κριτήριο των ελαχίστων τετραγώνων*, το οποίο ορίζει ότι η SRF μπορεί να καθοριστεί με τέτοιο τρόπο ώστε το

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

- να είναι όσο το δυνατόν μικρότερο, όπου $\hat{u}_i^2$ είναι το τετράγωνο των καταλοίπων.
- Τετραγωνίζοντας τα $\hat{u}_i$, αυτή η μέθοδος δίνει περισσότερη βαρύτητα σε κατάλοιπα όπως τα $\hat{u}_1$ και $\hat{u}_4$ στο Διάγραμμα 2.1 από ότι στα κατάλοιπα $\hat{u}_2$ και $\hat{u}_3$.

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΠΙΝΑΚΑΣ 2.1** Πειραματικός Προσδιορισμός της SRF

Yi	**Xt**	$\hat{Y}_{1i}$	$\hat{u}_{1i}$	$\hat{u}_{1i}^2$	$\hat{y}_{2i}$	$\hat{u}_{2i}$	$\hat{u}_{2i}^2$
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
4	1	2,929	1,071	1,147	4	0	0
5	4	7,000	−2,000	4,000	7	−2	4
7	5	8,357	−1,357	1,841	8	−1	1
12	6	9,714	2,286	5,226	9	3	9
Σύνολο: 28	16		0,0	12,214		0	14

Σημειώσεις: $\hat{Y}_{1i} = 1{,}572 + 1{,}357X_i$ (δηλ. $\hat{\beta}_1 = 1{,}572$ και $\hat{\beta}_2 = 1{,}357$), $\hat{Y}_{2i} = 3 + 1X_i$ (δηλ. $\hat{\beta}_1 = 3$ και $\hat{\beta}_2 = 1$), $\hat{u}_{1i} = (Y_i - \hat{Y}_{1i})$, $\hat{u}_{2i} = (Y_i - \hat{Y}_{2i})$.

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- Η διαδικασία της παραγώγισης δίνει τις παρακάτω εξισώσεις για την εκτίμηση των β_1 και β_2:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

- όπου *n* είναι το μέγεθος του δείγματος. Αυτές οι ταυτόχρονες εξισώσεις είναι γνωστές ως **κανονικές εξισώσεις** (normal equations).

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- Λύνοντας τις κανονικές εξισώσεις ταυτόχρονα, έχουμε

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum x_i y_i}{\sum x_i^2}$$

- όπου $\bar{X}$ και $\bar{Y}$ είναι οι μέσοι του δείγματος των *X* και *Y* και ορίζουμε $x_i = (X_i - \bar{X})$ και $y_i = (Y_i - \bar{Y})$. *Στο εξής, θεωρούμε ότι τα πεζά γράμματα υποδηλώνουν αποκλίσεις από τις μέσες τιμές.*

(Εξ. 2.1.7)

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

$$= \bar{Y} - \hat{\beta}_2 \bar{X}$$

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- Οι εκτιμητές που έχουν προκύψει είναι γνωστοί ως **εκτιμητές των ελαχίστων τετραγώνων** (least-squares estimators), καθώς προέρχονται από την αρχή των ελαχίστων τετραγώνων.
- Σημειώστε τις ακόλουθες **αριθμητικές ιδιότητες** των εκτιμητών που προκύπτουν από τη μέθοδο OLS:
- «Οι αριθμητικές ιδιότητες είναι εκείνες που ισχύουν λόγω της χρήσης των ελαχίστων τετραγώνων, ανεξάρτητα από το πώς προέκυψαν τα στοιχεία.»
- Εξετάζουμε επίσης τις **στατιστικές ιδιότητες** των εκτιμητών της OLS, δηλαδή, τις ιδιότητες «που ισχύουν μόνο με βάση ορισμένες υποθέσεις σχετικά με τον τρόπο που προέκυψαν τα στοιχεία.»

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

I. Οι εκτιμητές της OLS εκφράζονται μόνο σε όρους παρατηρήσιμων (δηλαδή, δειγματικών) ποσοτήτων (δηλαδή, των *X* και *Y*). Ως εκ τούτου, μπορούν να υπολογιστούν με ευκολία.

II. Είναι **σημειακοί εκτιμητές** (point estimators)˙ δηλαδή, δοθέντος του δείγματος, κάθε εκτιμητής θα παρέχει μόνο μία συγκεκριμένη (σημειακή) τιμή της σχετικής παραμέτρου του πληθυσμού.

III. Μόλις οι εκτιμητές της OLS προκύψουν από το δείγμα, η γραμμή παλινδρόμησης του δείγματος (Διάγραμμα 2.1) μπορεί να προκύψει πολύ εύκολα.

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- Η γραμμή παλινδρόμησης που προέκυψε έχει τις ακόλουθες ιδιότητες:

1. Διέρχεται από τις μέσες τιμές των δειγμάτων των *Y* και *X*. Το γεγονός αυτό είναι προφανές από την Εξ. (2.1.7), καθώς η τελευταία μπορεί να γραφεί ως $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$
2. Η μέση τιμή της εκτιμημένης $Y = \hat{Y}_i$ είναι ίση με τη μέση τιμή της πραγματικής *Y* για

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$
$$= \left(\bar{Y} - \hat{\beta}_2 \bar{X}\right) + \hat{\beta}_2 X_i$$
$$= \bar{Y} + \hat{\beta}_2 (X_i - \bar{X})$$

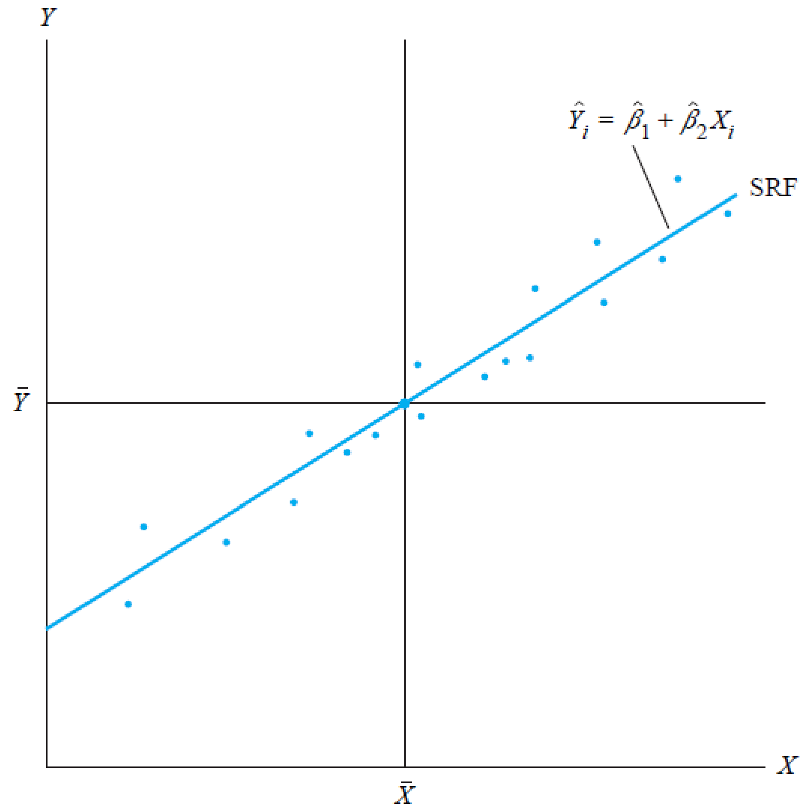
Αθροίζοντας τις δύο πλευρές αυτής της τελευταίας ισότητας με τις τιμές του δείγματος και διαιρώντας με το μέγεθος του δείγματος *n* παίρνουμε

$$\bar{\hat{Y}} = \bar{Y} \quad (2.1.10)$$

3. Η μέση τιμή των καταλοίπων $\hat{u}_i$ είναι μηδέν.
4. Τα κατάλοιπα $\hat{u}_i$ είναι ασυσχέτιστα με την προβλεπόμενη Y_i.
5. Το κατάλοιπα $\hat{u}_i$ είναι ασυσχέτιστα με τη *Xi.*

2.1 Η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΔΙΑΓΡΑΜΜΑ 2.2** Το διάγραμμα δείχνει ότι η γραμμή παλινδρόμησης του δείγματος διέρχεται από τις μέσες τιμές των δειγμάτων των *Y* και *X*.



2.2 Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- Το **Gaussian, τυπικό**, ή **κλασικό γραμμικό υπόδειγμα παλινδρόμησης** (CLRM), το οποίο είναι ο ακρογωνιαίος λίθος του μεγαλύτερου μέρους της οικονομετρικής θεωρίας, κάνει 7 υποθέσεις.

2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΥΠΟΘΕΣΗ 1**
- **Γραμμικό Υπόδειγμα Παλινδρόμησης**: Το υπόδειγμα παλινδρόμησης είναι **γραμμικό ως προς τις παραμέτρους**, ωστόσο μπορεί να είναι ή μπορεί να μην είναι γραμμικό ως προς τις μεταβλητές. Αυτό είναι το υπόδειγμα παλινδρόμησης:

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

- Το υπόδειγμα αυτό μπορεί να επεκταθεί ώστε να περιλαμβάνει περισσότερες ερμηνευτικές μεταβλητές.

2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΥΠΟΘΕΣΗ 2**
- **Σταθερές Τιμές της *X* ή Τιμές της *X* Ανεξάρτητες από τον Όρο Σφάλματος**: Οι τιμές που παίρνει ο παλινδρομητής *X* μπορούν να θεωρηθούν σταθερές σε επαναλαμβανόμενα δείγματα (η περίπτωση του σταθερού παλινδρομητή) ή μπορεί η δειγματοληψία να πραγματοποιείται σε συνδυασμό με την εξαρτημένη μεταβλητή *Y* (η περίπτωση του στοχαστικού παλινδρομητή).
- Στην τελευταία περίπτωση, θεωρούμε ότι η (οι) μεταβλητή (-ές) *X* και ο όρος σφάλματος είναι ανεξάρτητοι, δηλαδή, $\text{cov}(X_i, u_i) = 0$.

2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΥΠΟΘΕΣΗ 3**
- **Μηδενική Μέση Τιμή του Διαταρακτικού Όρου u_i**: Δεδομένης της τιμής της X_i, η μέση ή η προσδοκώμενη τιμή του τυχαίου διαταρακτικού όρου u_i είναι μηδέν. Συμβολικά, έχουμε

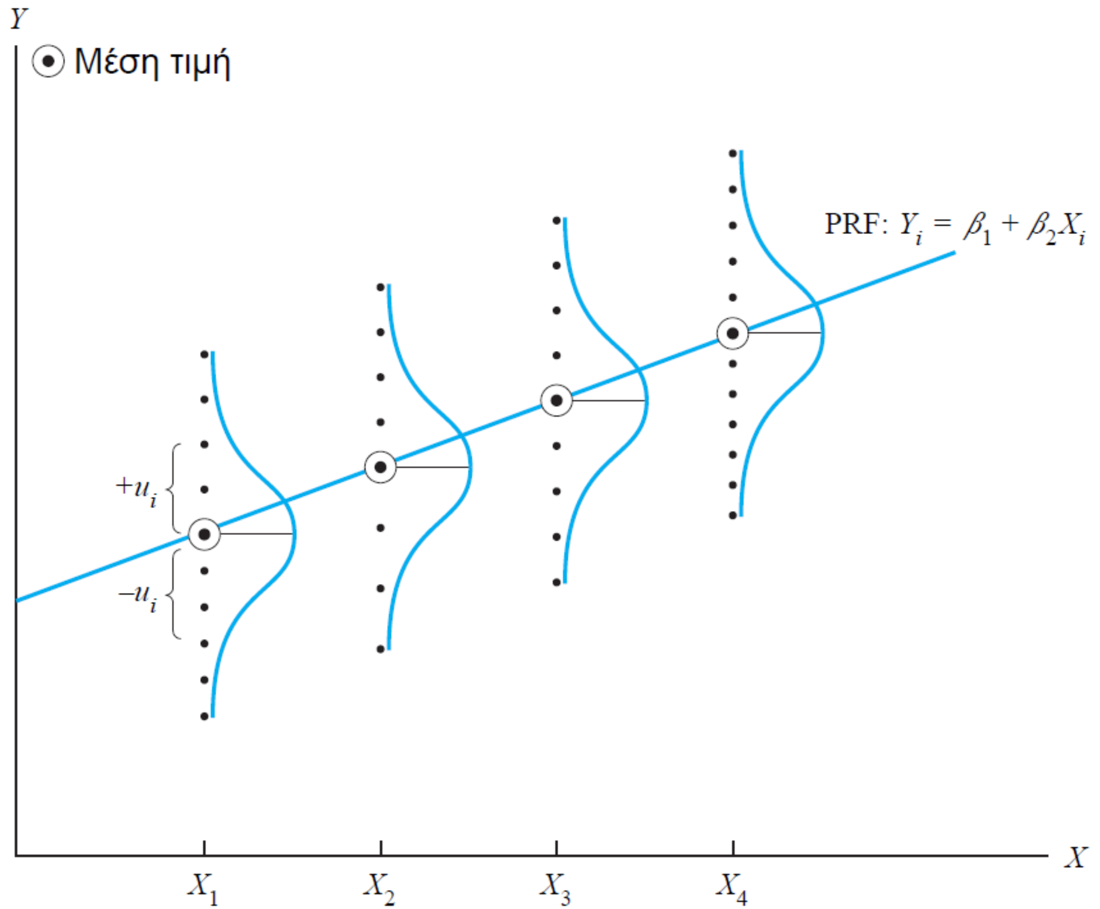
$$E(u_i \mid X_i) = 0$$

- Ή, εάν η X είναι μη-στοχαστική

$$E(u_i) = 0$$

2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΔΙΑΓΡΑΜΜΑ 2.3** Υπό συνθήκη κατανομή των διαταρακτικών όρων u_i.



2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

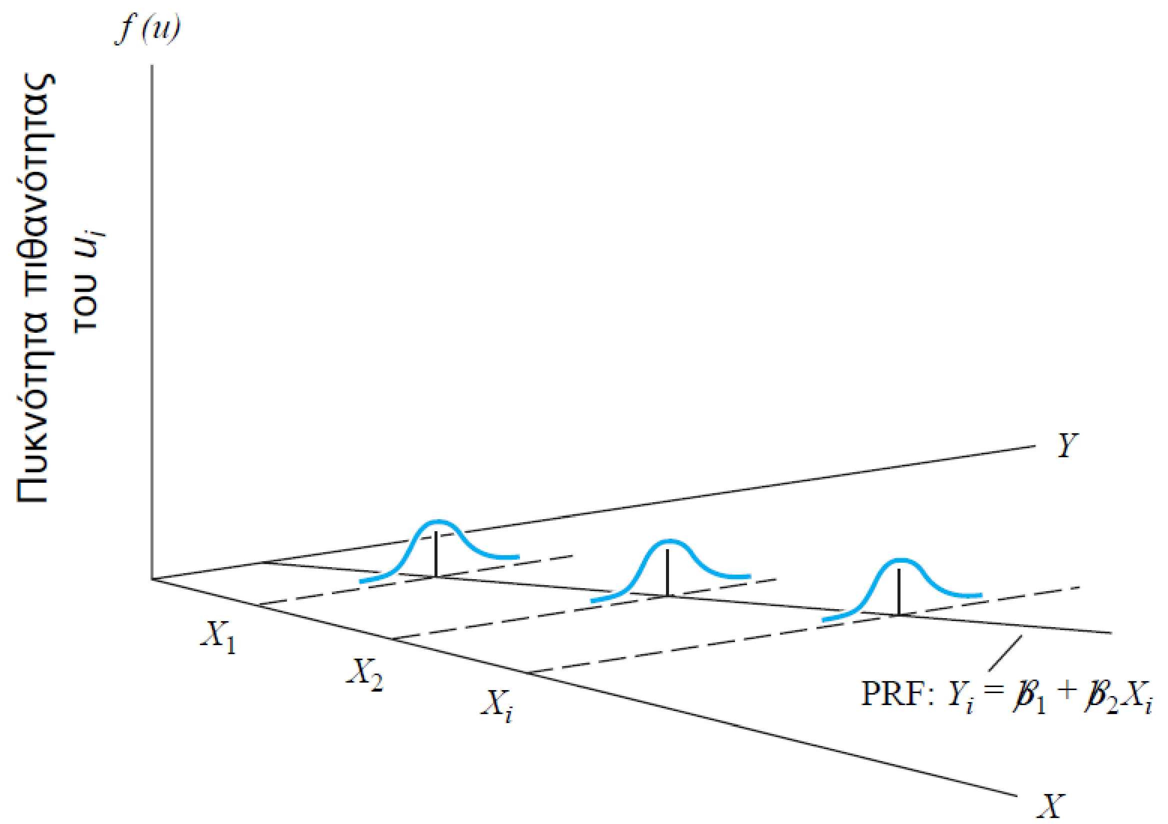
- **ΥΠΟΘΕΣΗ 4**
- **Ομοσκεδαστικότητα ή Σταθερή Διακύμανση του u_i**: Η διακύμανση του διαταρακτικού όρου ή όρου σφάλματος είναι η ίδια, ανεξάρτητα από την τιμή της X. Συμβολικά,

$$\begin{aligned} \text{var}(u_i) &= E[u_i - E(u_i \mid X_i)]^2 \\ &= E(u_i^2 \mid X_i), \text{ λόγω της Υπόθεσης 3} \\ &= E(u_i^2) \text{ αν η } X_i \text{ είναι μη-στοχαστική} \\ &= \sigma^2 \end{aligned}$$

- όπου var είναι η διακύμανση.

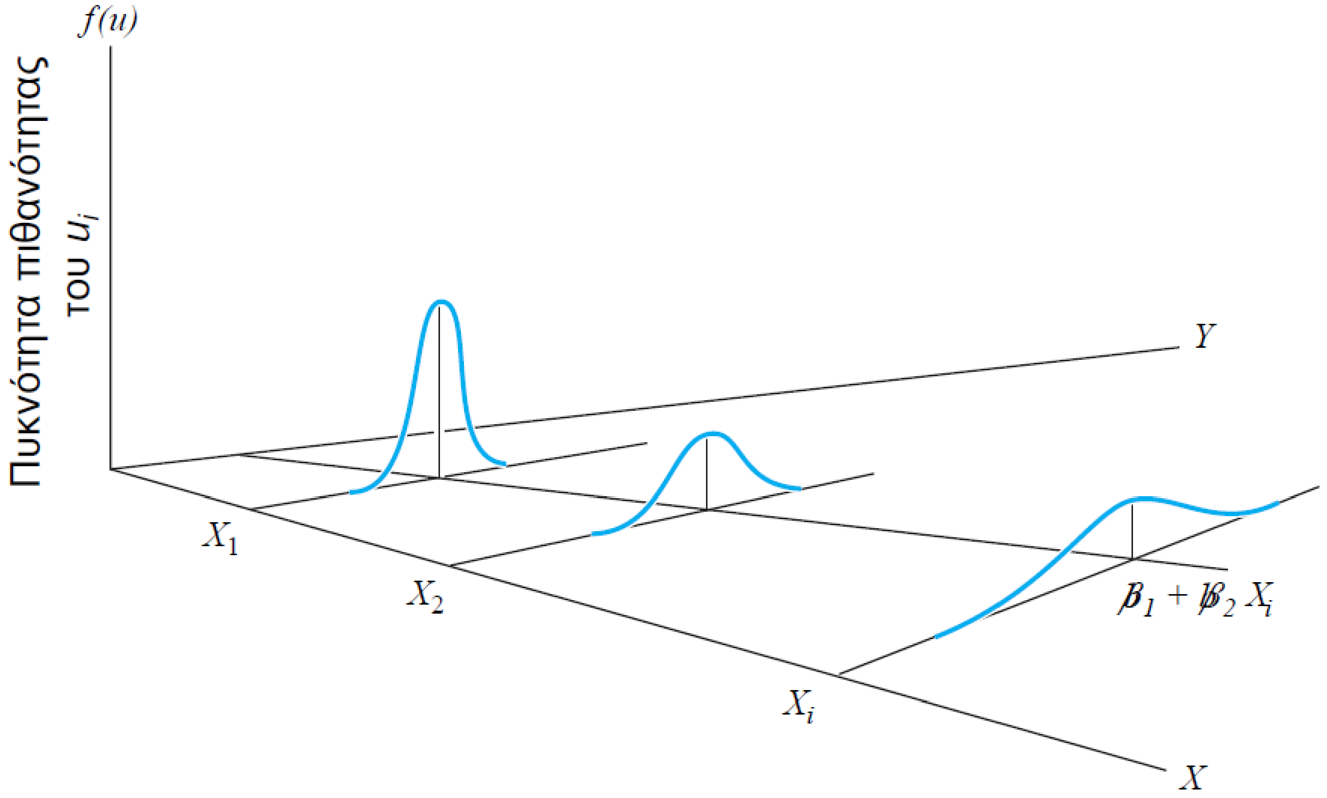
2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- ΔΙΑΓΡΑΜΜΑ 2.4 Ομοσκεδαστικότητα



2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΔΙΑΓΡΑΜΜΑ 2.5** Ετεροσκεδαστικότητα



2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΥΠΟΘΕΣΗ 5**
- **Δεν Υπάρχει Αυτοσυσχέτιση μεταξύ των Διαταρακτικών Όρων:** Δοθέντων οποιονδήποτε δύο τιμών των *X*, X_i και X_j ($i \neq j$), η συσχέτιση μεταξύ οποιονδήποτε δύο u_i και u_j ($i \neq j$) είναι μηδέν. Εν ολίγοις, οι παρατηρήσεις περιέχονται σε ανεξάρτητα μεταξύ τους δείγματα. Συμβολικά,

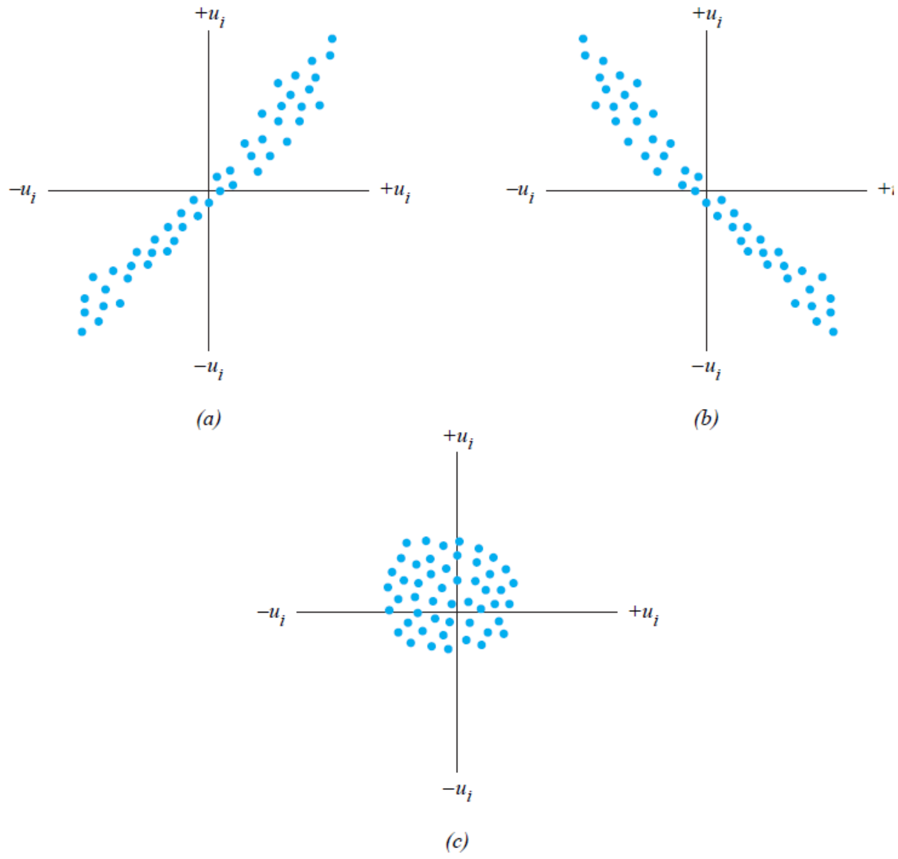
$$\text{cov}(u_i, u_j \mid X_i, X_j) = 0$$

$$\text{cov}(u_i, u_j) = 0 \text{ αν η } X \text{ είναι μη-στοχαστική}$$

- όπου *i* και *j* είναι δύο διαφορετικές παρατηρήσεις και όπου cov σημαίνει συνδιακύμανση.

2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΔΙΑΓΡΑΜΜΑ 2.6** Πρότυπα συσχέτισης μεταξύ των διαταρακτικών όρων. (α) θετική σειριακή συσχέτιση˙ (β) αρνητική σειριακή συσχέτιση˙ (γ) μηδενική συσχέτιση.



2.2 Το Κλασικό Γραμμικό Υπόδειγμα Παλινδρόμησης: Οι Υποθέσεις στις Οποίες Στηρίζεται η Μέθοδος των Ελαχίστων Τετραγώνων

- **ΥΠΟΘΕΣΗ 6**
- **Ο Αριθμός των Παρατηρήσεων *n* Πρέπει να Είναι Μεγαλύτερος από τον Αριθμό των Παραμέτρων που Θα Εκτιμηθούν:** Εναλλακτικά, ο αριθμός των παρατηρήσεων πρέπει να είναι μεγαλύτερος από τον αριθμό των ερμηνευτικών μεταβλητών.
- **ΥΠΟΘΕΣΗ 7 Η Φύση των Μεταβλητών *X*:** Οι τιμές της *X* σε ένα δεδομένο δείγμα δεν πρέπει να είναι όλες οι ίδιες. Τεχνικά, var (*X*) πρέπει να είναι ένας θετικός αριθμός. Επιπλέον, δε μπορεί να υπάρχουν **ακραίες τιμές** ή **έκτοπα** (outliers) στις τιμές της μεταβλητής *X*, δηλαδή, τιμές που είναι πολύ μεγάλες σε σχέση με τις υπόλοιπες παρατηρήσεις.

2.3 Ακρίβεια ή Τυπικά Σφάλματα των Εκτιμήσεων των Ελαχίστων Τετραγώνων

- Δεδομένου ότι τα στοιχεία είναι πιθανό να μεταβάλλονται από δείγμα σε δείγμα, οι εκτιμήσεις, εκ των πραγμάτων, θα αλλάξουν.
- Ως εκ τούτου, αυτό που χρειάζεται είναι ένα μέτρο της «αξιοπιστίας» ή της **ακρίβειας** των εκτιμητών $\hat{\beta}_1$ και $\hat{\beta}_2$.
- Στη στατιστική, η ακρίβεια μίας εκτίμησης μετράται με το τυπικό της σφάλμα (se).

2.3 Ακρίβεια ή Τυπικά Σφάλματα των Εκτιμήσεων των Ελαχίστων Τετραγώνων

Τα τυπικά σφάλματα των εκτιμήσεων της OLS προέρχονται από τις εξής σχέσεις:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

$$\operatorname{se}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$\operatorname{var}(\hat{\beta}_1) = \frac{\sum x_i^2}{n \sum x_i^2} \sigma^2$$

$$\operatorname{se}(\hat{\beta}_1) = \sqrt{\frac{\sum x_i^2}{n \sum x_i^2}} \sigma$$

όπου var = διακύμανση και se = τυπικό σφάλμα και όπου σ^2 είναι η συνεχής ή ομοσκεδαστική διακύμανση του u_i της Υπόθεσης 4.

2.3 Ακρίβεια ή Τυπικά Σφάλματα των Εκτιμήσεων των Ελαχίστων Τετραγώνων

- Η σ^2 εκτιμάται από τον ακόλουθο τύπο:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$$

- όπου $\hat{\sigma}^2$ είναι ο εκτιμητής της OLS της πραγματικής αλλά άγνωστης σ^2 και όπου η έκφραση $n-2$ είναι γνωστή ως **ο αριθμός των βαθμών ελευθερίας** (number of degrees of freedom – df), $\sum \hat{u}_i^2$ είναι το **άθροισμα των τετραγώνων των καταλοίπων ή RSS**.
- Μόλις γίνει γνωστό το $\sum \hat{u}_i^2$ ο $\hat{\sigma}^2$ μπορεί να υπολογιστεί εύκολα. Το $\sum \hat{u}_i^2$ μπορεί να υπολογιστεί από την ακόλουθη εξίσωση (βλ.Παράρτημα για την απόδειξη):

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2$$

2.3 Ακρίβεια ή Τυπικά Σφάλματα των Εκτιμήσεων των Ελαχίστων Τετραγώνων

- Εφόσον

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

- μία εναλλακτική έκφραση για τον υπολογισμό του $\sum \hat{u}_i^2$ είναι

$$\sum \hat{u}_i^2 = \sum y_i^2 - \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2}$$

- Η θετική τετραγωνική ρίζα του $\hat{\sigma}^2$

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}}$$

- Είναι γνωστή ως το **τυπικό σφάλμα εκτίμησης** (standard error of estimate) ή το **τυπικό σφάλμα της παλινδρόμησης** (standard error of the regression) (se).
- Είναι απλώς η τυπική απόκλιση των τιμών της *Y* από την εκτιμημένη γραμμή παλινδρόμησης και χρησιμοποιείται συχνά ως ένα μέτρο για την «καλή προσαρμογή» (goodness of fit) της εκτιμημένης γραμμής παλινδρόμησης

2.4 Ιδιότητες των Εκτιμητών των Ελαχίστων Τετραγώνων: Το Θεώρημα Gauss-Markov

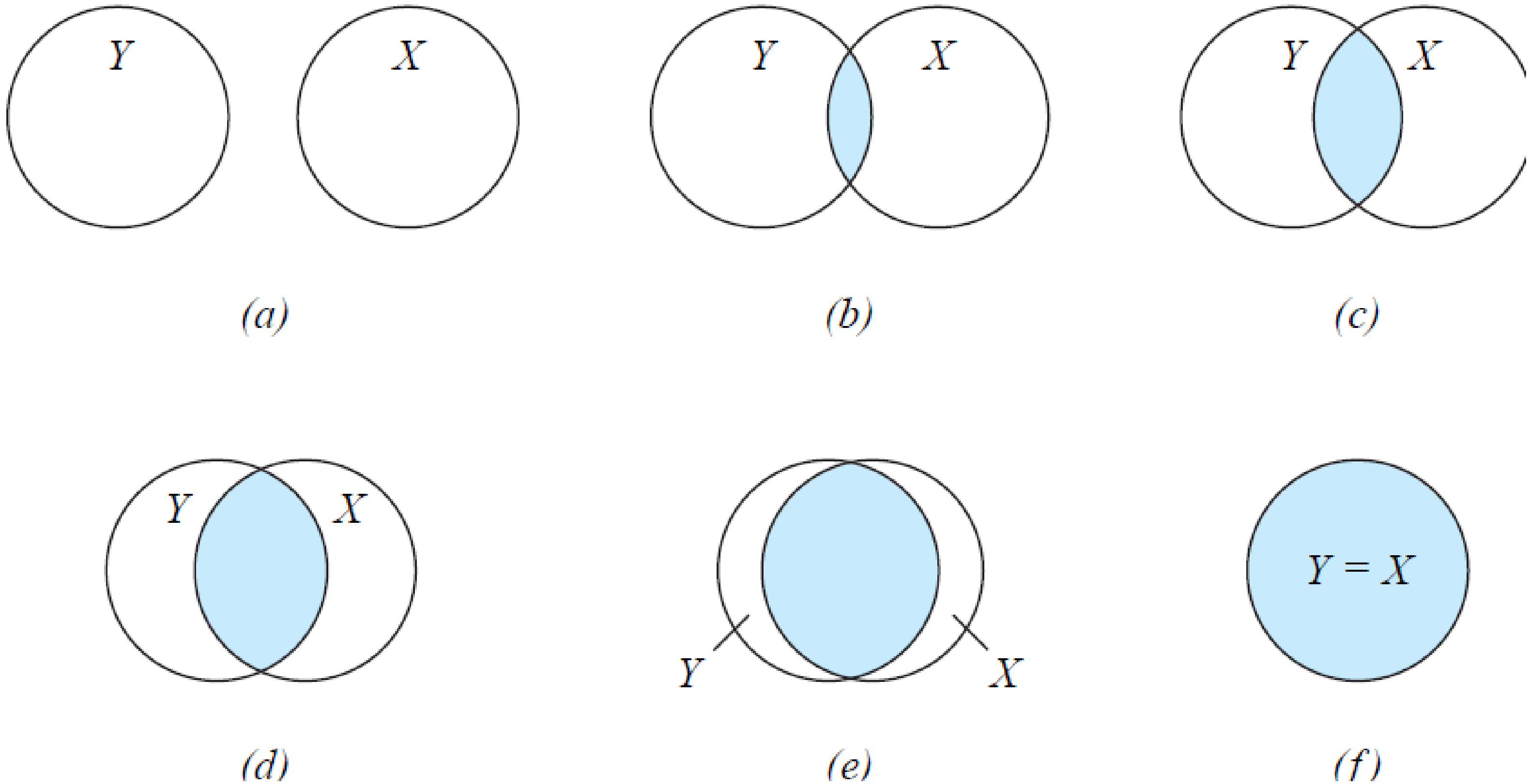
1. Είναι **γραμμικός**, δηλαδή, μία γραμμική συνάρτηση μίας τυχαίας μεταβλητής, όπως είναι η εξαρτημένη μεταβλητή Y στο υπόδειγμα παλινδρόμησης.
2. Είναι **αμερόληπτος**, δηλαδή, η μέση ή προσδοκώμενη τιμή του, $E(\hat{\beta}_2)$, είναι ίση με την πραγματική τιμή, β_2.
3. Έχει ελάχιστη διακύμανση της τάξης όλων των αντίστοιχων γραμμικών αμερόληπτων εκτιμητών˙ ένας αμερόληπτος εκτιμητής με τη μικρότερη διακύμανση είναι γνωστός ως ένας αποτελεσματικός εκτιμητής.
4. ***Το Θεώρημα Gauss-Markov*** Λαμβάνοντας υπόψη τις υποθέσεις του κλασσικού γραμμικού υποδείγματος παλινδρόμησης, οι εκτιμητές των ελαχίστων τετραγώνων, της τάξης των αμερόληπτων γραμμικών εκτιμητών, έχουν ελάχιστη διακύμανση, δηλαδή, είναι **BLUE.**

2.5 Ο Συντελεστής Προσδιορισμού r^2: Ένα Μέτρο της «Καλής Προσαρμογής»

- **Καλή προσαρμογή** (goodness of fit): πόσο «καλά» προσαρμόζεται στα στοιχεία η γραμμή παλινδρόμησης του δείγματος.
- Ο **συντελεστής προσδιορισμού** (coefficient of determination) r^2 (στην περίπτωση δύο μεταβλητών) ή R^2 (σε πολλαπλή παλινδρόμηση) είναι ένα μέτρο που παρέχει συνοπτική πληροφόρηση σχετικά με το πόσο καλά προσαρμόζεται στα στοιχεία η γραμμή παλινδρόμησης του δείγματος.

2.5 Ο Συντελεστής Προσδιορισμού r^2: Ένα Μέτρο της «Καλής Προσαρμογής»

- **ΔΙΑΓΡΑΜΜΑ 2.8** Η απεικόνιση του r^2 μέσω του διαγράμματος Ballentine: (*α*) $r^2 = 0$; (*στ*) $r^2 = 1$



2.5 Ο Συντελεστής Προσδιορισμού r^2: Ένα Μέτρο της «Καλής Προσαρμογής»

- Υψώνοντας και τις δύο πλευρές της Εξίσωσης $\mathrm{Y_i} = \hat{Y}_i + \hat{u}_i$ στο τετράγωνο και αθροίζοντας σε όλο το δείγμα, παίρνουμε

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i$$

$$= \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

$$= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2$$

- εφόσον $\sum \hat{y}_i \hat{u}_i = 0$ (γιατί;) και $\hat{y}_i = \hat{\beta}_2 x_i$

2.5 Ο Συντελεστής Προσδιορισμού r^2: Ένα Μέτρο της «Καλής Προσαρμογής»

- Τα διάφορα αθροίσματα των τετραγώνων που παρουσιάζονται στην ακριβώς προηγούμενη εξίσωση μπορούν να περιγραφούν ως εξής: $\sum y_i^2 = \sum(Y_i - \bar{Y})^2$ = συνολική μεταβλητότητα των πραγματικών τιμών του *Y* από το μέσο του δείγματος, που μπορεί να ονομαστεί **συνολικό άθροισμα τετραγώνων** (total sum of squares - TSS). $\sum \hat{y}_i^2 = \sum(\hat{Y}_i - \hat{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum x_i^2$ = μεταβλητότητα των εκτιμημένων τιμών της *Y* από το μέσο της $(\bar{\hat{Y}} = \bar{Y})$,
- Συνεπώς, η Εξίσωση γίνεται:

$$TSS = ESS + RSS$$

- η συνολική μεταβλητότητα των τιμών της *Y* γύρω από τη μέση τιμή της μπορεί να χωριστεί σε δύο μέρη, το ένα οφείλεται στη γραμμή παλινδρόμησης και το άλλο σε τυχαίες δυνάμεις επειδή δε βρίσκονται όλες οι πραγματικές παρατηρήσεις *Y* επί της γραμμής παλινδρόμησης.

2.5 Ο Συντελεστής Προσδιορισμού *r*²: Ένα Μέτρο της «Καλής Προσαρμογής»

- Διαιρώντας τώρα την Εξίσωση της TSSμε το TSS και στις δύο πλευρές, παίρνουμε

$$1 = \frac{\text{ESS}}{\text{TSS}} + \frac{\text{RSS}}{\text{TSS}}$$
$$= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}$$

Ορίζουμε το r^2 ως

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\text{ESS}}{\text{TSS}}$$

- ή εναλλακτικά ως

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

2.5 Ο Συντελεστής Προσδιορισμού r^2: Ένα Μέτρο της «Καλής Προσαρμογής»

- Η ποσότητα r^2 είναι γνωστή ως **συντελεστής προσδιορισμού** (coefficient of determination) (του δείγματος) και είναι το πιο συχνά χρησιμοποιούμενο μέτρο της ποιότητας προσαρμογής της γραμμής παλινδρόμησης.
- Ο r^2 *μετρά την αναλογία ή το ποσοστό της συνολικής μεταβλητότητας της Y που εξηγείται από το υπόδειγμα παλινδρόμησης*.
- Μπορούμε να επισημάνουμε δύο ιδιότητες του r^2:

1. Πρόκειται για μία μη αρνητική ποσότητα.
2. Τα όριά της είναι $0 \leq r^2 \leq 1$.

2.5 Ο Συντελεστής Συσχέτισης

- Μια ποσότητα που συνδέεται στενά με, αλλά είναι εννοιολογικά πολύ διαφορετική από το r^2 είναι ο **συντελεστής συσχέτισης** (coefficient of correlation), ο οποίος, όπως έχει επισημανθεί προηγουμένως, είναι ένα μέτρο του βαθμού συσχέτισης μεταξύ δύο μεταβλητών. Μπορεί να υπολογιστεί είτε από

$$r = \pm\sqrt{r^2}$$

- Είτε από τον ορισμό του

$$r = \frac{\sum x_i y_i}{\sqrt{\left(\sum x_i^2\right)\left(\sum y_i^2\right)}}$$

$$= \frac{n\sum X_i Y_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{\sqrt{\left[n\sum X_i^2 - \left(\sum X_i\right)^2\right]\left[n\sum Y_i^2 - \left(\sum Y_i\right)^2\right]}}$$

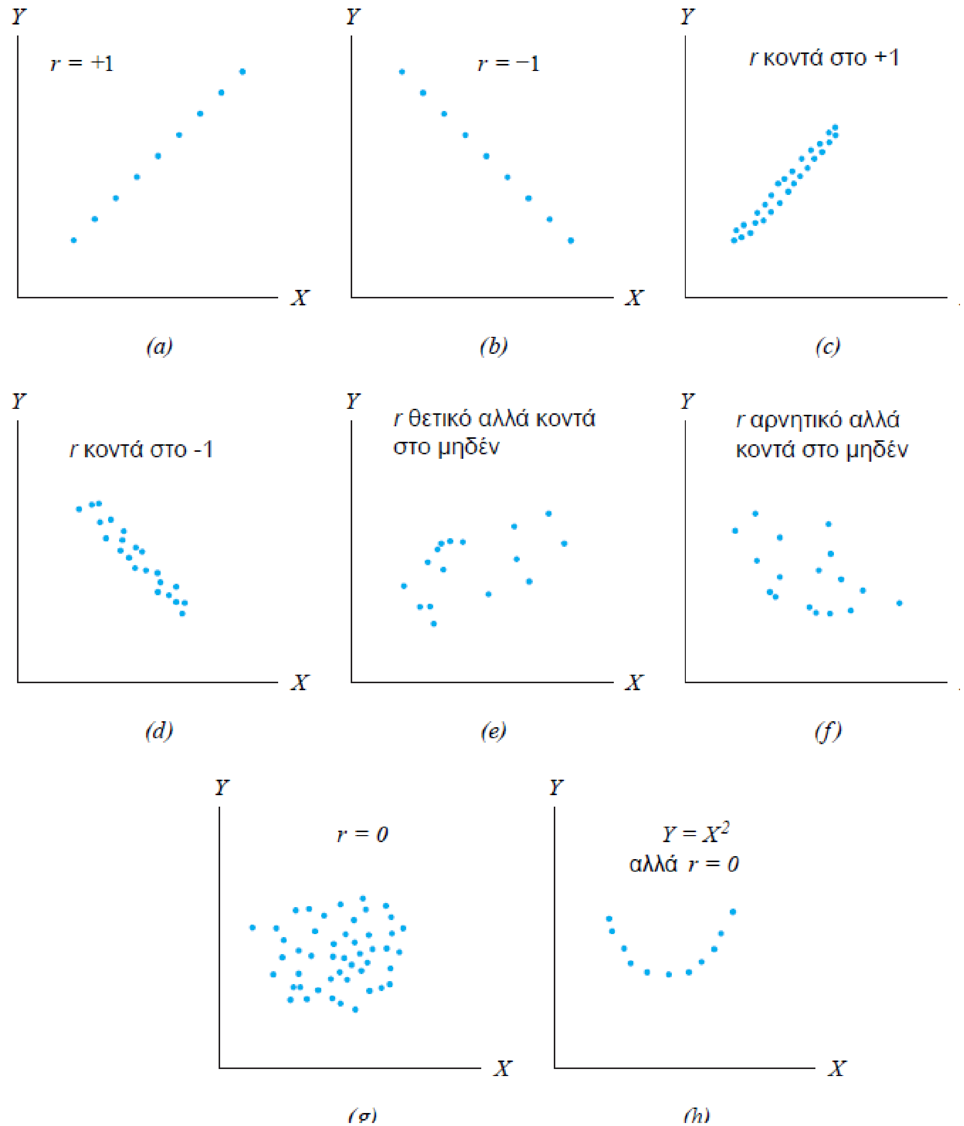
- ο οποίος είναι γνωστός ως ο **συντελεστής συσχέτισης του δείγματος** (sample correlation coefficient).

2.5 Ο Συντελεστής Συσχέτισης

- Ορισμένες από τις ιδιότητες του r έχουν ως εξής:

1. Μπορεί να είναι θετικός ή αρνητικός, το πρόσημο εξαρτάται από το πρόσημο του όρου στον αριθμητή, και μετρά τη *συμμεταβλητότητα (covariation)* στο δείγμα των δύο μεταβλητών.
2. Βρίσκεται μεταξύ των ορίων του -1 και του +1˙ δηλαδή, $-1 \leq r \leq 1$.
3. Είναι φύσει συμμετρικός˙ δηλαδή, ο συντελεστής συσχέτισης μεταξύ των X και Y (r_{XY}) είναι ο ίδιος με εκείνον μεταξύ των Y και X(r_{YX}).
4. Είναι ανεξάρτητος από το σταθερό όρο και την κλίμακα˙ δηλαδή, αν ορίζουμε $X_i^* = aXi + c$ και $Y_i^* = bY_i + d$, όπου $a > 0$, $b > 0$, και c και d είναι σταθερές, τότε ο r μεταξύ των X^* και Y^* είναι ο ίδιος με εκείνο μεταξύ των αρχικών μεταβλητών X και Y.
5. Αν οι X και Y είναι στατιστικά ανεξάρτητες, ο συντελεστής συσχέτισης μεταξύ τους είναι μηδέν˙ όμως αν $r = 0$, αυτό δε σημαίνει ότι οι δύο μεταβλητές είναι ανεξάρτητες. Με άλλα λόγια, η **μηδενική συσχέτιση δε σημαίνει απαραίτητα ανεξαρτησία**.
6. Είναι μόνο ένα μέτρο της *γραμμικής σύνδεσης* ή *γραμμικής εξάρτησης*˙ δεν έχει νόημα για την περιγραφή μη γραμμικών σχέσεων.
7. Παρόλο που είναι ένα μέτρο της γραμμικής σύνδεσης μεταξύ των δύο μεταβλητών, δε συνεπάγεται κατ’ ανάγκη κάποια σχέση αιτίας-αιτιατού.

2.5 Ο Συντελεστής Συσχέτισης



ΔΙΑΓΡΑΜΜΑ 2.10
Πρότυπα συσχέτισης (προσαρμογή από το Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, NJ, 1978, σελ. 86).

2.6 Ένα Αριθμητικό Παράδειγμα

- Από τα στοιχεία που δίνονται στον κάτωθι Πίνακα προκύπτει η ακόλουθη εκτιμημένη γραμμή παλινδρόμησης, όπου X τα έτη φοίτησης και Y το μέσο ώρομίσθιο που κερδίζουν οι άνθρωποι σε κάθε επίπεδο εκπαίδευσης:

$$\hat{Y}_i = -0{,}0144 + 0{,}7240X_i$$

Παρατηρήσεις	Y	X	x	y	x_i^2	$y_i x_i$
1	4,4567	6	−6	−4,218	36	25,308
2	5,77	7	−5	−2,9047	25	14,5235
3	5,9787	8	−4	−2,696	16	10,784
4	7,3317	9	−3	−1,343	9	4,029
5	7,3182	10	−2	−1,3565	4	2,713
6	6,5844	11	−1	−2,0903	1	2,0903
7	7,8182	12	0	−0,8565	0	0
8	7,8351	13	1	−0,8396	1	−0,8396
9	11,0223	14	2	2,3476	4	4,6952
10	10,6738	15	3	1,9991	9	5,9973
11	10,8361	16	4	2,1614	16	8,6456
12	13,615	17	5	4,9403	25	24,7015
13	13,531	18	6	4,8563	36	29,1378
Σύνολο	112,7712	156	0	0	182	131,7856

2.6 Ένα Αριθμητικό Παράδειγμα

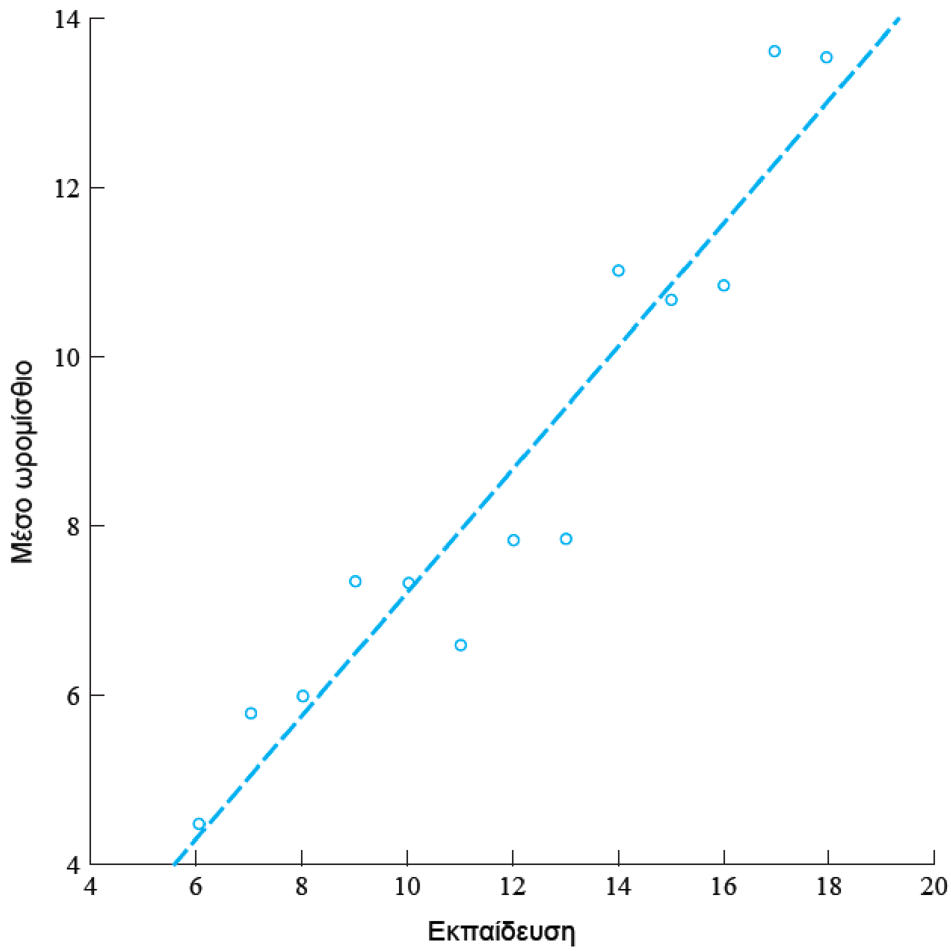
Παρατηρήσεις	X_i^2	Y_i^2	$\hat{Y}_i$	$\hat{u}_i = Y_i - \hat{Y}$	$\hat{u}_i^2$
1	36	19,86217	4,165294	0,291406	0,084917
2	49	33,2929	4,916863	0,853137	0,727843
3	64	35,74485	5,668432	0,310268	0,096266
4	81	53,75382	6,420001	0,911699	0,831195
5	100	53,55605	7,17157	0,14663	0,0215
6	121	43,35432	7,923139	−1,33874	1,792222
7	144	61,12425	8,674708	−0,85651	0,733606
8	169	61,38879	9,426277	−1,59118	2,531844
9	196	121,4911	10,17785	0,844454	0,713103
10	225	113,93	10,92941	−0,25562	0,065339
11	256	117,4211	11,68098	−0,84488	0,713829
12	289	185,3682	12,43255	1,182447	1,398181
13	324	183,088	13,18412	0,346878	0,120324
Σύνολο	2054	1083,376	112,7712	≈ 0	9,83017

2.6 Ένα Αριθμητικό Παράδειγμα

- *Σημείωση:*
- $x_i = X_i - \bar{X}$
- $y_i = Y_i = \bar{Y}$
- $\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} = \frac{131{,}7856}{182} = 0{,}7240967$
- $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 8{,}674708 - 0{,}7240967x12 = -0{,}01445$
- $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{9{,}83017}{11} = 0{,}893652$
- $\hat{\sigma}^2 = 0{,}945332$
- $\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{0{,}893652}{182} = 0{,}004910$
- $\text{se}(\hat{\beta}_2) = \sqrt{0{,}00490} = 0{,}070072$
- $r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{9{,}83017}{105{,}1188} = 0{,}9065$
- $r = \sqrt{r^2} = 0{,}9521$
- $\text{var}(\hat{\beta}_1) = \frac{\sum x_i^2}{n \sum x_i^2} = \frac{2054}{13(182)} = 0{,}868132$
- $\text{se}(\hat{\beta}_1) = \sqrt{0{,}868132} = 0{,}9317359$

2.6 Ένα Αριθμητικό Παράδειγμα

- **ΔΙΑΓΡΑΜΜΑ 2.11** Κατανομή δειγματοληψίας του εκτιμητή της OLS και εναλλακτικός εκτιμητής $\beta_2{}^*$.



2.6 Ένα Αριθμητικό Παράδειγμα

- Η τιμή $\hat{\beta}_2 = 0{,}7240$, η οποία μετρά την κλίση της γραμμής, δείχνει ότι, εντός του εύρους δείγματος της *X* μεταξύ 6 και 18 ετών εκπαίδευσης, καθώς η *X* αυξάνεται κατά 1, η εκτιμημένη αύξηση στο μέσο ωρομίσθιο είναι περίπου 72 cents. Δηλαδή, κάθε επιπλέον έτος σχολικής φοίτησης, κατά μέσο όρο, αυξάνει το ωρομίσθιο κατά περίπου 72 cents.
- Η τιμή του $\hat{\beta}_1$ = -0,0144, που είναι ο συντελεστής σταθεράς της γραμμής, δείχνει το μέσο μισθολογικό επίπεδο όταν το επίπεδο της εκπαίδευσης είναι μηδενικό. Η συγκεκριμένη, κυριολεκτική, ερμηνεία του συντελεστή σταθεράς στην προκειμένη περίπτωση δεν έχει κανένα νόημα. Πώς θα μπορούσαν να υπάρξουν αρνητικοί μισθοί;
- Η τιμή του r^2, ύψους 0,90 δείχνει ότι η εκπαίδευση εξηγεί περίπου το 90 τοις εκατό της μεταβολής στο ωρομίσθιο. Λαμβάνοντας υπόψη ότι ο r^2 μπορεί να έχει μέγιστη τιμή το 1, η γραμμή παλινδρόμησής μας παρουσιάζει πολύ καλή προσαρμογή στα στοιχεία. Ο συντελεστής συσχέτισης, r = 0,9521, δείχνει ότι οι μισθοί και η εκπαίδευση εμφανίζουν υψηλή θετική συσχέτιση.

2.7 Επεξηγηματικά Παραδείγματα

- **Παράδειγμα 2.1** *Η Σχέση Κατανάλωσης - Εισοδήματος στις Η.Π.Α., 1960-2005*

$$\hat{Y}_t = -299{,}5913 + 0{,}7218X_t$$

$$\mathrm{var}(\hat{\beta}_1) = 827{,}4195 \quad \mathrm{se}(\hat{\beta}_1) = 28{,}7649$$

$$\mathrm{var}(\hat{\beta}_2) = 0{,}0000195 \quad \mathrm{se}(\hat{\beta}_2) = 0{,}004423$$

$$r^2 = 0{,}9983 \quad \hat{\sigma}^2 = 73{,}56689$$

2.7 Επεξηγηματικά Παραδείγματα

- **ΠΑΡΑΔΕΙΓΜΑ 2.2** *Δαπάνες για Τρόφιμα στην Ινδία*

$$\widehat{\text{FoodExp}}_i = 94{,}2087 + 0{,}4368\text{TotalExp}_i$$

$$\text{var}(\hat{\beta}_1) = 2560{,}9401 \quad \text{se}(\hat{\beta}_1) = 50{,}8563$$

$$\text{var}(\hat{\beta}_2) = 0{,}0061 \quad \text{se}(\hat{\beta}_2) = 0{,}0783$$

$$r^2 = 0{,}3698 \quad \hat{\sigma}^2 = 4469{,}6913$$

2.7 Επεξηγηματικά Παραδείγματα

- **ΠΑΡΑΔΕΙΓΜΑ 2.3** *Η Ζήτηση για Κινητά Τηλέφωνα και Η/Υ σε Σχέση με το Κατά Κεφαλήν Προσωπικό Εισόδημα*

$$\hat{Y}_i = 14{,}4773 + 0{,}0022X_i$$

$$\text{se}\left(\hat{\beta}_1\right) = 6{,}1523$$

$$\text{se}\left(\hat{\beta}_2\right) = 0{,}00032$$

$$r^2 = 0{,}6023$$

2.7 Επεξηγηματικά Παραδείγματα

ΠΑΡΑΔΕΙΓΜΑ 2.3 *Η Ζήτηση για Κινητά Τηλέφωνα και Η/Υ σε Σχέση με το Κατά Κεφαλήν Προσωπικό Εισόδημα*

ΠΙΝΑΚΑΣ 2.3 Αριθμός Συνδρομητών Κινητής Τηλεφωνίας ανά Εκατό Άτομα, Αριθμός Η/Υ ανά 100 Άτομα και Κατά Κεφαλήν Εισόδημα σε Επιλεγμένες Χώρες (2003).

Χώρα	Κινητά Τηλέφωνα	Η/Υ	Κατά Κεφαλήν Εισόδημα
Αργεντινή	17,76	8,2	11410
Αυστραλία	71,95	60,18	28780
Βέλγιο	79,28	31,81	28920
Βραζιλία	26,36	7,48	7510
Βουλγαρία	46,64	5,19	75.4
Καναδάς	41,9	48,7	30040
Κίνα	21,48	2,76	4980
Κολομβία	14,13	4,93	6410
Τσεχία	96,46	17,74	15600
Εκουαδόρ	18,92	3,24	3940
Αίγυπτος	8,45	2,91	3940
Γαλλία	69,59	34,71	27640
Γερμανία	78,52	48,47	27610
Ελλάδα	90,23	8,17	19900
Γουατεμάλα	13,15	1,44	4090
Ουγγαρία	76,88	10,84	13840
Ινδία	2,47	0,72	2880
Ινδονησία	8,74	1,19	3210
Ιταλία	101,76	23,07	26,830
Ιαπωνία	67,9	38,22	28450
Μεξικό	29,47	8,3	8980
Ολλανδία	76,76	46,66	28560
Πακιστάν	1,75	0,42	2040
Πολωνία	45,09	14,2	11210
Ρωσία	24,93	8,87	8950
Σαουδική Αραβία	32,11	13,67	13230
Νότια Αφρική	36,36	7,26	10130
Ισπανία	91,61	19,6	22150
Σουηδία	98,05	62,13	26710
Ελβετία	84,34	70,87	32220
Ταϊλάνδη	39,42	3,98	7450
Η.Β.	91,17	40,57	27690
Η.Π.Α.	54,58	65,98	37750
Βενεζουέλα	27,3	6,09	4750

2.8 Μία Σημείωση Αναφορικά με τα Πειράματα Monte Carlo

- Πώς όμως γνωρίζει κάποιος ότι η ιδιότητα BLUE ισχύει;
- Πώς μπορεί κάποιος να καταλήξει σε συγκεκριμένο συμπέρασμα σχετικά με το κατά πόσο είναι αμερόληπτοι οι εκτιμητές της OLS;
- Η απάντηση παρέχεται από τα λεγόμενα πειράματα **Monte Carlo**, τα οποία είναι ουσιαστικά πειράματα προσομοιώσεων σε Η/Υ, ή πειράματα δειγματοληψίας.
- Για να παρουσιάσουμε την κεντρική ιδέα, ας εξετάσουμε τη διμεταβλητή PRF μας:

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

2.8 Μία Σημείωση Αναφορικά με τα Πειράματα Monte Carlo

- Ένα πείραμα Monte Carlo περιλαμβάνει τα εξής στάδια:

1. Ας υποθέσουμε ότι οι πραγματικές τιμές των παραμέτρων είναι οι εξής: $\beta_1 = 20$ και $\beta_2 = 0{,}6$.
2. Επιλέγουμε το μέγεθος του δείγματος, π.χ. $n = 25$.
3. Καθορίζουμε τις τιμές της X για κάθε παρατήρηση. Συνολικά θα έχουμε 25 τιμές για τη X.
4. Ας υποθέσουμε ότι εξετάζετε ένα πίνακα τυχαίων αριθμών, επιλέγετε 25 τιμές και τις ονομάζετε u_i (πλέον τα περισσότερα στατιστικά πακέτα έχουν ενσωματωμένες «γεννήτριες» τυχαίων αριθμών).
5. Καθότι *γνωρίζουμε* τα β_1, β_2, X_i, και u_i, παίρνουμε 25 τιμές για τη Y_i. Στην πράξη, θεωρούμε ότι ο u_i ακολουθεί μία συγκεκριμένη κατανομή πιθανότητας, ας πούμε, την κανονική, με ορισμένες παραμέτρους (π.χ., τη μέση τιμή και τη διακύμανση). Αφότου καθοριστούν οι τιμές των παραμέτρων, μπορεί κανείς εύκολα να υπολογίσει το u_i με τη χρήση στατιστικών προγραμμάτων.

2.8 Μία Σημείωση Αναφορικά με τα Πειράματα Monte Carlo

- Ένα πείραμα Monte Carlo περιλαμβάνει τα εξής στάδια:

6. Χρησιμοποιώντας τώρα τις 25 τιμές που προέκυψαν για τη Y_i, τις παλινδρομούμε με τις 25 τιμές της X που επιλέχθηκαν στο βήμα 3, αποκτώντας τους $\hat{\beta}_1$ και $\hat{\beta}_2$, τους εκτιμητές των ελαχίστων τετραγώνων.
7. Ας υποθέσουμε ότι επαναλαμβάνουμε αυτό το πείραμα 99 φορές, χρησιμοποιώντας κάθε φορά τις ίδιες τιμές για τα β_1, β_2 και X_i. Ασφαλώς, οι τιμές του u_i θα διαφέρουν από πείραμα σε πείραμα. Ως εκ τούτου, θα έχουμε 100 πειράματα, τα οποία θα δημιουργούν 100 τιμές για κάθε έναν από τους β_1 και β_2. (Στην πράξη, πραγματοποιούνται πολλά αντίστοιχα πειράματα, 1000 ή και 2000.)
8. Οι μέσες τιμές αυτών των 100 εκτιμήσεων ονομάζονται $\bar{\hat{\beta}}_1$ και $\bar{\hat{\beta}}_2$.
9. Εάν αυτές οι μέσες τιμές είναι περίπου οι ίδιες με τις πραγματικές τιμές των β_1 και β_2 που υποθέσαμε στο βήμα 1, αυτό το πείραμα Monte Carlo «αποδεικνύει» ότι οι εκτιμητές των ελαχίστων τετραγώνων είναι πράγματι αμερόληπτοι. Υπενθυμίζουμε ότι σύμφωνα με το κλασικό γραμμικό υπόδειγμα παλινδρόμησης $E(\hat{\beta}_1) = \beta_1$ και $E(\hat{\beta}_2) = \beta_2$.

Περίληψη και Συμπεράσματα

1. Η βασική ιδέα πίσω από την ανάλυση παλινδρόμησης είναι η στατιστική εξάρτηση μίας μεταβλητής, της εξαρτημένης μεταβλητής, από μία ή περισσότερες άλλες μεταβλητές, τις ερμηνευτικές μεταβλητές.
2. Ο στόχος αυτής της ανάλυσης είναι η εκτίμηση και/ή η πρόβλεψη του μέσου ή της μέσης τιμής της εξαρτημένης μεταβλητής με βάση τις γνωστές ή σταθερές τιμές των ερμηνευτικών μεταβλητών.
3. Η επιτυχία της ανάλυσης παλινδρόμησης εξαρτάται από τη διαθεσιμότητα των κατάλληλων στοιχείων. Στο παρόν κεφάλαιο συζητήσαμε τη φύση, τις πηγές και τους περιορισμούς των στοιχείων που είναι γενικά διαθέσιμα για την εκάστοτε έρευνα, ειδικά στις κοινωνικές επιστήμες.
4. Σε κάθε έρευνα, ο ερευνητής πρέπει να αναφέρει με σαφήνεια τις πηγές των στοιχείων που χρησιμοποιούνται στην ανάλυση, τους ορισμούς τους, τις μεθόδους συλλογής τους, και τυχόν κενά ή παραλείψεις στα στοιχεία καθώς και τυχόν αναθεωρήσεις στα στοιχεία. Τα μακροοικονομικά στοιχεία που δημοσιεύονται από την κυβέρνηση συχνά αναθεωρούνται.

Περίληψη και Συμπεράσματα

1. Το βασικό πλαίσιο της ανάλυσης παλινδρόμησης είναι το κλασικό γραμμικό υπόδειγμα παλινδρόμησης.
2. Το κλασικό γραμμικό υπόδειγμα παλινδρόμησης βασίζεται σε μία σειρά υποθέσεων.
3. Με βάση αυτές τις υποθέσεις, οι εκτιμητές των ελαχίστων τετραγώνων έχουν ορισμένες ιδιότητες οι οποίες συνοψίζονται στο θεώρημα Gauss-Markov, το οποίο αναφέρει ότι στην τάξη των γραμμικών αμερόληπτων εκτιμητών, οι εκτιμητές των ελαχίστων τετραγώνων έχουν ελάχιστη διακύμανση. Με λίγα λόγια, είναι BLUE.
4. Η *ακρίβεια* των εκτιμητών της OLS μετράται μέσω των τυπικών τους σφαλμάτων. Στο Κεφάλαιο 4 και 5, θα δούμε πώς τα τυπικά σφάλματα επιτρέπουν την εξαγωγή συμπερασμάτων για τις παραμέτρους του πληθυσμού, τους συντελεστές β.

Περίληψη και Συμπεράσματα

5. Η συνολική καλή προσαρμογή του υποδείγματος παλινδρόμησης μετράται με το **συντελεστή προσδιορισμού**, r^2. Καταδεικνύει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής, ή παλινδρομούμενης μεταβλητής, που εξηγείται από την ερμηνευτική μεταβλητή, ή παλινδρομητή. Το r^2 βρίσκεται μεταξύ του 0 και του 1˙ όσο πιο κοντά βρίσκεται στο 1, τόσο καλύτερη είναι η προσαρμογή.
6. Μία έννοια που σχετίζεται με το συντελεστή προσδιορισμού είναι ο **συντελεστής συσχέτισης**, r. Πρόκειται για ένα μέτρο της *γραμμικής σχέσης* μεταξύ δύο μεταβλητών και βρίσκεται μεταξύ του -1 και του +1.
7. Το κλασικό γραμμικό υπόδειγμα παλινδρόμησης είναι ένα θεωρητικό κατασκεύασμα επειδή βασίζεται σε ένα σύνολο υποθέσεων που μπορεί να είναι αυστηρές ή «μη ρεαλιστικές». Όμως κάτι τέτοιο είναι συχνά αναγκαίο στα πρώτα στάδια της μελέτης του οποιουδήποτε επιστημονικού κλάδου. Μόλις το κλασικό γραμμικό υπόδειγμα παλινδρόμησης καταστεί απολύτως κατανοητό, μπορεί κανείς να ανακαλύψει τι θα συμβεί εάν μία ή περισσότερες από τις υποθέσεις του δεν ικανοποιούνται.

Παράρτημα 2Α

- **2Α.1 Υπολογισμός των Εκτιμήσεων των Ελαχίστων Τετραγώνων**
- Παραγωγίζοντας μερικώς την Εξίσωση που μας δίνει το εκτιμημένο $\hat{\beta}_2$ ως προς τα $\hat{\beta}_1$ και $\hat{\beta}_2$ παίρνουμε

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_1} = -2\sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) = -2\sum \hat{u}_i$$

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_2} = -2\sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)X_i = -2\sum \hat{u}_i X_i$$

Παράρτημα 2Α

- **2Α.2 Γραμμικότητα και Ιδιότητες Αμεροληψίας των Εκτιμητών των Ελάχιστων Τετραγώνων**
- Από την Εξίσωση που μας δίνει το εκτιμημένο $\hat{\beta}_2$ έχουμε

$$\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i Y_i$$

όπου

$$k_i = \frac{x_i}{\left(\sum x_i^2\right)}$$

Παράρτημα 2Α

- **2Α.2 Γραμμικότητα και Ιδιότητες Αμεροληψίας των Εκτιμητών των Ελάχιστων Τετραγώνων**
- Επισημαίνονται οι παρακάτω ιδιότητες των σταθμών k_i:
- Εφόσον οι X_i θεωρούνται μη-στοχαστικές, τα k_i είναι επίσης μη-στοχαστικά.

$$\sum k_i = 0$$

$$\sum k_i^2 = {}^{1}\!/_{\sum x_i^2}$$

$$\sum k_i x_i = \sum k_i X_i = 1$$

Αυτές οι ιδιότητες επιβεβαιώνονται από τον ορισμό του k_i.
Για παράδειγμα,

$$\sum k_i = \sum \left(\frac{x_i}{\sum x_i^2} \right) = \frac{1}{\sum x_i^2} \sum x_i$$

καθώς για ένα δείγμα το $\sum x_i^2$ είναι γνωστό
= 0
καθώς το $\sum x_i^2$, το άθροισμα των αποκλίσεων από τη μέση τιμή είναι πάντα μηδέν.

Παράρτημα 2Α

- **2Α.2 Γραμμικότητα και Ιδιότητες Αμεροληψίας των Εκτιμητών των Ελάχιστων Τετραγώνων**
- Αντικαθιστούμε την PRF $Y_i = \beta_1 + \beta_2 X_i + u_i$ στην Εξίσωση (3) και παίρνουμε

$$\hat{\beta}_2 = \sum k_i \left(\beta_1 + \beta_2 X_i + u_i\right)$$

$$= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i$$

$$= \beta_2 + \sum k_i u_i$$

- όπου χρησιμοποιούμε τις ιδιότητες του k_i όπως τις επισημάναμε νωρίτερα.
- Χρησιμοποιώντας τώρα την προσδοκία της Εξίσωσης (4) και στις δύο πλευρές και επισημαίνοντας ότι τα k_i, δεδομένου ότι είναι μη-στοχαστικά, μπορούν να θεωρηθούν σταθερά, παίρνουμε

$$E\left(\hat{\beta}_2\right) = \beta_2 + \sum k_i \, E(u_i) = \beta_2$$

- εφόσον εξορισμού $E(u_i) = 0$. Επομένως, ο $\hat{\beta}_2$ είναι ένας αμερόληπτος εκτιμητής του β_2. Ομοίως μπορεί να αποδειχθεί ότι ο $\hat{\beta}_1$ είναι επίσης ένας αμερόληπτος εκτιμητής του β_1.

Παράρτημα 2Α

- **2Α.3 Διακυμάνσεις και Τυπικά Σφάλματα των Εκτιμητών των Ελάχιστων Τετραγώνων**
- Με βάση τον ορισμό της διακύμανσης, μπορούμε να γράψουμε

$$\operatorname{var}(\hat{\beta}_2) = E[\hat{\beta}_2 - E(\hat{\beta}_2)]^2$$

$= E(\hat{\beta}_2 - \beta_2)^2$ καθώς $E(\hat{\beta}_2) = \beta_2$

$$= E\left(\sum k_i u_i\right)^2$$

- χρησιμοποιώντας την παραπάνω Εξ. (4)

$$= E(k_1^2 u_1^2 + k_2^2 u_2^2 + \cdots + k_n^2 u_n^2 + 2k_1 k_2 u_1 u_2 + \cdots + 2k_{n-1} k_n u_{n-1} u_n)$$

-
- Καθώς εξ' ορισμού, $E(u_i^2) = \sigma^2$ για κάθε *i* και $E(u_i u_j) = 0$, *i* ≠ *j*, προκύπτει ότι

$$\operatorname{var}(\hat{\beta}_2) = \sigma^2 \sum k_i^2 = \frac{\sigma^2}{\sum x_i^2} = \text{Εξ. } (3.3.1)$$

- Η διακύμανση του $\hat{\beta}_1$ προκύπτει εάν ακολουθήσουμε την ίδια λογική με προηγουμένως. Μόλις υπολογιστούν οι διακυμάνσεις των $\hat{\beta}_1$ και $\hat{\beta}_2$, οι θετικές τετραγωνικές τους ρίζες δίνουν τα αντίστοιχα τυπικά σφάλματα.

Παράρτημα 2A

- **2A.4 Συνδιακύμανση μεταξύ $\widehat{\boldsymbol{\beta}}_1$ και $\widehat{\boldsymbol{\beta}}_2$**
- Εξ ορισμού,

$$\begin{aligned}\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) &= E\{[\hat{\beta}_1 - E(\hat{\beta}_1)][\hat{\beta}_2 - E(\hat{\beta}_2)]\} \\ &= E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2)\end{aligned}$$

(Γιατί;)

$$\begin{aligned}&= -\bar{X}E(\hat{\beta}_2 - \beta_2)^2 \\ &= -\bar{X}\operatorname{var}(\hat{\beta}_2) \\ &= \text{Εξ. }(3.3.9)\end{aligned}$$

όπου χρησιμοποιούμε τις σχέσεις $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$ και $E(\hat{\beta}_1) = \bar{Y} - \beta_2\bar{X}$, και προκύπτει

$\hat{\beta}_1 - E(\hat{\beta}_1) = -\bar{X}(\hat{\beta}_2 - \beta_2)$. *Σημείωση:* η $\operatorname{var}(\hat{\beta}_2)$ προκύπτει από την Εξ. (3.3.1).

Παράρτημα 2Α

- **2Α.5 Ο Εκτιμητής των Ελαχίστων Τετραγώνων της σ^2**
- Υπενθυμίζεται ότι

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

Επομένως,

$$\bar{Y} = \beta_1 + \beta_2 \bar{X} + \bar{u}$$

Αφαιρώντας την Εξίσωση (10) από την Εξίσωση (9) παίρνουμε

$$y_i = \beta_2 x_i + (u_i - \bar{u})$$

Επίσης υπενθυμίζεται ότι

$$\hat{u}_i = y_i - \hat{\beta}_2 x_i$$

Επομένως, αντικαθιστώντας την Εξίσωση (11) στην Εξίσωση (12) έχουμε

$$\hat{u}_i = \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i$$

Παράρτημα 2Α

- **2Α.5 Ο Εκτιμητής των Ελαχίστων Τετραγώνων της σ^2**
- Τετραγωνίζοντας και προσθέτοντας και στα δύο μέλη, παίρνουμε

$$\sum \hat{u}_i^2 = \left(\hat{\beta}_2 - \beta_2\right)^2 \sum x_i^2 + \sum (u_i - \bar{u})^2 - 2\left(\hat{\beta}_2 - \beta_2\right) \sum x_i \, (u_i - \bar{u})$$

Παίρνοντας τις προσδοκίες και στα δύο μέλη προκύπτει

$$E\left(\sum \hat{u}_i^2\right) = \sum x_i^2 \, E\left(\hat{\beta}_2 - \beta_2\right)^2 + E\left[\sum (u_i - \bar{u})^2\right] - 2E\left[\left(\hat{\beta}_2 - \beta_2\right) \sum x_i \, (u_i - \bar{u})\right]$$

$$= \sum x_i^2 \operatorname{var}\left(\hat{\beta}_2\right) + (n-1)\operatorname{var}(u_i) - 2E\left[\sum k_i u_i (x_i u_i)\right]$$

$$= \sigma^2 + (n-1)\sigma^2 - 2E\left[\sum k_i x_i u_i^2\right]$$

$$= \sigma^2 + (n-1)\sigma^2 - 2\sigma^2$$

$$= (n-2)\sigma^2$$

Παράρτημα 2Α

- **2Α.5 Ο Εκτιμητής των Ελαχίστων Τετραγώνων της σ^2**
- Στο προτελευταίο βήμα γίνεται χρήση του ορισμού του k_i που δίνεται στην Εξ. (3) και στη σχέση που δίνεται στην Εξ. (4). Επισημαίνεται επίσης ότι

$$
\begin{aligned}
E\sum(u_i-\bar{u})^2 &= E\left[\sum u_i^2 - n\bar{u}^2\right] \\
&= E\left[\sum u_i^2 - n\left(\frac{\sum u_i}{n}\right)^2\right] \\
&= E\left[\sum u_i^2 - \frac{1}{n}\sum\left(u_i^2\right)\right] \\
&= n\sigma^2 - \frac{n}{n}\sigma^2 = (n-1)\sigma^2
\end{aligned}
$$

όπου γίνεται χρήση του γεγονότος ότι οι u_i είναι ασυσχέτιστοι και η διακύμανση κάθε u_i είναι σ^2.

Συνεπώς, παίρνουμε

$$E\left(\sum \hat{u}_i^2\right) = (n-2)\sigma^2$$

Επομένως, εάν ορίσουμε

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$$

η προσδοκώμενη τιμή είναι

$$E(\hat{\sigma}^2) = \frac{1}{n-2}E\left(\sum \hat{u}_i^2\right) = \sigma^2$$

χρησιμοποιώντας την Εξίσωση (16)

η οποία δείχνει ότι $\hat{\sigma}^2$ είναι ένας αμερόληπτος εκτιμητής του πραγματικού σ^2.

Παράρτημα 2Α

- **2Α.6 Ιδιότητα της Ελάχιστης Διακύμανσης των Εκτιμητών των Ελάχιστων Τετραγώνων**
- Όπως δείξαμε στο Παράρτημα 3Α, Ενότητα 3Α.2, ο εκτιμητής των ελαχίστων τετραγώνων $\hat{\beta}_2$ είναι γραμμικός καθώς και αμερόληπτος (αυτό ισχύει και για το $\hat{\beta}_1$). Για να δείξουμε ότι αυτοί οι εκτιμητές έχουν επίσης μηδενική διακύμανση της τάξης όλων των αμερόληπτων γραμμικών εκτιμητών, εξετάζουμε τον εκτιμητή των ελαχίστων τετραγώνων $\hat{\beta}_2$:

$$\hat{\beta}_2 = \sum k_i Y_i$$

όπου

$$k_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} = \frac{x_i}{\sum x_i^2}$$

Παράρτημα 2Α

- **2Α.6 Ιδιότητα της Ελάχιστης Διακύμανσης των Εκτιμητών των Ελάχιστων Τετραγώνων**

Ας ορίσουμε έναν εναλλακτικό γραμμικό εκτιμητή του β_2 ως ακολούθως:

$$\beta_2^* = \sum w_i Y_i$$

όπου w_i είναι επίσης σταθμά, τα οποία δεν είναι απόλυτα ίσα με το k_i. Τώρα

$$E(\beta_2^*) = \sum w_i E(Y_i)$$

$$= \sum w_i (\beta_1 + \beta_2 X_i)$$

$$= \beta_1 \sum w_i + \beta_2 \sum w_i X_i$$

Επομένως, για να είναι ο β_2^* αμερόληπτος, πρέπει να έχουμε

$$\sum w_i = 0$$

και

$$\sum w_i X_i = 1$$

Παράρτημα 2Α

- **2Α.6 Ιδιότητα της Ελάχιστης Διακύμανσης των Εκτιμητών των Ελάχιστων Τετραγώνων**
- Επίσης, μπορούμε να γράψουμε

$$\operatorname{var}(\beta_2^*) = \operatorname{var}\sum w_i X_i$$

$$= \sum w_i^2 \operatorname{var} Y_i$$

[*Σημείωση:* $\operatorname{var} Y_i = \operatorname{var} u_i = \sigma^2$]

$$= \sigma^2 \sum w_i^2$$

[*Σημείωση:* $\operatorname{cov}(Y_i, Y_j) = 0 (i \neq j)$]

$$= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2} \right)^2$$

(Επισημαίνεται το μαθηματικό τέχνασμα)

$$= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \frac{\sum x_i^2}{\left(\sum x_i^2\right)^2} + 2\sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)\left(\frac{x_i}{\sum x_i^2} \right)$$

$$= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\frac{1}{\sum x_i^2} \right)$$

Παράρτημα 2Α

- **2Α.6 Ιδιότητα της Ελάχιστης Διακύμανσης των Εκτιμητών των Ελάχιστων Τετραγώνων**
- Επίσης, μπορούμε να γράψουμε

$$\text{var}(\beta_2^*) = \text{var}\sum w_i X_i$$

$$= \sum w_i^2 \,\text{var} Y_i$$

[*Σημείωση:* $\text{var} Y_i = \text{var} u_i = \sigma^2$]

$$= \sigma^2 \sum w_i^2$$

[*Σημείωση:* $\text{cov}(Y_i, Y_j) = 0 (i \neq j)$]

$$= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2} \right)^2$$

(Επισημαίνεται το μαθηματικό τέχνασμα)

$$= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \frac{\sum x_i^2}{\left(\sum x_i^2\right)^2} + 2\sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)\left(\frac{x_i}{\sum x_i^2} \right)$$

$$= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\frac{1}{\sum x_i^2} \right)$$

Βιβλιογραφία

- Θοδωρής Παπαδόγγονας, «Ποσοτικές μέθοδοι ανάλυσης επιχειρηματικών αποφάσεων", 2η Έκδοση, Εκδόσεις Τσότρας, 2022