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

6η Εκπαιδευτική Συνάντηση:

Ανάλυση Πολλαπλής Παλινδρόμησης , Πολυσυγγραμικότητα, Ετεροσκεδαστικότητα

Αθήνα, 2024

1.1 Το Υπόδειγμα Τριών-Μεταβλητών: Συμβολισμός και Υποθέσεις

- Γενικεύοντας τη διμεταβλητή συνάρτηση παλινδρόμησης του πληθυσμού, μπορούμε να γράψουμε την PRF τριών-μεταβλητών ως

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (1.1)$$

- όπου Y είναι η εξαρτημένη μεταβλητή, X_2 και X_3 οι ερμηνευτικές μεταβλητές (ή παλινδρομητές), u ο στοχαστικός διαταρακτικός όρος, και i η i^{η} παρατήρηση˙ στην περίπτωση που τα στοιχεία είναι χρονοσειρές, ο δείκτης t θα χαρακτηρίζει την t^{η} παρατήρηση.

1.1 Το Υπόδειγμα Τριών-Μεταβλητών: Συμβολισμός και Υποθέσεις

- Ο β_1 είναι η σταθερά. Ως συνήθως, εκφράζει το μέσο ή τη μέση επίδραση στη Y του συνόλου των μεταβλητών που εξαιρούνται από το υπόδειγμα, αν και η ερμηνεία της είναι ότι αποτελεί τη μέση τιμή της Y όταν οι X_2 και X_3 ορίζονται ότι ισούνται με το μηδέν.
- Οι συντελεστές β_2 και β_3 ονομάζονται **συντελεστές μερικής παλινδρόμησης** (partial regression coefficients)

1.1 Το Υπόδειγμα Τριών-Μεταβλητών: Συμβολισμός και Υποθέσεις

- **ΥΠΟΘΕΣΕΙΣ**

1. Γραμμικό υπόδειγμα παλινδρόμησης, ή *γραμμικό ως προς τις παραμέτρους*.
2. Σταθερές τιμές της X ή τιμές της X ανεξάρτητες από τον όρο σφάλματος. Αυτό σημαίνει ότι ζητούμε μηδενική συνδιακύμανση μεταξύ u_i και κάθε μεταβλητής X.

$$\mathrm{cov}\,(u_i, X_{2i}) = \mathrm{cov}\,(u_i, X_{3i}) = 0 \quad (1.1.3)$$

3. Μηδενική μέση τιμή του διαταρακτικού όρου u_i.

$$E(u_i \mid X_{2i}, X_{3i}) = 0 \text{ για κάθε } i \quad (1.1.4)$$

1.1 Το Υπόδειγμα Τριών-Μεταβλητών: Συμβολισμός και Υποθέσεις

- **ΥΠΟΘΕΣΕΙΣ**

4. Ομοσκεδαστικότητα ή σταθερή διακύμανση του u_i.

$$\text{var}\ (u_i) = \sigma^2 \quad (1.1.5)$$

5. Απουσία αυτοσυσχέτισης, ή σειριακής συσχέτισης, μεταξύ των διαταρακτικών όρων.

$$\text{cov}\ (u_i, u_j) = 0 \quad i \neq j \quad (1.1.6)$$

6. Ο αριθμός των παρατηρήσεων *n* πρέπει να είναι μεγαλύτερος από τον αριθμό των παραμέτρων που πρέπει θα εκτιμηθούν, οι οποίες στην προκειμένη περίπτωση είναι 3. (1.1.7)

1.1 Το Υπόδειγμα Τριών-Μεταβλητών: Συμβολισμός και Υποθέσεις

- **ΥΠΟΘΕΣΕΙΣ**

7. Πρέπει να υπάρχει διακύμανση στις τιμές των μεταβλητών *X*. Θα εξετάσουμε επίσης άλλες δύο προϋποθέσεις.
8. Δεν υπάρχει ακριβής συγγραμικότητα μεταξύ των μεταβλητών *X*. Δεν υπάρχει **ακριβής γραμμική σχέση** μεταξύ X_2 και X_3
9. Δεν υπάρχει *σφάλμα εξειδίκευσης*. Το υπόδειγμα έχει προσδιοριστεί σωστά.

1.2 Ερμηνεία της Εξίσωσης Πολλαπλής Παλινδρόμησης

- Λαμβάνοντας υπόψη τις υποθέσεις του κλασικού υποδείγματος παλινδρόμησης, προκύπτει η κάτωθι εξίσωση:

$$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} \quad (1.2.1)$$

- Με άλλα λόγια, η Εξ. (1.2.1) δίνει την **υπό συνθήκη μέση ή προσδοκώμενη τιμή της *Y* η οποία εξαρτάται από τις δοθείσες ή σταθερές τιμές των** X_2 και X_3.

1.3 Η Έννοια των Συντελεστών Μερικής Παλινδρόμησης

- Όπως προαναφέρθηκε, οι συντελεστές παλινδρόμησης β_2 και β_3 είναι γνωστοί ως **συντελεστές μερικής παλινδρόμησης** ή **μερικής κλίσης**.
- Η έννοια του συντελεστή μερικής παλινδρόμησης είναι η εξής: ο β_2 μέτρα τη *μεταβολή* της μέσης τιμής της Y, $E(Y)$, ανά μονάδα μεταβολής της X_2, διατηρώντας την τιμή της X_3 σταθερή. Με άλλα λόγια, δίνει το «άμεσο» ή «καθαρό» αποτέλεσμα μίας μοναδιαίας μεταβολής στη X_2 στη μέση τιμή της Y, μετά την αφαίρεση τυχόν επιδράσεων που μπορεί να έχει η X_3 στο μέσο της Y.
- Ομοίως το β_3 μέτρα τη μεταβολή της μέσης τιμής της Y για κάθε μοναδιαία μεταβολή στη X_3, διατηρώντας την τιμή της X_2 σταθερή. Δηλαδή, δίνει το «άμεσο» ή «καθαρό» αποτέλεσμα μίας μοναδιαίας μεταβολής στη X_3 στη μέση τιμή της Y, μετά την αφαίρεση τυχόν επιδράσεων που μπορεί να έχει η X_2 στο μέσο της Y.
- β_2 και β_3 είναι οι μερικές παράγωγοι της $E(Y \mid X_2, X_3)$ ως προς τις X_2 και X_3.

1.4 Εκτίμηση OLS και ML των Συντελεστών Μερικής Παλινδρόμησης

- **Οι Εκτιμητές της OLS**
- Για να βρούμε τους εκτιμητές της OLS, γράφουμε αρχικά τη συνάρτηση παλινδρόμησης του δείγματος (SRF) που αντιστοιχεί στην PRF της Εξίσωσης (1.1.):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \quad (1.4.1)$$

- όπου $\hat{u}_i$ είναι τα κατάλοιπα, το αντίστοιχο του στοχαστικού διαταρακτικού όρου u_i.

1.4 Εκτίμηση OLS και ML των Συντελεστών Μερικής Παλινδρόμησης

- **Οι Εκτιμητές της OLS**
- Η διαδικασία OLS αποτελείται από την επιλογή των τιμών των αγνώστων παραμέτρων, ούτως ώστε το άθροισμα των τετραγώνων των καταλοίπων (RSS) $\sum \hat{u}_i^2$ να είναι όσο το δυνατόν μικρότερο. Συμβολικά,

$$\min \sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2$$

- όπου η έκφραση για το RSS προκύπτει με απλούς αλγεβρικούς χειρισμούς της εξ. (1.4.1)

1.4 Εκτίμηση OLS και ML των Συντελεστών Μερικής Παλινδρόμησης

- **Οι Εκτιμητές της OLS**
- Η πιο απλή διαδικασία για την απόκτηση των εκτιμητών που θα ελαχιστοποιήσουν την προηγούμενη Εξίσωση είναι να παραγωγίσουμε τη σχέση ως προς τους αγνώστους, να θέσουμε αυτό που θα προκύψει ίσο με το μηδέν, και να τις επιλύσουμε ταυτόχρονα.
- Η διαδικασία αυτή δίνει τις ακόλουθες *κανονικές εξισώσεις* [Εξισώσεις (1.1.4) και (1.1.5)]:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2$$

1.4 Εκτίμηση OLS και ML των Συντελεστών Μερικής Παλινδρόμησης

- $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$

- $\hat{\beta}_2 = \dfrac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i X_{3i})(\sum X_{2i} X_{3i})}{(\sum x_{2i}^2) \sum x_{3i}^2) - (\sum X_{2i} X_{3i})^2}$

- $\hat{\beta}_3 = \dfrac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i X_{2i})(\sum X_{2i} X_{3i})}{(\sum x_{2i}^2) \sum x_{3i}^2) - (\sum X_{2i} X_{3i})^2}$

1.4 Εκτίμηση OLS και ML των Συντελεστών Μερικής Παλινδρόμησης

- **Οι Εκτιμητές της OLS**
- Πρέπει να επισημάνουμε τα ακόλουθα:

1. Οι Εξισώσεις που μας δίνουν τους μερικούς συντελεστές είναι φύσει συμμετρικές επειδή η μία μπορεί να προκύψει από την άλλη με εναλλαγή των ρόλων των X_2 και X_3
2. Οι παρονομαστές αυτών των δύο εξισώσεων είναι ίδιοι
3. Η περίπτωση των τριών μεταβλητών είναι μία φυσική επέκταση της διμεταβλητής περίπτωσης.

1.4 Εκτίμηση OLS και ML των Συντελεστών Μερικής Παλινδρόμησης

- **Διακυμάνσεις και Τυπικά Σφάλματα των Εκτιμητών της OLS**

$$\mathrm{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 \left(1 - r_{23}^2\right)}$$

$$\mathrm{se}(\hat{\beta}_3) = +\sqrt{\mathrm{var}(\hat{\beta}_3)}$$

$$\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{\left(1 - r_{23}^2\right)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}}$$

- Σε όλους αυτούς τους τύπους σ^2 είναι η (ομοσκεδαστική) διακύμανση των διαταρακτικών όρων του πληθυσμού u_i.

1.4 Εκτίμηση OLS και ML των Συντελεστών Μερικής Παλινδρόμησης

- **Ιδιότητες των Εκτιμητών της OLS**

1. Η γραμμή παλινδρόμησης τριών μεταβλητών διατρέχει τους μέσους $\bar{Y}$, $\bar{X}_2$, και $\bar{X}_3$.
2. Η μέση τιμή του εκτιμημένου Y_i (=$\hat{Y}_i$) είναι ίση με τη μέση τιμή του πραγματικού Y_i.
3. $\sum u_i = \bar{\hat{u}} = 0,$
4. Τα κατάλοιπα $\hat{u}_i$ είναι ασυσχέτιστα με τις X_{2i} και X_{3i}

1.5 Ο Συντελεστής Πολλαπλού Προσδιορισμού R^2 και ο Συντελεστής Πολλαπλής Συσχέτισης R

- Στο υπόδειγμα τριών μεταβλητών θα θέλαμε να γνωρίζουμε την αναλογία της διακύμανσης στη Y που εξηγείται από κοινού από τις μεταβλητές X_2 και X_3. Η ποσότητα που δίνει αυτές τις πληροφορίες είναι γνωστή ως ο **συντελεστής πολλαπλού προσδιορισμού** (multiple coefficient of determination) και συμβολίζεται με R^{2}· εννοιολογικά είναι παρόμοιος με τον r^2.
- Για να προκύψει το R^2, μπορούμε να ακολουθήσουμε τον υπολογισμό του r^2 όπως τον παρουσιάσαμε στην προηγούμενη συνάντησή μας.

- Θυμηθείτε ότι

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned}$$

- όπου $\hat{Y}_i$ είναι η εκτιμημένη τιμή του Y_i από την εκτιμημένη καμπύλη παλινδρόμησης και είναι ένας εκτιμητής του πραγματικού $E(Y_i \mid X_{2i}, X_{3i})$. Μετατρέποντας τη σχέση σε πεζά γράμματα για να δείξουμε τις αποκλίσεις από τις μέσες τιμές, η Εξ. (1.5.1) μπορεί να γραφεί ως

$$\begin{aligned} y_i &= \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \\ &= \hat{y}_i + \hat{u}_i \end{aligned}$$

- Τετραγωνίζοντας την προηγούμενη εξίσωση και στις δύο πλευρές και αθροίζοντας τις τιμές του δείγματος, παίρνουμε

$$\begin{aligned} \sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i \\ &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \end{aligned}$$

1.5 Ο Συντελεστής Πολλαπλού Προσδιορισμού R^2 και ο Συντελεστής Πολλαπλής Συσχέτισης R

- Και $\mathrm{R}^2 = \dfrac{\widehat{\beta}_2 \sum y_i x_{2i}) - \widehat{\beta}_3 (\sum y_i X_{3i})}{(\sum y_i^2)} = \dfrac{\mathrm{ESS}}{\mathrm{TSS}}$
- Υπενθυμίζεται ότι στη διμεταβλητή περίπτωση ορίσαμε την ποσότητα *r* ως το συντελεστή συσχέτισης και αναφέραμε ότι μετρά το βαθμό της (γραμμικής) σχέσης μεταξύ δύο μεταβλητών.
- Το ανάλογο τριών ή περισσότερων μεταβλητών του *r* είναι ο συντελεστής **πολλαπλής συσχέτισης** (multiple correlation), που συμβολίζεται με το *R*, και είναι ένα μέτρο του βαθμού συσχέτισης μεταξύ της *Y* και όλων των ερμηνευτικών μεταβλητών, από κοινού.
- Παρόλο που ο *r* μπορεί να είναι θετικός ή αρνητικός, ο *R* θεωρείται πάντα ότι είναι θετικός. Ωστόσο, στην πράξη, ο *R* έχει μικρή σημασία. Το πιο σημαντικό μέγεθος είναι το R^2.

1.6 R^2 και Διορθωμένο R^2

- Μια σημαντική ιδιότητα του R^2 είναι ότι πρόκειται για μία μη φθίνουσα συνάρτηση του αριθμού των ερμηνευτικών μεταβλητών ή παλινδρομητών που υπάρχουν στο υπόδειγμα, εκτός εάν η επιπλέον μεταβλητή είναι τέλεια συγγραμμική με τους άλλους παλινδρομητές.
- Καθώς ο αριθμός των παλινδρομητών αυξάνεται, ο R^2 αυξάνεται σχεδόν σταθερά και δε μειώνεται ποτέ.

1.6 R^2 και Διορθωμένο R^2

- Για να συγκρίνουμε δύο όρους R^2, πρέπει να λάβουμε υπόψη τον αριθμό των μεταβλητών *X* που υπάρχουν στο υπόδειγμα. Αυτό μπορεί να γίνει εύκολα αν σκεφτούμε έναν εναλλακτικό συντελεστή προσδιορισμού, ο οποίος είναι ο εξής:

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 /(n-k)}{\sum y_i^2 /(n-1)}$$

- όπου *k* = ο αριθμός των παραμέτρων του υποδείγματος, *συμπεριλαμβανομένου του σταθερού όρου*.
- Ο R^2 όταν ορίζεται έτσι, είναι γνωστός ως ο **διορθωμένος *R*²**, συμβολίζεται με $\bar{R}^2$.
- Ο όρος *διορθωμένος* σημαίνει διόρθωση ως προς τους βαθμούς ελευθερίας που σχετίζονται με τα αθροίσματα των τετραγώνων που εισέρχονται: Το $\sum \hat{u}_i^2$ έχει $n - k$ βαθμούς ελευθερίας σε ένα υπόδειγμα που αφορά *k* παραμέτρους, το οποίο περιλαμβάνει το σταθερό όρο, και το $\sum y_i^2$ έχει $n - 1$ βαθμούς ελευθερίας.

1.6 R^2 και Διορθωμένο R^2

- Ποιον R^2 θα πρέπει να χρησιμοποιήσει κάποιος στην πράξη; Όπως σημειώνει ο Theil:

- … είναι καλή πρακτική να χρησιμοποιούμε το $\bar{R}^2$ και όχι το R^2 επειδή ο R^2 δίνει μία υπερβολικά αισιόδοξη εικόνα της προσαρμογής της παλινδρόμησης, ιδιαίτερα όταν ο αριθμός των ερμηνευτικών μεταβλητών δεν είναι πολύ μικρός σε σχέση με τον αριθμό των παρατηρήσεων.

1.6 R^2 και Διορθωμένο R^2

- Όμως η άποψη του Theil δεν είναι κοινά αποδεκτή, γιατί δεν έχει προσφέρει καμία γενική θεωρητική αιτιολόγηση για την «ανωτερότητα» του $\bar{R}^2$. Για παράδειγμα, ο Goldberger υποστηρίζει ότι ο ακόλουθος R^2, που μπορεί να ονομαστεί **τροποποιημένος** (modified) R^2, θα είναι εξίσου κατάλληλος:

$$\text{Τροποποιημένος } R^2 = (1 - k/n)R^2$$

- Η συμβουλή του είναι να αναφέρουμε τα R^2, n, και k και να επιλέξει ο αναγνώστης πώς θα διορθώσει το R^2, λαμβάνοντας υπόψη τα n και k.

1.6 R^2 και Διορθωμένο R^2

- Εκτός από το R^2 και το διορθωμένο R^2 ως μέτρα της καλής προσαρμογής του υποδείγματος, υπάρχουν και άλλα κριτήρια που χρησιμοποιούνται συχνά για να καθοριστεί η καταλληλότητα ενός υποδείγματος παλινδρόμησης.
- Δύο από αυτά είναι το **κριτήριο Πληροφορίας του Akaike** (Akaike's Information criterion) και τα **κριτήρια Πρόβλεψης του Amemiya** (Amemiya Prediction criteria), τα οποία χρησιμοποιούνται για την επιλογή μεταξύ ανταγωνιστικών υποδειγμάτων.

1.6 R^2 και Διορθωμένο R^2

- **Συγκρίνοντας Δύο Τιμές R^2**
- Είναι σημαντικό να σημειωθεί ότι κατά τη σύγκριση δύο υποδειγμάτων με βάση το συντελεστή προσδιορισμού, είτε διορθωμένο είτε όχι, *το μέγεθος του δείγματος n και η εξαρτημένη μεταβλητή πρέπει να είναι ίδια*˙ οι ερμηνευτικές μεταβλητές μπορούν να πάρουν οποιαδήποτε μορφή. Επομένως, για τα υποδείγματα

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

$$Y_i = a_1 + a_2 X_{2i} + a_3 X_{3i} + u_i$$

- οι όροι R^2 που έχουν υπολογιστεί δε μπορούν να συγκριθούν.

1.6 R^2 και Διορθωμένο R^2

- **Το «Παιχνίδι» της Μεγιστοποίησης του $\bar{R}^2$**
- Ολοκληρώνοντας την παρούσα ενότητα, πρέπει να γίνει μία επισήμανση: Ορισμένες φορές οι ερευνητές «παίζουν το παιχνίδι της μεγιστοποίησης του $\bar{R}^2$», δηλαδή, επιλέγουν το υπόδειγμα που δίνει τον υψηλότερο $\bar{R}^2$. Αυτό όμως μπορεί να είναι επικίνδυνο, καθώς στην ανάλυση παλινδρόμησης ο στόχος μας δεν είναι να υπάρξει υψηλό $\bar{R}^2$ *per se*, αλλά κυρίως να αποκτήσουμε αξιόπιστες εκτιμήσεις των πραγματικών συντελεστών παλινδρόμησης του πληθυσμού και να εξάγουμε στατιστικά συμπεράσματα γι’ αυτούς.

1.6 R^2 και Διορθωμένο R^2

- **Το «Παιχνίδι» της Μεγιστοποίησης του $\bar{R}^2$**
- Στην εμπειρική ανάλυση δεν είναι ασυνήθιστο να έχουμε ένα πολύ υψηλό $\bar{R}^2$ αλλά αντιθέτως να διαπιστώσουμε ότι ορισμένοι από τους συντελεστές παλινδρόμησης είτε είναι στατιστικά μη-σημαντικοί ή έχουν πρόσημα που είναι αντίθετα με τις *a priori* προσδοκίες. Ως εκ τούτου, ο ερευνητής θα πρέπει να ανησυχεί περισσότερο για τη λογική ή θεωρητική σχέση των ερμηνευτικών μεταβλητών με την εξαρτημένη μεταβλητή και τη στατιστική τους σημαντικότητα. Αν σε αυτή τη διαδικασία έχουμε έναν υψηλό $\bar{R}^2$, αυτό είναι σίγουρα καλό˙ από την άλλη πλευρά, αν ο $\bar{R}^2$ είναι χαμηλός, δε σημαίνει ότι το υπόδειγμα είναι κατ’ ανάγκην κακό.

1.9 Συντελεστές Μερικής Συσχέτισης

- Ορίζουμε
- $r_{1\,2.3}$ = συντελεστής μερικής συσχέτισης μεταξύ της Y και της X_2, διατηρώντας τη X_3 σταθερή
- $r_{1\,3.2}$ = συντελεστής μερικής συσχέτισης μεταξύ της Y και της X_3, διατηρώντας τη X_2 σταθερή
- $r_{2\,3.1}$ = συντελεστής μερικής συσχέτισης μεταξύ της X_2 και της X_3, διατηρώντας τη Y σταθερή

1.9 Συντελεστές Μερικής Συσχέτισης

- Αυτές οι μερικές συσχετίσεις μπορούν να εξαχθούν με ευκολία από τους απλούς ή μηδενικής τάξης, συντελεστές συσχέτισης ως εξής (για τις αποδείξεις, βλ. τις ασκήσεις):

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

- Οι μερικές συσχετίσεις που αναφέρονται στις Εξισώσεις (παραπάνω) ονομάζονται **συντελεστές συσχέτισης πρώτης τάξης**.

- **ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ**

Εισαγωγή

- Αναζητούμε απαντήσεις στα ακόλουθα ερωτήματα:

1. Ποια είναι η φύση της ετεροσκεδαστικότητας;
2. Ποιες είναι οι συνέπειές της;
3. Πώς μπορεί κάποιος να την εντοπίσει;
4. Ποια είναι τα μέτρα αντιμετώπισης;

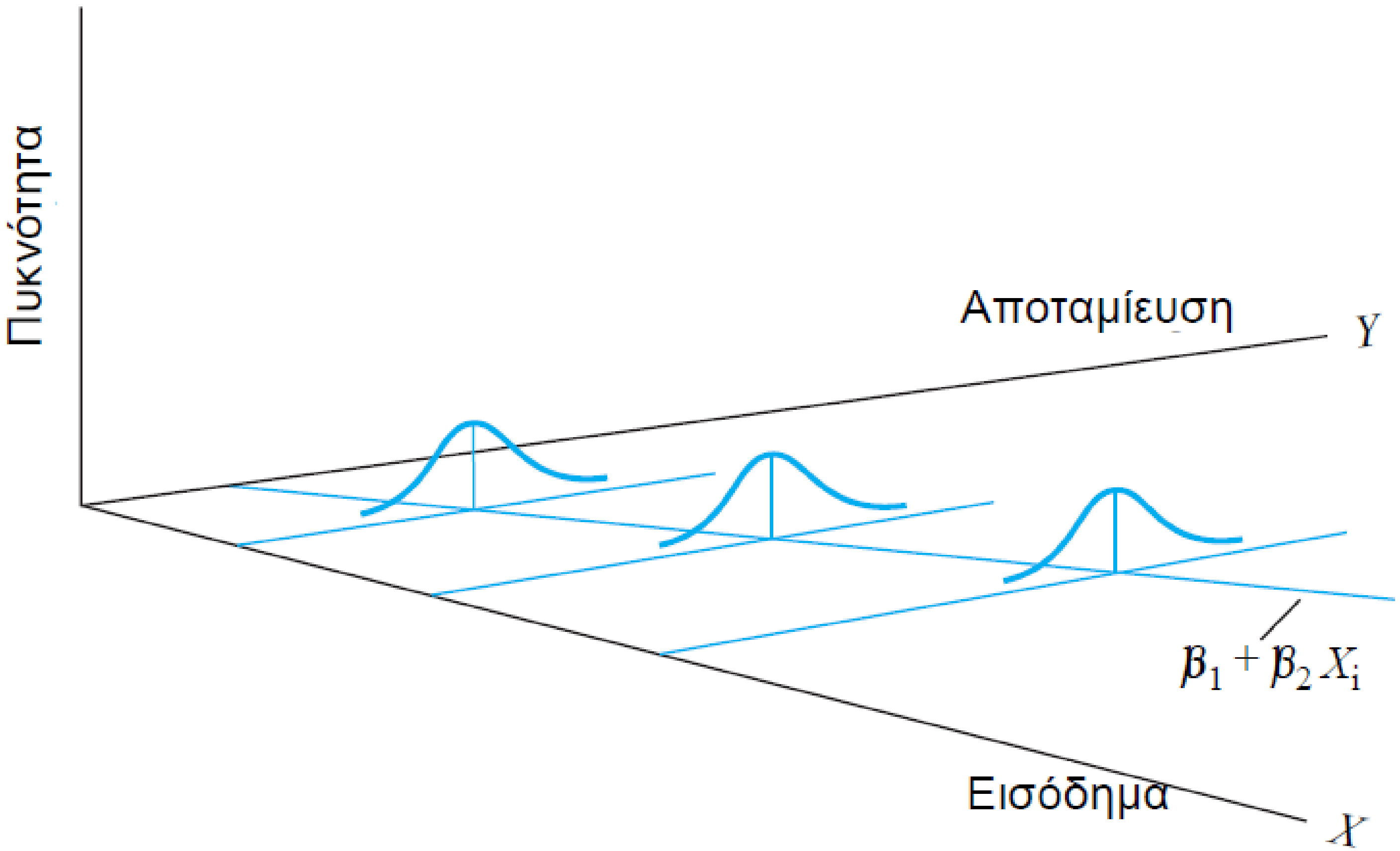
3.1 Η Φύση της Ετεροσκεδαστικότητας

- Μία από τις σημαντικότερες υποθέσεις του κλασικού γραμμικού υποδείγματος παλινδρόμησης είναι ότι η διακύμανση του κάθε διαταρακτικού όρου u_i, ο οποίος αποτελεί συνάρτηση των τιμών των ερμηνευτικών μεταβλητών, είναι κάποιος σταθερός αριθμός ίσος με σ^2. Αυτή είναι η υπόθεση της **ομοσκεδαστικότητας**, ή *ίση* (ομο) *εξάπλωση* (σκεδαστικότητα), δηλαδή, *ίση διακύμανση*. Συμβολικά,

$$E(u_i^2) = \sigma^2 \quad i = 1, 2, \ldots, n$$

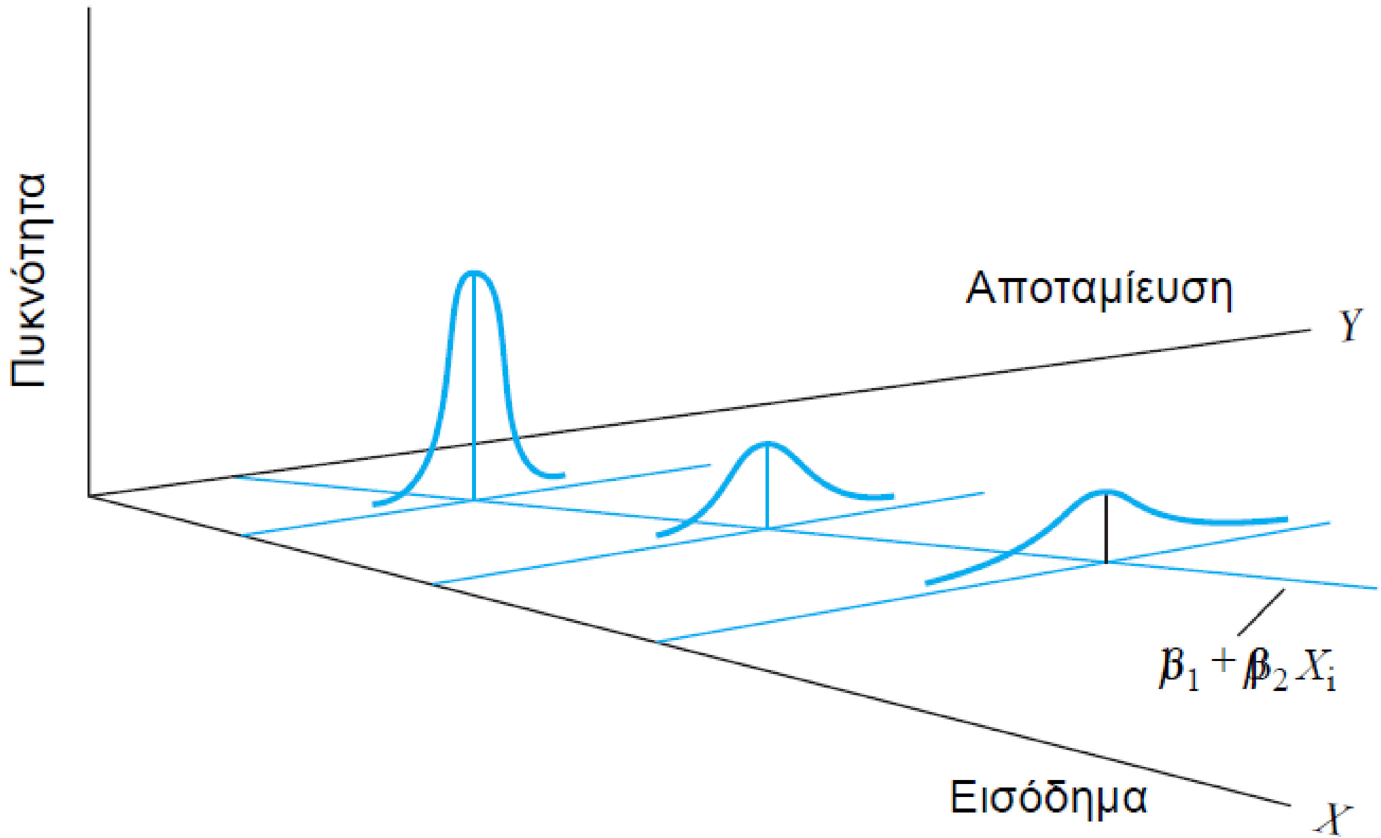
3.1 Η Φύση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.1** Ομοσκεδαστικοί διαταρακτικοί όροι



3.1 Η Φύση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.2** Ετεροσκεδαστικοί διαταρακτικοί όροι



3.1 Η Φύση της Ετεροσκεδαστικότητας

- Υπάρχουν διάφοροι λόγοι για τους οποίους οι διακυμάνσεις του u_i μπορεί να μεταβάλλονται, μερικοί εκ των οποίων είναι οι ακόλουθοι.

1. Ακολουθώντας τα *υποδείγματα σφάλματος-γνώσης*, καθώς οι άνθρωποι μαθαίνουν, τα λάθη στη συμπεριφορά τους περιορίζονται με την πάροδο του χρόνου ή ο αριθμός των σφαλμάτων τους γίνεται πιο συνεπής. Στην περίπτωση αυτή, η σ_i^2 αναμένεται να μειωθεί.

2. Καθώς το εισόδημα αυξάνεται, οι άνθρωποι έχουν περισσότερο *διαθέσιμο εισόδημα* (discretionary income) και ως εκ τούτου μεγαλύτερο περιθώριο επιλογών σχετικά με τη διάθεση του εισοδήματός τους. Ως εκ τούτου, η σ_i^2 είναι πιθανό να αυξηθεί με την αύξηση του εισοδήματος. Συνεπώς, στην παλινδρόμηση της αποταμίευσης με το εισόδημα είναι πιθανό να βρούμε ότι η σ_i^2 αυξάνεται με το εισόδημα

3.1 Η Φύση της Ετεροσκεδαστικότητας

3. Καθώς οι τεχνικές συλλογής στοιχείων βελτιώνονται, η σ_i^2 είναι πιθανό να μειώνεται. Συνεπώς, οι τράπεζες που έχουν προηγμένο εξοπλισμό για την επεξεργασία στοιχείων είναι πιθανόν να διαπράξουν λιγότερα λάθη στις μηνιαίες ή τριμηνιαίες καταστάσεις των πελατών τους σε σχέση με τράπεζες που δε διαθέτουν τον αντίστοιχο εξοπλισμό.

4. Η ετεροσκεδαστικότητα μπορεί επίσης να προκύψει ως αποτέλεσμα της παρουσίας **έκτοπων** (ακραίων τιμών). Μία παρατήρηση που αποτελεί έκτοπο, είναι μία παρατήρηση που είναι πολύ διαφορετική (είτε πολύ μικρή είτε πολύ μεγάλη) σε σχέση με τις παρατηρήσεις του δείγματος. Πιο συγκεκριμένα, ένα έκτοπο είναι μία παρατήρηση από ένα διαφορετικό πληθυσμό σε σχέση με αυτόν που παράγει τις υπόλοιπες δειγματικές παρατηρήσεις.

3.1 Η Φύση της Ετεροσκεδαστικότητας

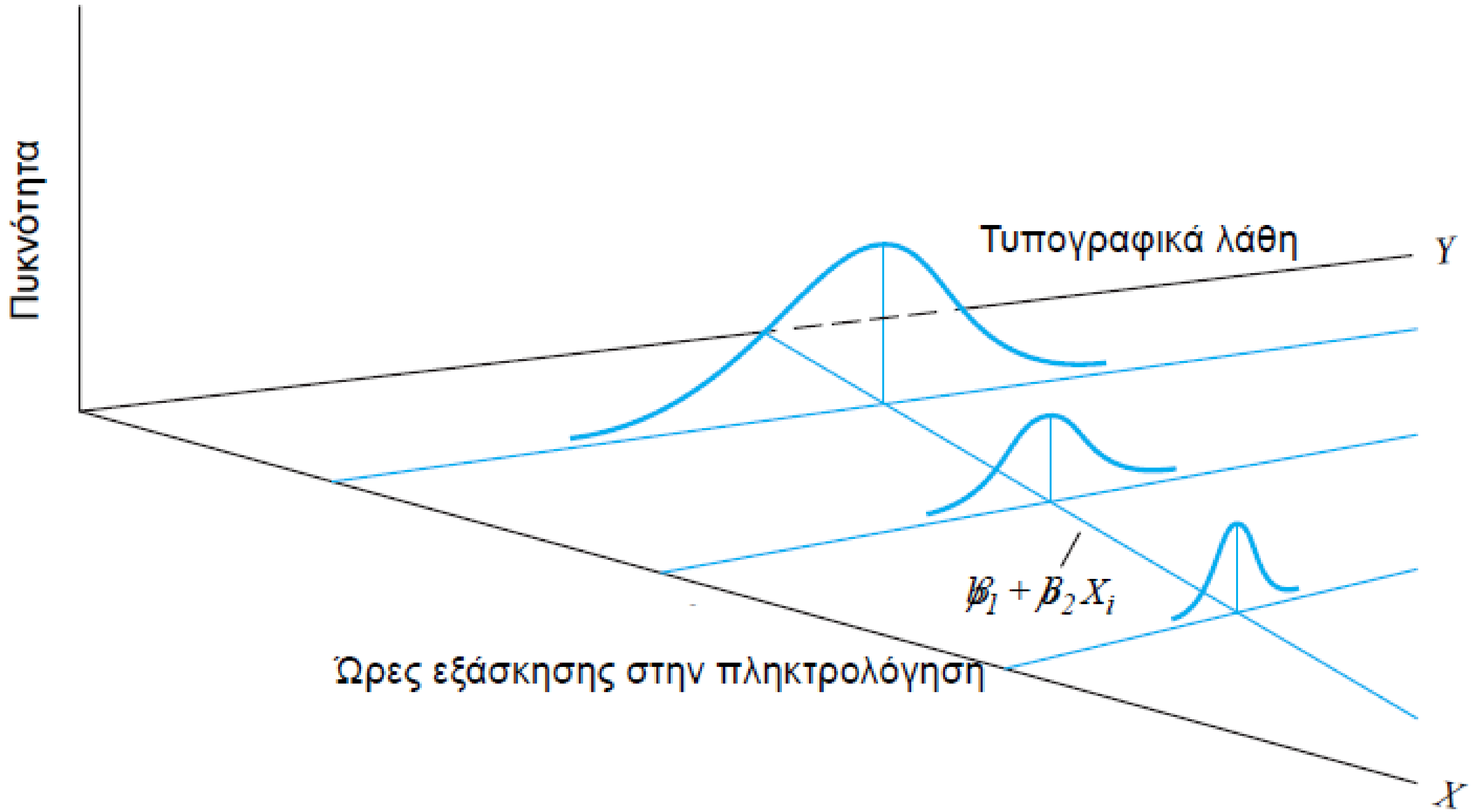
5. Μία άλλη πηγή ετεροσκεδαστικότητας προκύπτει από την παραβίαση της Υπόθεσης 9 του κλασικού γραμμικού υποδείγματος παλινδρόμησης (CLRM), δηλαδή, ότι το υπόδειγμα παλινδρόμησης παρουσιάζει τη σωστή εξειδίκευση.

6. Μία άλλη πηγή ετεροσκεδαστικότητας είναι η **ασυμμετρία** (skewness) στην κατανομή ενός ή περισσοτέρων παλινδρομητών που περιλαμβάνονται στο υπόδειγμα. Παραδείγματα είναι οι οικονομικές μεταβλητές όπως το εισόδημα, ο πλούτος, και η εκπαίδευση. Είναι γνωστό ότι η κατανομή του εισοδήματος και του πλούτου στις περισσότερες κοινωνίες είναι άνιση, με το μεγαλύτερο μέρος του εισοδήματος και του πλούτου να ανήκουν σε λίγα άτομα.

7. Άλλες πηγές ετεροσκεδαστικότητας: Όπως επισημαίνει ο David Hendry, η ετεροσκεδαστικότητα μπορεί επίσης να προκύψει λόγω (1) λανθασμένου μετασχηματισμού των στοιχείων (π.χ., μετασχηματισμοί σε λόγους ή σε πρώτες διαφορές) και (2) σε λανθασμένη συναρτησιακή μορφή (π.χ., γραμμικά έναντι λογαριθμικών-γραμμικών υποδειγμάτων).

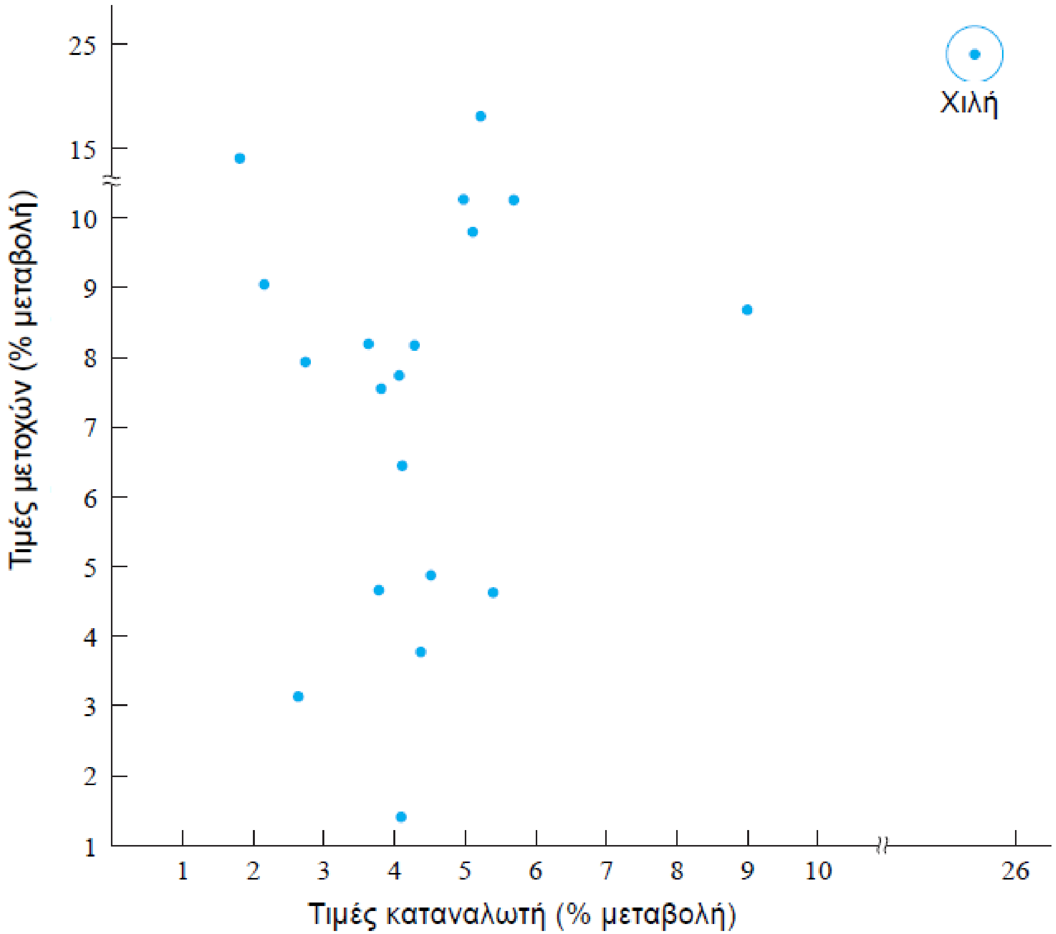
3.1 Η Φύση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.3** Διαγραμματική απεικόνιση της ετεροσκεδαστικότητας.



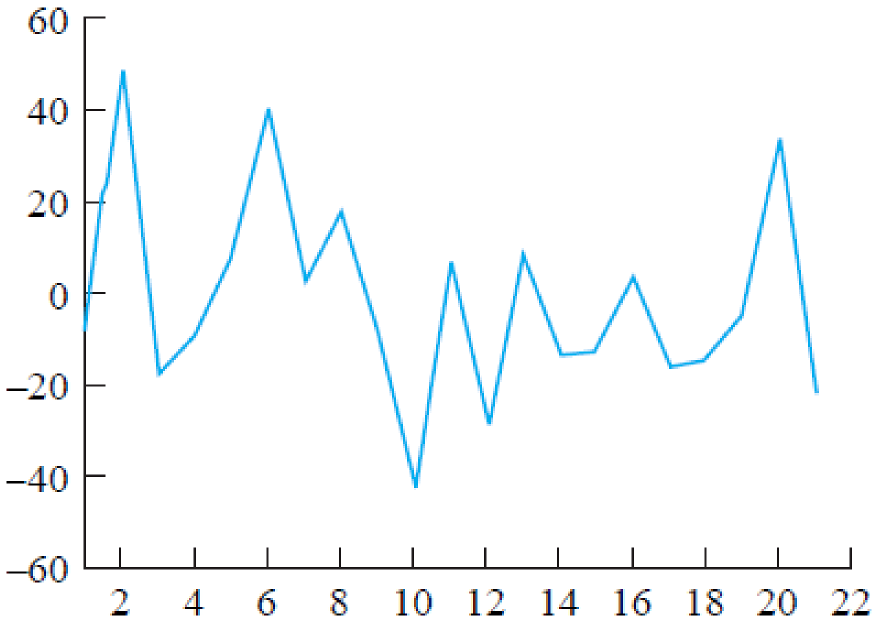
3.1 Η Φύση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.4** Η σχέση μεταξύ των τιμών των μετοχών και των τιμών καταναλωτή.

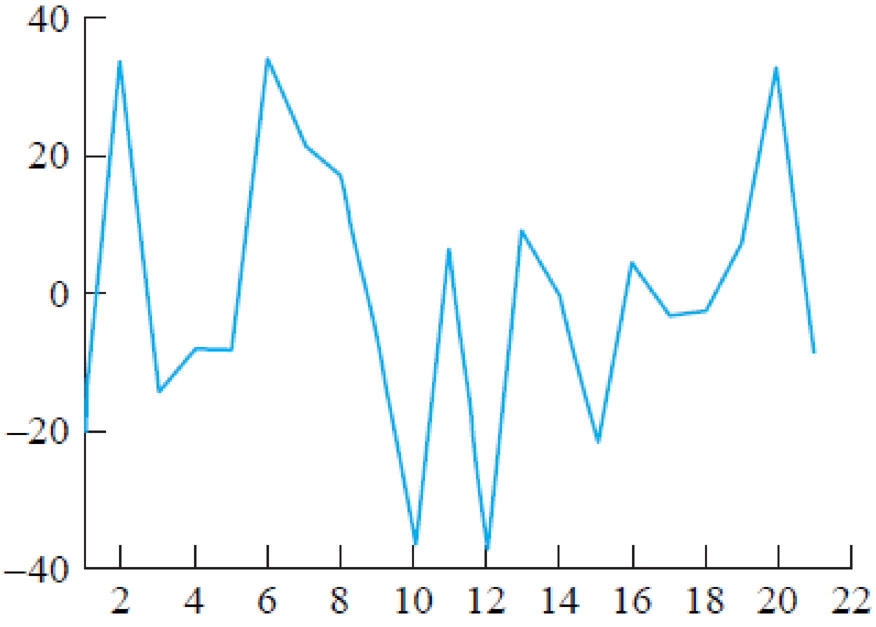


3.1 Η Φύση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.5** Κατάλοιπα από την παλινδρόμηση (α) των εμφανίσεων των διαφημίσεων με τις διαφημιστικές δαπάνες και (β) των εμφανίσεων με τις μεταβλητές Adexp και Adexp



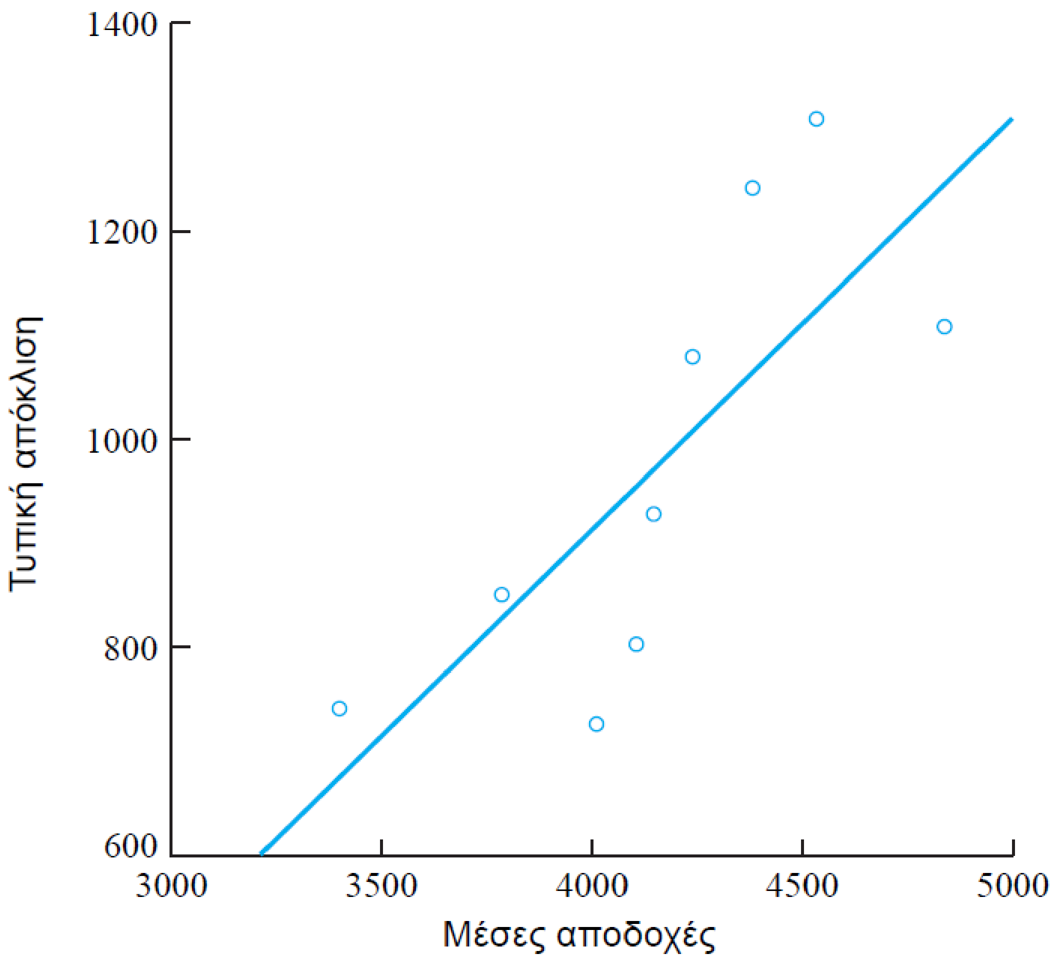
(α)



(β)

3.1 Η Φύση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.6** Τυπική απόκλιση των αποδοχών και μέσες αποδοχές.



3.2 Εκτίμηση OLS με την Παρουσία Ετεροσκεδαστικότητας

- Τι θα συμβεί στους εκτιμητές των ελαχίστων τετραγώνων (OLS) και στις διακυμάνσεις τους, εάν εισάγουμε την ετεροσκεδαστικότητα, με $E(u_i^2) = \sigma_i^2$, αλλά διατηρήσουμε όλες τις άλλες υποθέσεις του κλασικού υποδείγματος; Για να απαντήσουμε στο ερώτημα αυτό, ας επανέλθουμε στο διμεταβλητό υπόδειγμα:

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

- Εφαρμόζοντας το συνήθη τύπο, ο εκτιμητής της OLS του β_2 είναι

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

- αλλά η διακύμανσή του δίνεται τώρα από τον ακόλουθο τύπο :

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

- που είναι προφανώς διαφορετικός από το συνηθισμένο τύπο της διακύμανσης που προκύπτει υπό την υπόθεση της ομοσκεδαστικότητας, δηλαδή,

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

- Ασφαλώς, εάν $\sigma_i^2 = \sigma^2$ για κάθε i, οι δύο τύποι θα είναι ίδιοι.

3.3 Η Γενικευμένη Μέθοδος των Ελαχίστων Τετραγώνων (GLS)

- Δυστυχώς, η συνήθης μέθοδος της OLS δεν ακολουθεί αυτή τη στρατηγική και, συνεπώς, δε χρησιμοποιεί τις «πληροφορίες» που περιέχονται στην άνιση μεταβλητότητα της εξαρτημένης μεταβλητής *Y*. Προσδιορίζει την ίδια βαρύτητα ή σημασία σε κάθε παρατήρηση.
- Όμως, μία μέθοδος εκτίμησης, που είναι γνωστή ως **γενικευμένη μέθοδος των ελαχίστων τετραγώνων** (generalized least squares - GLS), λαμβάνει τις πληροφορίες αυτές υπόψη ρητά και, συνεπώς, έχει τη δυνατότητα να οδηγήσει σε εκτιμητές που είναι BLUE.

3.3 Η Γενικευμένη Μέθοδος των Ελαχίστων Τετραγώνων (GLS)

- **Διαφορά μεταξύ OLS και GLS**
- Υπενθυμίζεται ότι στην OLS ελαχιστοποιούμε

$$\sum \hat{u}_i^2 = \sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2 \quad (3.3.1)$$

- αλλά στη GLS μπορεί επίσης να γραφεί ως

$$\sum w_i \hat{u}_i^2 = \sum w_i \left(Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i\right)^2 \quad (3.3.2)$$

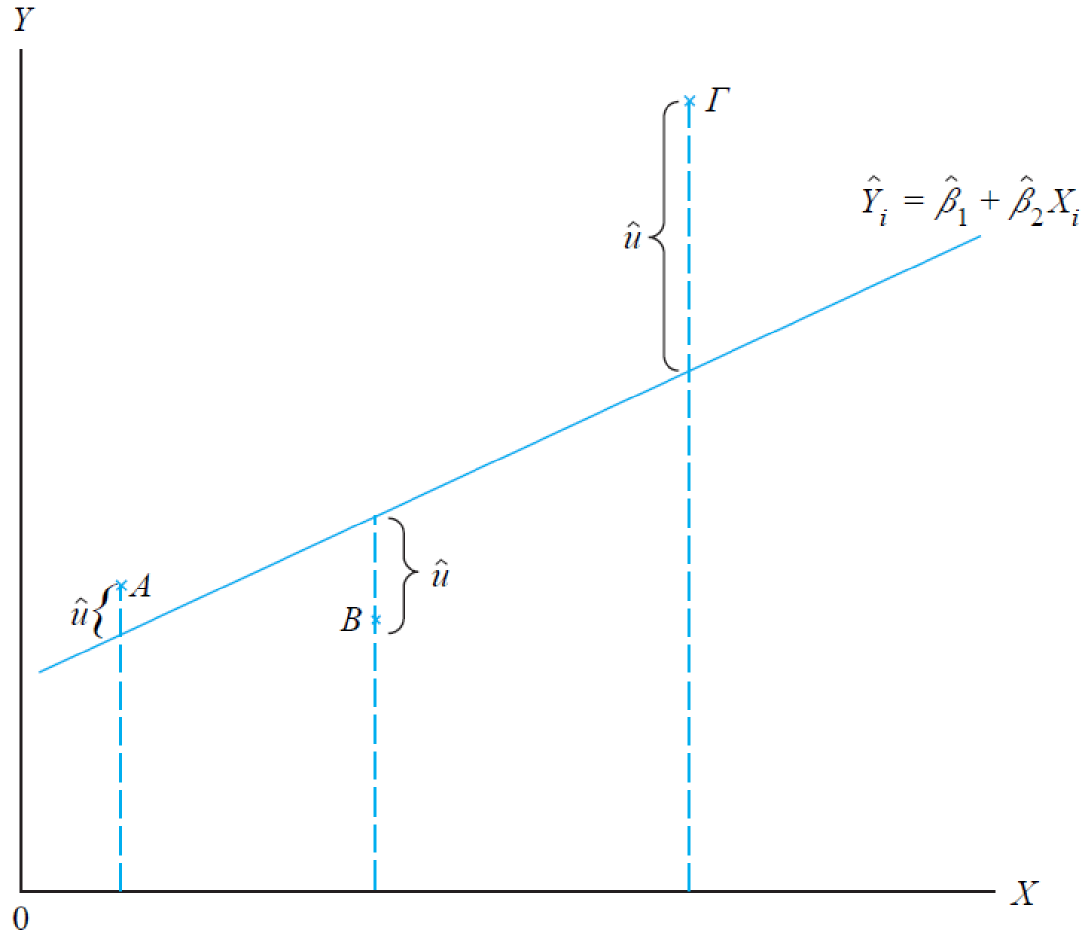
- όπου $w_i = 1/\sigma_i^2$.
- Επομένως, στη GLS ελαχιστοποιούμε ένα *σταθμισμένο άθροισμα των καταλοίπων τετραγώνων* με $w_i = 1/\sigma_i^2$ να είναι τα σταθμά, αλλά στην OLS ελαχιστοποιούμε ένα μη σταθμισμένο ή (αυτό που ισοδυναμεί με το ίδιο) εξίσου σταθμισμένο άθροισμα των τετραγώνων των καταλοίπων (RSS).

3.3 Η Γενικευμένη Μέθοδος των Ελαχίστων Τετραγώνων (GLS)

- Δεδομένου ότι η ανωτέρω Εξ. ελαχιστοποιεί ένα σταθμισμένο RSS, είναι γνωστή επίσης ως **σταθμισμένη μέθοδος των ελαχίστων τετραγώνων** (weighted least squares - WLS), και οι εκτιμητές που έχουν είναι γνωστοί ως **σταθμισμένοι εκτιμητές των ελαχίστων τετραγώνων** (εκτιμητές της WLS).
- Αλλά η WLS είναι απλά μία ειδική περίπτωση της γενικότερης τεχνικής εκτίμησης, GLS. Στο πλαίσιο της ετεροσκεδαστικότητας, κάποιος μπορεί να αντιμετωπίζει τους δύο όρους WLS και GLS ως ταυτόσημους.

3.3 Η Γενικευμένη Μέθοδος των Ελαχίστων Τετραγώνων (GLS)

- **ΔΙΑΓΡΑΜΜΑ 3.7** Υποθετικό διάγραμμα διασποράς.



3.4 Συνέπειες της Χρήσης της OLS με την Παρουσία Ετεροσκεδαστικότητας

- Όπως είδαμε, και ο $\hat{\beta}_2^*$ και ο $\hat{\beta}_2$ είναι (γραμμικοί) αμερόληπτοι εκτιμητές: Στην επαναλαμβανόμενη δειγματοληψία, κατά μέσο όρο, οι $\hat{\beta}_2^*$ και $\hat{\beta}_2$ θα ισούνται με τον πραγματικό β_2˙ δηλαδή, είναι και οι δύο αμερόληπτοι εκτιμητές. Αλλά ξέρουμε ότι είναι ο $\hat{\beta}_2^*$ που είναι αποτελεσματικός, δηλαδή, έχει τη μικρότερη διακύμανση. Τι συμβαίνει με το διάστημα εμπιστοσύνης, τον έλεγχο υποθέσεων, και άλλες διαδικασίες εάν συνεχίσουμε να χρησιμοποιούμε τον εκτιμητή της OLS $\hat{\beta}_2$; Διακρίνουμε δύο περιπτώσεις.

1. Εκτίμηση OLS με την Παρουσία Ετεροσκεδαστικότητας
2. Εκτίμηση OLS Χωρίς να Λαμβάνεται Υπόψη η Ετεροσκεδαστικότηταμε την Παρουσία Ετεροσκεδαστικότητας

3.4 Συνέπειες της Χρήσης της OLS με την Παρουσία Ετεροσκεδαστικότητας

- **Εκτίμηση OLS με την Παρουσία Ετεροσκεδαστικότητας**
- Οι έλεγχοι *t* και *F* είναι πιθανό να μας δώσουν ανακριβή αποτελέσματα σε ότι αφορά το γεγονός ότι η $\mathrm{var}(\hat{\beta}_2)$ είναι υπερβολικά μεγάλη και αυτό που φαίνεται να είναι ένας στατιστικά μη σημαντικός συντελεστής (επειδή η τιμή *t* είναι μικρότερη από ότι επιτρέπεται) μπορεί στην πραγματικότητα να είναι σημαντικός εάν είχαν οριστεί τα σωστά διαστήματα εμπιστοσύνης με βάση τη διαδικασία GLS.

3.4 Συνέπειες της Χρήσης της OLS με την Παρουσία Ετεροσκεδαστικότητας

- **Εκτίμηση OLS Χωρίς να Λαμβάνεται Υπόψη η Ετεροσκεδαστικότητα**
- Με λίγα λόγια, αν επιμένουμε στη χρήση των συνηθισμένων διαδικασιών ελέγχου παρά την ύπαρξη ετεροσκεδαστικότητας, οποιαδήποτε είναι τα αποτελέσματα ή τα συμπεράσματα στα οποία θα καταλήξουμε αυτά μπορεί να είναι πολύ παραπλανητικά.

3.4 Συνέπειες της Χρήσης της OLS με την Παρουσία Ετεροσκεδαστικότητας

- **Εκτίμηση OLS Χωρίς να Λαμβάνεται Υπόψη η Ετεροσκεδαστικότητα**

- *Το πιο εντυπωσιακό χαρακτηριστικό αυτών των αποτελεσμάτων είναι ότι η OLS, με ή χωρίς διόρθωση για ετεροσκεδαστικότητα, συστηματικά υπερεκτιμά το πραγματικό τυπικό σφάλμα που προκύπτει από τη (σωστή) διαδικασία GLS, ειδικά για τις μεγάλες τιμές του α, αποδεικνύοντας έτσι την ανωτερότητα της GLS*. Αυτά τα αποτελέσματα δείχνουν επίσης ότι αν δε χρησιμοποιήσουμε τη GLS και βασιστούμε στην OLS-επιτρέποντας ή όχι την ύπαρξη ετεροσκεδαστικότητας-η εικόνα είναι σύνθετη.
- Τα συνήθη τυπικά σφάλματα της OLS είναι είτε πολύ μεγάλα (για το συντελεστή σταθεράς) ή *γενικά* πολύ μικρά (για το συντελεστή κλίσης) σε σχέση με εκείνα που προκύπτουν από την OLS με την παρουσία ετεροσκεδαστικότητας.
- Το μήνυμα είναι σαφές: Όταν υπάρχει ετεροσκεδαστικότητα, χρησιμοποιούμε GLS. Ωστόσο, για λόγους που θα αναλυθούν αργότερα δεν είναι πάντα πρακτικά εύκολο να εφαρμοστεί η GLS.

3.4 Συνέπειες της Χρήσης της OLS με την Παρουσία Ετεροσκεδαστικότητας

- **Μια Τεχνική Σημείωση**
- Παρόλο που όπως έχουμε δηλώσει, σε περιπτώσεις ετεροσκεδαστικότητας, είναι η GLS και όχι η OLS, που είναι BLUE, υπάρχουν παραδείγματα όπου η OLS μπορεί να είναι BLUE, παρά την ετεροσκεδαστικότητα.
- Τέτοια παραδείγματα είναι σπάνια.

3.5 Ανίχνευση της Ετεροσκεδαστικότητας

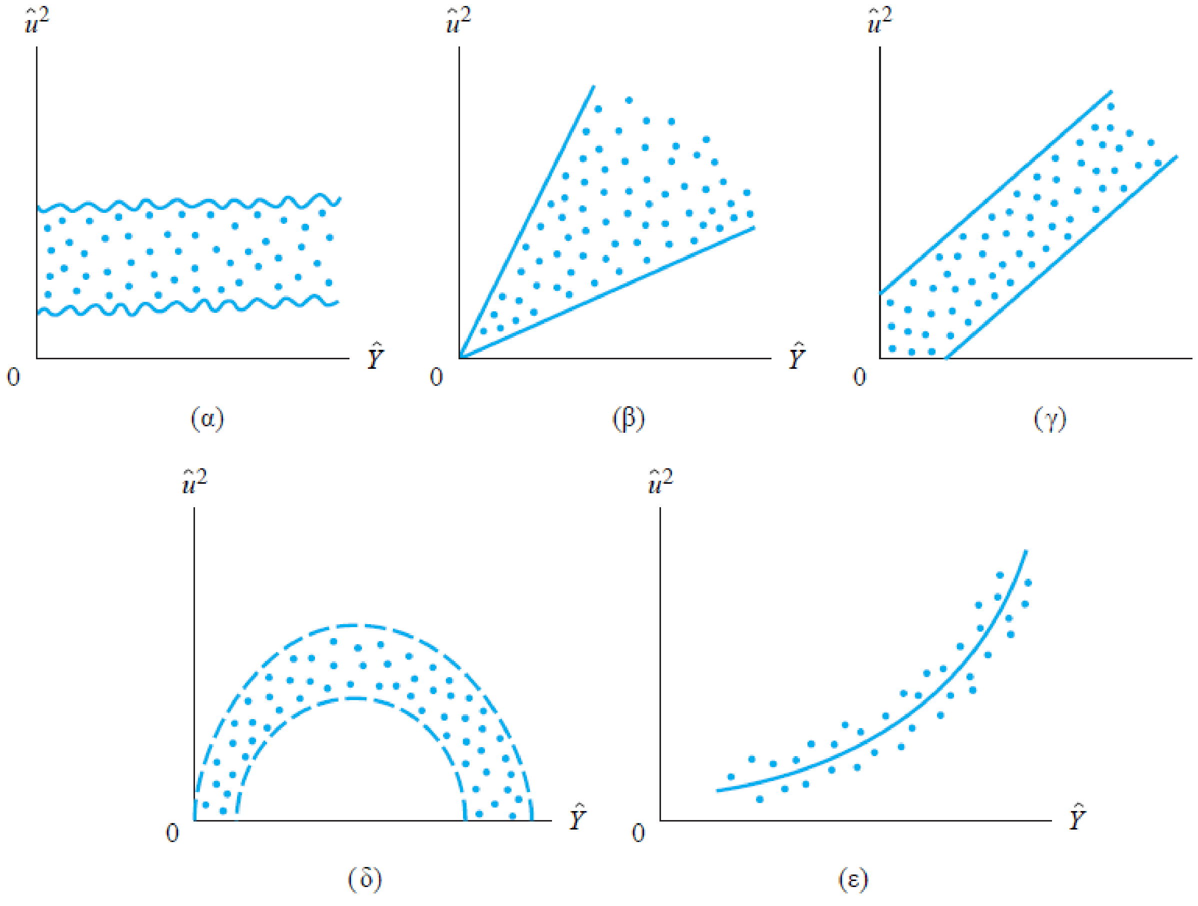
- Πώς μπορεί κάποιος να γνωρίζει ότι η ετεροσκεδαστικότητα είναι παρούσα σε μία συγκεκριμένη περίσταση; Και πάλι, όπως στην περίπτωση της πολυσυγγραμμικότητας, δεν υπάρχουν συγκεκριμένοι κανόνες για την ανίχνευση της ετεροσκεδαστικότητας, μόνο ορισμένοι εμπειρικοί κανόνες. Όμως η κατάσταση αυτή είναι αναπόφευκτη, επειδή η σ_i^2 μπορεί να είναι γνωστή μόνο αν έχουμε το συνολικό πληθυσμό *Y* που αντιστοιχεί στα επιλεγμένα *X*.
- Έχοντας κατά νου την προηγούμενη προειδοποίηση, ας εξετάσουμε ορισμένες από τις τυπικές και άτυπες μεθόδους για την ανίχνευση της ετεροσκεδαστικότητας.

3.5 Ανίχνευση της Ετεροσκεδαστικότητας

- **Άτυπες Μέθοδοι**
- *Η Φύση του Προβλήματος*
- *Γραφική Μέθοδος*

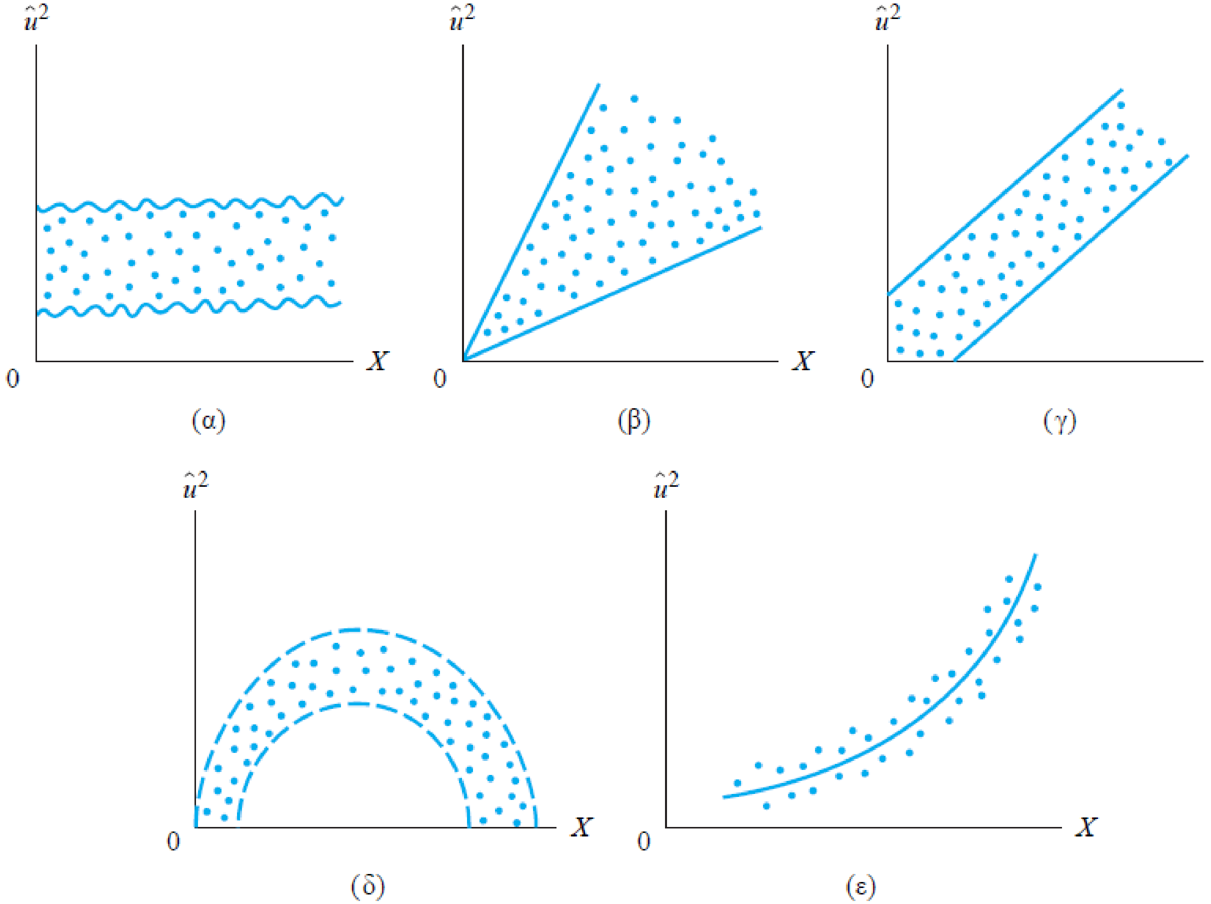
3.5 Ανίχνευση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.8** Υποθετικά πρότυπα των εκτιμημένων τετραγώνων των καταλοίπων.



3.5 Ανίχνευση της Ετεροσκεδαστικότητας

- **ΔΙΑΓΡΑΜΜΑ 3.9** Διάγραμμα διασποράς των εκτιμημένων τετραγώνων των καταλοίπων έναντι της *X*.



3.5 Ανίχνευση της Ετεροσκεδαστικότητας

- **Τυπικές Μέθοδοι**
- *Έλεγχος Park*
- *Έλεγχος Glejser*
- *Έλεγχος Συσχέτισης του Spearman*
- *Έλεγχος Goldfeld-Quandt*
- *Έλεγχος Breusch-Pagan-Godfrey*
- *Γενικός Έλεγχος Ετεροσκεδαστικότητας του White*
- *Άλλοι Έλεγχοι Ετεροσκεδαστικότητας*

3.5 Ανίχνευση της Ετεροσκεδαστικότητας

- *Μία Επισήμανση Αναφορικά με τους Ελέγχους Ετεροσκεδαστικότητας*
- Πώς θα αποφασίσουμε επομένως, ποιος είναι ο καλύτερος έλεγχος; Η απάντηση σε αυτό το ερώτημα δεν είναι εύκολη, καθώς αυτοί οι έλεγχοι βασίζονται σε διάφορες υποθέσεις. Όταν συγκρίνουμε αυτούς τους ελέγχους, πρέπει να επιστήσουμε την προσοχή μας στο μέγεθός τους (ή το επίπεδο σημαντικότητας), στη ισχύ τους (την πιθανότητα να απορρίψουμε μία ψευδή υπόθεση), και την ευαισθησία τους σε ακραίες τιμές.

3.6 Διορθωτικά Μέτρα

- **Υπάρχουν δύο προσεγγίσεις για τη διόρθωση:**
- όταν η σ_i^2 είναι γνωστή και
- όταν η σ_i^2 δεν είναι γνωστή

3.6 Διορθωτικά Μέτρα

- **Όταν η σ_i^2 Είναι Γνωστή: Η Σταθμισμένη Μέθοδος των Ελαχίστων Τετραγώνων**
- Αν η σ_i^2 είναι γνωστή, η πιο απλή μέθοδος διόρθωσης της ετεροσκεδαστικότητας είναι μέσω των σταθμισμένων ελαχίστων τετραγώνων, καθώς οι εκτιμητές που προκύπτουν είναι BLUE.

3.6 Διορθωτικά Μέτρα

- **Όταν η σ_i^2 Δεν Είναι Γνωστή**
- Αν η πραγματική σ_i^2 είναι γνωστή, μπορούμε να χρησιμοποιήσουμε τη μέθοδο WLS για την απόκτηση BLUE εκτιμητών.
- Δεδομένου ότι η πραγματική σ_i^2 είναι σπάνια γνωστή, υπάρχει κάποιος τρόπος να επιτύχουμε συνεπείς (με τη στατιστική έννοια) εκτιμήσεις των διακυμάνσεων και των συνδιακυμάνσεων των εκτιμητών της OLS, ακόμη και αν δεν υπάρχει ετεροσκεδαστικότητα; Η απάντηση είναι ναι.

3.6 Διορθωτικά Μέτρα

- Για να ολοκληρώσουμε την ανάλυσή μας για τα διορθωτικά μέτρα, θα επαναλάβουμε ότι όλοι οι μετασχηματισμοί που συζητήθηκαν προηγουμένως είναι *ad hoc*˙ ουσιαστικά, κάνουμε εικασίες σχετικά τη φύση της σ_i^2.
- Υπάρχουν ορισμένα πρόσθετα προβλήματα αναφορικά με τους μετασχηματισμούς τα οποία πρέπει να έχουμε κατά νου:

3.6 Διορθωτικά Μέτρα

1. Όταν κινούμαστε πέρα από το διμεταβλητό υπόδειγμα, μπορεί να μη γνωρίζουμε εκ των προτέρων ποια από τις μεταβλητές *X* θα πρέπει να επιλεγεί για το μετασχηματισμό των στοιχείων.
2. Ο λογαριθμικός μετασχηματισμός, όπως συζητήθηκε δε μπορεί να εφαρμοστεί εάν κάποιες από τις τιμές των *Y* και *X* είναι μηδενικές ή αρνητικές.

3.7 Μία Επιφύλαξη Σχετικά με Υπερβολικές Αντιδράσεις σε Σχέση με την Ετεροσκεδαστικότητα

- Όπως ισχυρίζεται ένας συγγραφέας, «η ετεροσκεδαστικότητα δεν υπήρξε ποτέ επαρκής λόγος για να απορρίψουμε ένα, κατά τα άλλα, καλό υπόδειγμα.»
- Στο σημείο αυτό μπορεί να είναι χρήσιμο να έχουμε κατά νου την επισήμανση του John Fox:
- …η άνιση διακύμανση σφάλματος αξίζει να διορθώνει μόνο όταν το πρόβλημα είναι σοβαρό.
- Ο αντίκτυπος της μη σταθερής διακύμανσης σφάλματος στην αποτελεσματικότητα του εκτιμητή των ελαχίστων τετραγώνων και στην εγκυρότητα της συμπερασματολογίας των ελαχίστων τετραγώνων εξαρτάται από πολλούς παράγοντες, συμπεριλαμβανομένου του μεγέθους του δείγματος, του βαθμού της μεταβλητότητας της σ_i^2, της διαμόρφωσης των τιμών της X [δηλαδή, των παλινδρομητών], και της σχέσης μεταξύ της διακύμανσης σφάλματος και των X. Ως εκ τούτου, δεν είναι δυνατόν να καταλήξουμε σε γενικά συμπεράσματα σχετικά με τη βλάβη που προκαλείται από την ετεροσκεδαστικότητα.

Περίληψη και Συμπεράσματα

1. Μία βασική υπόθεση του κλασικού γραμμικού υποδείγματος παλινδρόμησης είναι ότι οι διαταρακτικοί όροι u_i έχουν όλοι την ίδια διακύμανση, σ^2. Αν η υπόθεση αυτή δεν ισχύει υπάρχει ετεροσκεδαστικότητα.
2. Η ετεροσκεδαστικότητα δεν καταστρέφει τις ιδιότητες αμεροληψίας και συνέπειας των εκτιμητών της OLS.
3. Όμως αυτοί οι εκτιμητές δεν είναι πλέον ελάχιστης διακύμανσης ή αποτελεσματικοί. Δηλαδή, δεν είναι BLUE.
4. Οι εκτιμητές BLUE προκύπτουν από τη σταθμισμένη μέθοδο των ελαχίστων τετραγώνων, με την προϋπόθεση οι ετεροσκεδαστικές διακυμάνσεις σφάλματος, σ_i^2, να είναι γνωστές.

Περίληψη και Συμπεράσματα

5. Παρουσία της ετεροσκεδαστικότητας, οι διακυμάνσεις των εκτιμητών της OLS δεν παρέχονται από τους συνήθεις τύπους της OLS. Αν όμως επιμένουμε στη χρησιμοποίηση των συνήθων τύπων της OLS, οι έλεγχοι *t* και *F* που βασίζονται σε αυτούς μπορεί να είναι άκρως παραπλανητικοί, οδηγώντας σε εσφαλμένα συμπεράσματα.
6. Η τεκμηρίωση των συνεπειών της ετεροσκεδαστικότητας είναι πιο εύκολη από τον εντοπισμό της. Υπάρχουν διαθέσιμοι πολλοί διαγνωστικοί έλεγχοι, κανείς δε μπορεί όμως να είναι σίγουρος ποιος θα λειτουργήσει σε κάθε δεδομένη κατάσταση.

Περίληψη και Συμπεράσματα

7. Ακόμα κι αν η παρουσία ετεροσκεδαστικότητας εικάζεται και εξακριβώνεται, δεν είναι εύκολο να διορθώσουμε το πρόβλημα. Αν το δείγμα είναι μεγάλο, κάποιος μπορεί να αποκτήσει ετεροσκεδαστικά-διορθωμένα τυπικά σφάλματα του White των εκτιμητών της OLS και να προχωρήσει στην εξαγωγή στατιστικών συμπερασμάτων με βάση αυτά τα τυπικά σφάλματα.
8. Σε αντίθετη περίπτωση, βάσει των καταλοίπων της OLS, κάποιος μπορεί να κάνει μία εικασία των πιθανών προτύπων της ετεροσκεδαστικότητας και να μετασχηματίσει τα αρχικά στοιχεία με τέτοιο τρόπο, ώστε να μην υπάρχει ετεροσκεδαστικότητα στα μετασχηματισμένα στοιχεία.

Περίληψη και Συμπεράσματα

1. Μία από τις υποθέσεις της κλασικού γραμμικού υποδείγματος παλινδρόμησης είναι ότι δεν υπάρχει πολυσυγγραμμικότητα μεταξύ των ερμηνευτικών μεταβλητών, των *X*. Σε γενικές γραμμές, η πολυσυγγραμμικότητα αναφέρεται στην κατάσταση κατά την οποία υπάρχει είτε μία ακριβής είτε μία περίπου ακριβής γραμμική σχέση μεταξύ των μεταβλητών *X*.
2. Οι συνέπειες της πολυσυγγραμμικότητας είναι οι εξής: Αν υπάρχει τέλεια συγγραμμικότητα μεταξύ των *X*, οι συντελεστές παλινδρόμησής τους είναι απροσδιόριστοι και τα τυπικά σφάλματα τους δεν ορίζονται. Αν η συγγραμμικότητα είναι υψηλή, αλλά όχι τέλεια, η εκτίμηση των συντελεστών παλινδρόμησης είναι δυνατή, αλλά τα τυπικά σφάλματά τους τείνουν να είναι μεγάλα. Ως αποτέλεσμα αυτών, οι τιμές των συντελεστών του πληθυσμού δε μπορούν να εκτιμηθούν με ακρίβεια. Ωστόσο, εάν ο στόχος είναι να εκτιμηθούν οι γραμμικοί συνδυασμοί αυτών των συντελεστών, *οι εκτιμήσιμες συναρτήσεις*, αυτό μπορεί να γίνει παρά την παρουσία τέλειας πολυσυγγραμμικότητας.

Περίληψη και Συμπεράσματα

4. Η ανίχνευση της πολυσυγγραμμικότητας είναι μόνο ένα μέρος της μάχης. Το άλλο ασχολείται με το πώς μπορούμε να απαλλαγούμε από το πρόβλημα. Και πάλι δεν υπάρχουν δεδομένες μέθοδοι, παρά μόνο ορισμένοι εμπειρικοί κανόνες. Μερικοί από αυτούς τους κανόνες είναι οι εξής: (1) η χρήση προηγουμένων πληροφοριών, (2) ο συνδυασμός διαστρωματικών στοιχείων και χρονοσειρών στοιχείων, (3) η απαλοιφή μίας μεταβλητής που παρουσιάζει υψηλή συγγραμμικότητα, (4) μετασχηματισμός των στοιχείων, και (5) η χρησιμοποίηση επιπλέον ή νέων στοιχείων. Φυσικά, το ποιος από αυτούς τους κανόνες θα λειτουργήσει στην πράξη, θα εξαρτηθεί από τη φύση των στοιχείων και τη σοβαρότητα του προβλήματος της συγγραμμικότητας.
5. Επισημάναμε το ρόλο της πολυσυγγραμμικότητας στην πρόβλεψη και τονίσαμε ότι εάν η δομή της συγγραμμικότητας δε συνεχίζει να υφίσταται στο μελλοντικό δείγμα είναι επικίνδυνο να χρησιμοποιήσουμε την εκτιμημένη παλινδρόμηση, η οποία μαστίζεται από πολυσυγγραμμικότητα, για το σκοπό της πρόβλεψης.

Περίληψη και Συμπεράσματα

6. Παρά το γεγονός ότι η πολυσυγγραμμικότητα έχει λάβει εκτεταμένη (για άλλους υπερβολική) προσοχή στη βιβλιογραφία, ένα εξίσου σημαντικό πρόβλημα που εμφανίζεται στην εμπειρική έρευνα είναι αυτό της micronumerosity, λόγω του περιορισμένου μεγέθους του δείγματος. Σύμφωνα με τον Goldberger, «Όταν ένα ερευνητικό άρθρο προβληματίζεται αναφορικά με την ενδεχόμενη παρουσία πολυσυγγραμμικότητας, οι αναγνώστες πρέπει να εξετάσουν αν οι προβληματισμοί αυτοί θα συνέχιζαν να ισχύουν αν αντικαθιστούσαμε τον όρο «πολυσυγγραμμικότητα» με τον όρο «micronumerosity.» Σύμφωνα με τα λεγόμενά του ο αναγνώστης θα πρέπει να αποφασίσει πόσο μικρό είναι το n, ο αριθμός των παρατηρήσεων, πριν αποφασίσει ότι έχει πρόβλημα λόγω μικρού δείγματος, όπως ακριβώς πρέπει να αποφασίσει πόσο υψηλή είναι μία τιμή R^2 σε μία βοηθητική παλινδρόμηση πριν δηλώσει ότι το πρόβλημα της συγγραμμικότητας είναι πολύ σοβαρό.