$y_i = \alpha + \beta x_i + e_i \rightarrow$ απλό γραμμικό υπόδειγμα

y_i: food_exp (καταναλωτική δαπάνη του i νοικοκυριού)

x_i: income (εισόδημα του i νοικοκυριού)

- Θέλουμε να εκτιμήσουμε

$E(y_i / x_i) = \alpha + \beta x_i \rightarrow$ πληθυσμιακή γραμμή παλινδρόμησης

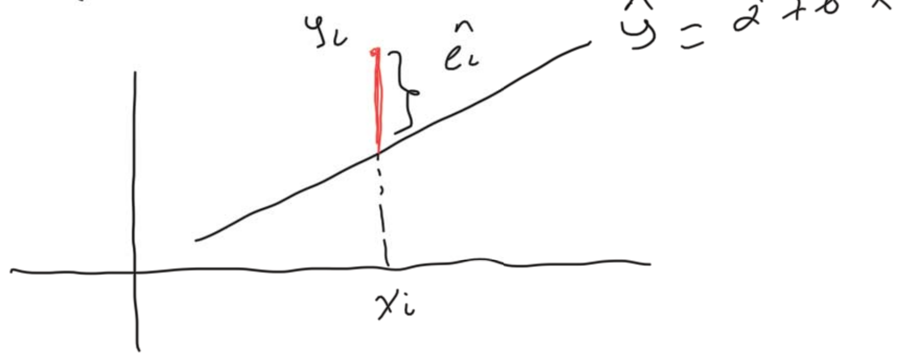
- Το εκτιμηθέν υπόδειγμα

$$\hat{y} = \hat{\alpha} + \hat{\beta} x_i \rightarrow \text{δειγματική γραμμή παλινδρόμησης}$$

$se(\hat{\alpha})$ $\quad$ $se(\hat{\beta}) \rightarrow$ Τυπικά σφάλματα εκτιμητών

$$t_{\hat{\alpha}} = \frac{\hat{\alpha}}{se(\hat{\alpha})} \sim t_{n-2}, \quad t_{\hat{\beta}} = \frac{\hat{\beta}}{se(\hat{\beta})} \sim t_{n-2}$$

- $y_i - \hat{y}_i = \hat{e}_i \Rightarrow$ κατάλοιπο της παλινδρόμησης για την παρατήρηση i



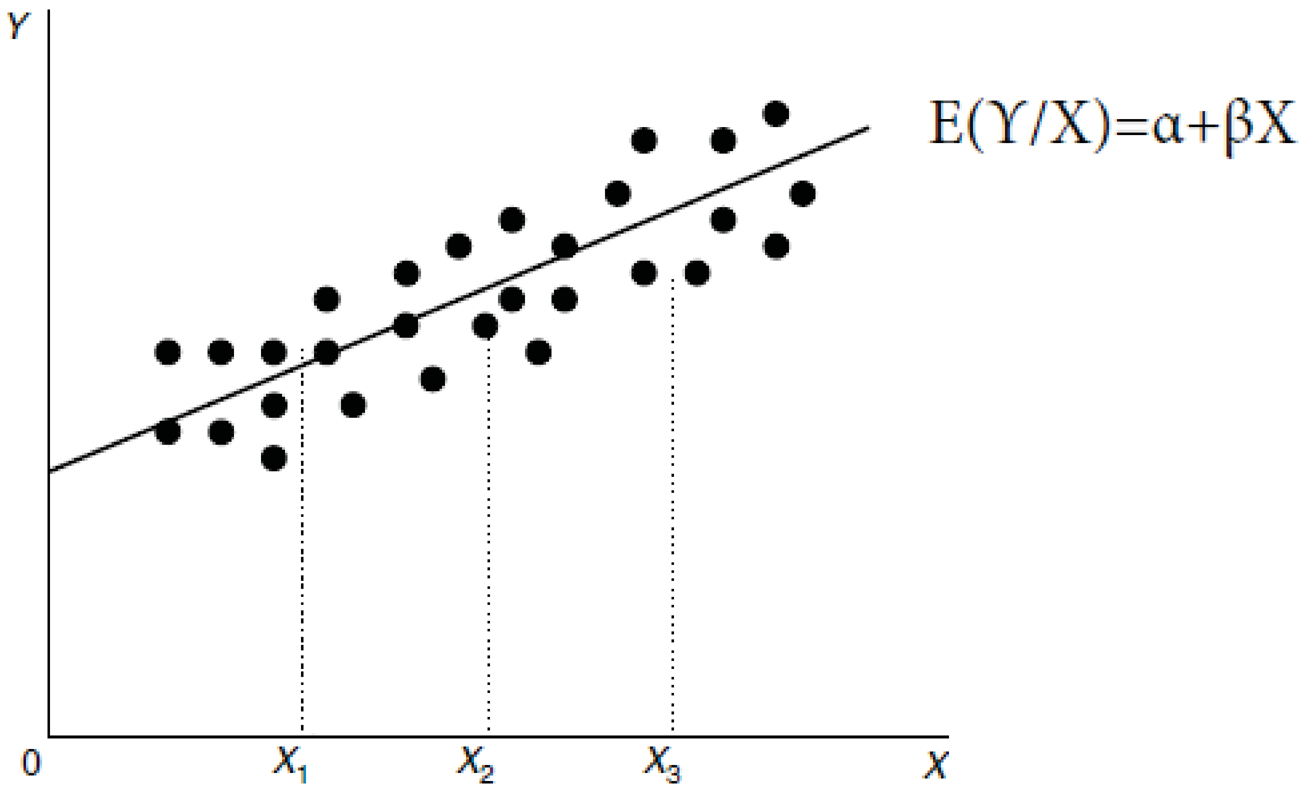
- $\hat{\sigma}^2 = \sum \hat{e}_i^2 / n-2$ → εκτιμητής διακύμανσης διαταρακτικού όρου

- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ ⇒ Τυπικό σφάλμα της παλινδρόμησης (standard error of the regression)

Ετεροσκεδαστικότητα

$$y_i = \alpha + \beta x_i + e_i$$

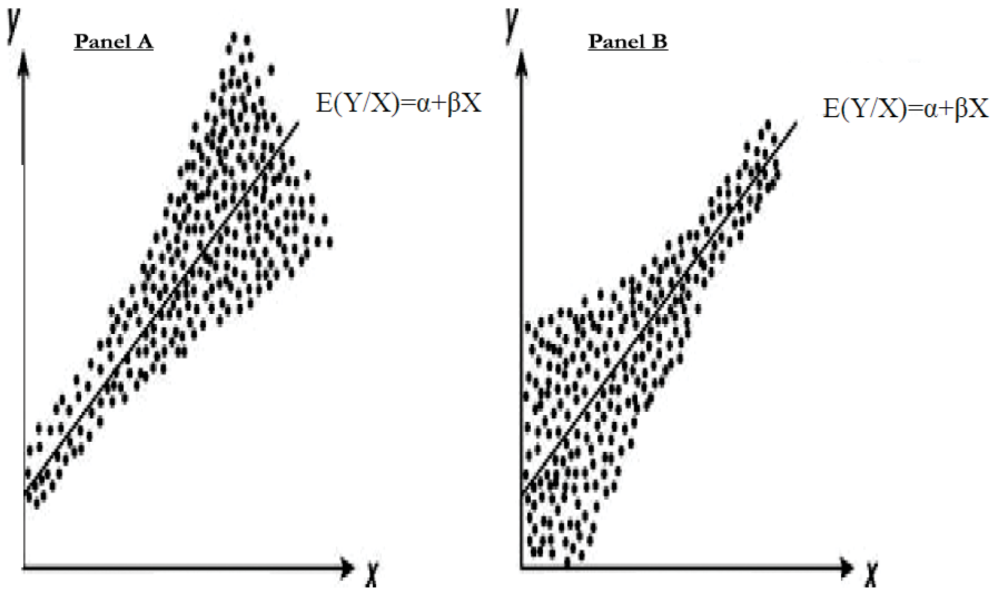
- $var(y_i/x_i) = var(e_i/x_i) = \sigma^2 \quad \forall i$ → ομοσκεδαστικότητα



Δεδομενα με σταθερη διακυμανση-

ομοσκεδαστικοτητα

• $var(y_i/X_i) = var(e_i/X_i) = \sigma_i^2 \rightarrow$ ετεροσκεδαστικότητα



Αυξανομενες και μειουμενες ετεροσκεδαστικες διακυμανσεις

Έλεγχοι ετεροσκεδαστικότητας

Goldfeld–Quandt

$$y_i = \alpha + \beta x_i + e_i$$

1) Καταταγούμε τις παρατηρήσεις της X κατά αύξουσα τάξη μεγέθους και αντίστοιχα αναδιατάσσονται και οι παρατηρήσεις της Y

2) Διαιρούμε το δείγμα σε 2 τμήματα, N_1, N_2
Το πρώτο (N_1) περιέχει χαμηλές τιμές της X
Το δεύτερο (N_2) >> υψηλές >> >>

3) Τρέχουμε την παλινδρόμηση για καθένα από τα 2 τμήματα και υπολογίζουμε το εκτιμητή της διακύμανσης του διαταρακτικού όρου

1η παλινδρόμηση $\rightarrow \hat{\sigma}_1^2$

2η >> $\rightarrow \hat{\sigma}_2^2$

4) $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_2^2 > \sigma_1^2$

5) $$F_{GQ} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{RSS_2 / N_2-k-1}{RSS_1 / N_1-k-1} \sim F_{N_2-k-1,\, N_1-k-1}$$

όπου $k = \#$ των ανεξάρτητων μεταβλητών στο παλινδρομικό εκτός του σταθερού όρου

Άρα εδώ $k=1$

6) $F_{GQ} > F_{critical} \Rightarrow$ Απορρίπτουμε H_0
Άρα έχουμε ετεροσκεδαστικότητα

ΕΜΕΙΣ ΘΕΤΟΥΜΕ $N_1=N_2=20$

Breusch-Pagan

$y_i = \alpha + \beta x_i + e_i$

1) Τρέχουμε τη παλινδρόμηση και υπολογίζουμε τα $\hat{e}_i^2$

2) ορίζουμε $p_i = \hat{e}_i^2 / \hat{\sigma}^2$, όπου $\hat{\sigma}^2 = \dfrac{\sum \hat{e}_i^2}{N}$

3) Τρέχουμε την βοηθητική παλινδρόμηση

$$p_i = \delta_0 + \delta_1 x_i + v_i$$

4) $H_0: \delta_1 = 0$ (ομοσκεδαστικότητα)
$H_1: \delta_1 \neq 0$ (ετεροσκεδαστικότητα)

5) $LM_{BG} = \dfrac{1}{2} ESS \sim \chi^2_r$, r = # ανεξάρτητων μεταβλητών στη βοηθητική παλινδρόμηση εκτός του σταθερού όρου. Άρα εδώ r=1

όπου $ESS = \sum_{i=1}^{N} (\hat{p}_i - \bar{p})^2$

6) $LM_{BG} > \chi^2_{critical} \Rightarrow$ Απορρίπτουμε H_0
Άρα έχουμε ετεροσκεδαστικότητα

$$R^2 = \frac{ESS}{TSS}$$ όπου $TSS = \sum (P_i - \bar{P})^2$

Άρα $ESS = R^2 \cdot TSS$

Godfrey

Επαναλαμβάνουμε τα ίδια βήματα με πριν μόνο που χρησιμοποιούμε διαφορετική βοηθητική παλινδρόμηση

$$\hat{e}_i^2 = \delta_0 + \delta_1 X_i + u_i$$

και άρα η στατιστική που σχετίζεται με την ανωτέρω παλινδρόμηση είναι

$$LM_G = NR^2 \sim \chi^2_{(1)}$$

Εάν $LM_G > \chi^2_{(1)}$ απορρίπτουμε Ho

και έχουμε ετεροσκεδαστικότητα

White

$y_i = a + \beta x_i + e_i$

1) τρέχουμε η υπόδειγμα και υπολογίζουμε τα $\hat{e}_i$

2) τρέχουμε τη βοηθητική παλινδρόμηση

$$\hat{e}_i^2 = \delta_0 + \delta_1 x_i + \delta_2 x_i^2 + v_i$$

3) $H_0: \delta_1 = \delta_2 = 0$ (ομοσκεδαστικότητα)

H_1: ένα από αυτά τουλάχιστον $\neq 0$. (ετεροσκεδαστικότητα)

4) $LM_w = NR^2 \sim \chi^2_r$. εδώ $r = 2$

5) $LM_w > LM_{critical} \Rightarrow$ Απορρίπτουμε H_0

Άρα έχουμε ετεροσκεδαστικότητα

Συνέπειες ετεροσκεδαστικότητας

$$y_i = \alpha + \beta x_i + e_i$$

Έστω ότι το μοντέλο αυτό πάσχει από ετεροσκεδαστικότητα

Τότε $se(\hat{\alpha}) \neq se^{E}(\hat{\alpha})$

$$se(\hat{\beta}) \neq se^{E}(\hat{\beta})$$

$se(\hat{\beta})$ → όταν αγνοούμε την ύπαρξη ετεροσκεδαστικότητας → όλοι οι στατιστικοί έλεγχοι και τα διαστήματα εμπιστοσύνης είναι αναξιόπιστα

$se^{E}(\hat{\beta})$ → όταν λαμβάνουμε υπόψιν την ύπαρξη ετεροσκεδαστικότητας → όλοι οι στατιστικοί έλεγχοι και τα διαστήματα εμπιστοσύνης είναι αξιόπιστα

Αντιμετώπιση ετερο-σκεδαστικότητας

A) Τυπικά σφάλματα του white

- Αυτή η μέθοδος χρησιμοποιείται όταν δεν γνωρίζουμε την μορφή της ετεροσκεδαστικότητας

- Ο white προτείνει μια μαθηματική έκφραση για τα τυπικά σφάλματα των εκτιμητών που είναι αξιόπιστα απέναντι σε κάθε μορφή ετεροσκεδαστικότητας

- Έστω ότι τα τυπικά σφάλματα του white είναι:

  $se_w(\hat{\alpha})$ και $se_w(\hat{\beta})$

- Η μέθοδος δουλεύει ως εξής που έχει ετερο-σκεδαστικότητα

1) Τρέχουμε το μοντέλο

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

$se(\hat{\alpha})$ $se(\hat{\beta})$ → λάθος τιμές

$\hat{\alpha}, \hat{\beta}$ → αμερόληπτοι και συνεπείς εκτιμητές

2) Αντικαθιστούμε τα $se(\hat{\alpha})$ και $se(\hat{\beta})$ από τα σωστά $\underbrace{se_w(\hat{\alpha}) \text{ και } se_w(\hat{\beta})}_{\text{τυπικά σφάλματα του white}}$

γ) όλοι οι στατιστικοί έλεγχοι (t) διαστήματα εμπιστοσύνης κατασκευάζονται με βάση τα τυπικά σφάλματα του white [$se_w(\hat{\alpha})$ και $se_w(\hat{\beta})$]

β) Λογαριθμικός μετασχηματισμός του υποδείγματος

γ) GLS

Αυτή η μέθοδος χρησιμοποιείται όταν γνωρίζουμε τη μορφή ετεροσκεδαστικότητας

Παράδειγμα:

$$y_i = \alpha + \beta x_i + e_i$$

Ας υποθέσουμε ότι $V(e_i) = \sigma_i^2 = \sigma^2 x_i$ τότε μετασχηματίζουμε το υπόδειγμα ως εξής

$$\frac{y_i}{\sqrt{x_i}} = \frac{\alpha}{\sqrt{x_i}} + \beta \frac{x_i}{\sqrt{x_i}} + \frac{e_i}{\sqrt{x_i}}$$

$$E\left(\frac{e_i}{\sqrt{x_i}}\right) = 0 \quad \text{και} \quad V\left(\frac{e_i}{\sqrt{x_i}}\right) = \frac{1}{x_i} \cdot V(e_i) = \frac{\sigma^2 x_i}{x_i} = \sigma^2$$

Άρα 2 μετατρέψαμε σε αυτό υπόδειγμα

$$\frac{y}{\sqrt{x_i}} = \frac{a}{\sqrt{x_i}} + b\frac{x_i}{\sqrt{x_i}} + \frac{e_i}{\sqrt{x_i}}$$

Είναι ομοσκεδαστικό